



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

Dursun Akaslan¹, Umut Kuran²

Abstract

It is relatively important to understand terms in order to be successful in Higher Education since they are used to refer a particular thing in a technical or scientific object, especially in engineering. The terms originated from English and French seem to be the most prevalent ones in comparison with other foreign languages in the fields of Electrical, Electronic and Computer Engineering in Turkey. However, majority of those terms have their own synonyms in the Turkish language and they are used as a substitute for their Turkish counterparts. The utilization of the terms in this manner leads to multilingualism in many fields such as engineering, medicine and other disciplines. Henceforth, it is evident that the number of academic studies pertaining to the impacts of the use of the terms originated from foreign languages on the Turkish language has been increasing in recent years. For instance, the academic studies point out that the use of foreign originated terms impairs the Turkish language in terms of various aspects such as vowel harmony, spelling and utterance. Additionally, these issues aggravate the process of teaching and learning the Turkish language as a foreign language. The goal of this study to understand the impacts of multilingualism by measuring the prevalence of foreign originated terms (e.g. server) and their synonyms (e.g. sunucu) using various search engines namely Bing and Yandex, especially in the fields of Electric, Electronic and Computer Engineering. To achieve our objectives, the Foreign Word Equivalence Manual, which was made available by the Turkish Language Association, was used in the study. The number of foreign originated words and their equivalence were computed as 1948 and 2265 and categorized into 41 groups (e.g. medicine, astronomy & chemistry) because a specific term may have more than one equivalence in the Turkish language. Descriptive and inferential statistics were also computed to find out whether there is a significant difference among categories. Additionally, the findings of our study also compares the significant differences among search engines.

1. Introduction

Terms are substantially used in Higher Education with the purpose of referring to a word or expression used in relation to a particular thing in a subject and often used to describe something official or technical, especially in engineering. However, majority of terms used in the Turkish Language are originated from foreign languages and English and French seem to be the most prevalent ones in comparison with other foreign languages such as German and Spanish [1]. Moreover, the Foreign Word Equivalence Manual published by the Turkish Language Organization indicates that subjects such as philosophy, physics, medicine, sport and economy are the most prevalent ones in terms of the utilisation of foreign originated terms in the Turkish language. Akaslan, et al. (2015) also emphasize that recent advancements in Information and Communication Technologies (ICTs) with the apprance of the Internet are also considered as a reason for the prevalence of foreign originated terms because latest developments in ICTs are usually discovered by developed countries and those result in the need for new words in Turkish [2]. On the other hand, majority of those terms have their own synonyms in the Turkish language and they are used as a substitute for their Turkish counterparts. The simultaneous utilization of foreign originated terms and their synonyms together leads to multilingualism in many fields such as physics, computing, electric, electronic and so on in terms of the utilization of terms, especially in Higher Education. Hence, it is deemed relevant to understand the impacts of multilingualism on the Turkish language in terms of the utilization of terms in various disciplines. In this regard, the goal of this study is two-fold. First, we aim to understand the prevalence of the foreign originated terms and their synonyms in the Turkish language using search engines namely the Bing and Yandex Search Engine with the support of descriptive statistics. Second, based on the prevalence of the foreign terms, we investigate whether there are significant differences among terms in various disciplines by using inferential statistics.

¹ Harran University, Computer Engineering Department, (Turkey)

² Harran University, Computer Engineering Department, (Turkey)



2. Methodology

This section is divided into four parts: The first part reports the properties of the data set used in the paper. The second part illustrates how search engines namely Bing and Yandex were customized and filtered to ensure that the findings are associated with the impacts of multilingualism on the Turkish language. The third part develop a formula for computing the impact factor of terms within the Turkish language with the purpose of indicating the relative importance of them. The last part classifies the terms within various levels to conduct descriptive and inferential statistics.

2.1. Dataset

The Turkish Language Organization (TDK) prints a number of great dictionaries (e.g. the Great Turkish Dictionary) in order to notify Turkish readers and writers regarding the origins of the foreign originated words in Turkish to ensure that they notice the derivation of foreign originated from various languages such as Arabic, English and French. However, the use of the Turkish equivalence of the foreign originated terms has been perceived as an important solution in academia to deal with the impacts of multilingualism on the Turkish language instead of studying the origins of words in Turkish [3-4]. The TDK made available the Turkish equivalence of 1948 foreign originated words using a Foreign Word Equivalence Manual (FWEM) in 2008 with their disciplines and languages. The FWEM they will be used as a dataset to understand the impacts of foreign originated terms on the Turkish language.

2.2. Search Engines

The amount of information on the web is growing rapidly because of hundreds of millions of web pages are connected to the Internet [5]. Today, search engines such as Google, Yandex, Bing and Yahoo are used to conduct searchers nearly everywhere on the Web using many tools such as toolbars and widgets. Search engines are mainly designed as a program in order to find information on the Internet by looking for words which are typed in. Many analytics (e.g. comScore & Statista) captures all of the search behaviour at search engines in various countries. The July Report of the Statista in 2016 illustrates that the majority of market share of search queries in the United States are handled by Google (63.4%), Microsoft (21.9%) and Yahoo (12.1%). Similarly, search engines might be also used to measure the number of the foreign originated terms and its synonyms on the Turkish language to understand the impact of multilingualism on the Turkish language in various disciplines. 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 Google, Yandex, Bing and Yahoo. Moreover, search results can be also filtered and customized using search operators namely quotes to include the pages with the same words in the same order as the ones inside the quotes. For example, the Bing Search were customized using search operators and filtered selecting only the Turkish language and Turkey with 26.600 results because we are only interested in impacts of the word “abluka” on the Turkish language rather than its one on the earth.

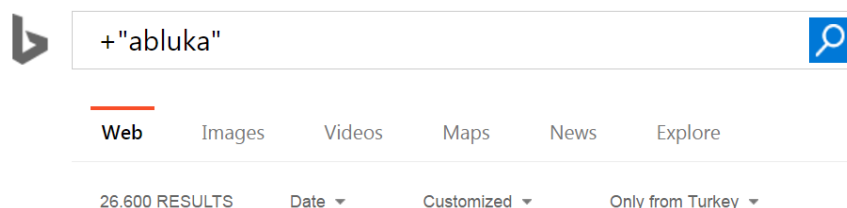


Figure 1. Search results

2.3. Impact Factor

The number of foreign originated terms and its synonyms appearing in the search engines must be evaluated to understand the impacts of multilingualism on the Turkish language in terms of the aspects of various disciplines such as chemistry and medicine. An impact factor might be described as a measure reflecting the importance or power of a term within the Turkish language. It can be used as a proxy for relative importance of a term within the Turkish language, with terms with higher impact factors deemed to be more important than those with lower ones like impact factors used in journals. 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. In order to calculate the impact factors of terms, as illustrated in Fig. 2, the formula is:



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

Figure 2. Impact Factor Formula

where R is the ratio of the number of search results in foreign languages to the one in Turkish language or vice versa, C is a constant in order to standardize the impact factor of terms between -5 and +5 and IF is the power or impact factor of terms in Turkish language. Table 1 illustrates the values of C of the terms in the Turkish language. Moreover, the R should be multiplied with -1 if the number of the search results of a foreign originated term is higher than its Turkish synonym or with +1 if vice versa due to its positive or negative effect on the Turkish language. For example, the word “adaptasyon” is a foreign originated term and is derived from the word “adaptation” in the French language and its synonym is written as “uyarlama” in the Turkish language. According to the Yandex Search Engine, the word “adaptasyon” appears 229,000 times and its synonym “uyarlama” 396,000 times in the search results. According to the formula as given in Figure 3, the impact factor of the term can be calculated as:

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

where 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 and IF = 0.8646.

Table 1. Constant Values

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

2.4. Classification

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. The C and R values might be used to determine levels of the terms as illustrated in Table 2.

Table 2. Constant Values

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

For example, the foreign originated term “stilizasyon” and its synonym “biçemleme” appears 15,000 and 3,000 times in the Yandex Search Engine respectively. On the other hand, the foreign originated term “klasik” and its synonym “kökleşik” appears 3,000,000 and 6,000 times in the Yandex Search Engine in order. However, the impact factor of both terms is calculated as -2.500 whereas their levels are labelled as A1 and B1. Hence, it is important to note that terms within the same level should be evaluated or assessed together. Statistical tests such as descriptive and inferential statistics should not be used or applied to terms within the dissimilar levels.



3. Findings

This section is divided into two parts: The first part reports statistics among terms in the study associated with their disciplines whereas the second part compares the impact factors of terms within the disciplines related to their original languages. The number and mean scores of the impact factors of terms within various languages are presented in the Table 3 using search results from the Bing and Yandex Search Engines. This study revealed that the majority of the terms in the FWEM were derived from the French language (i.e. common) with 1836 terms. Besides, the terms derived from the French language are mainly classified within levels of the A1 and A2 as illustrated in the Table 3. The overall impact factors indicate that there is no positive impact factor. This concludes that the usage of the foreign originated terms is higher than their Turkish synonyms for all the levels from A1 to C2. From the table, it can be observed that there are few terms in the level of C2 according to the Bing and Yandex search results namely: forum (forum-toplu tartışma), injection (enjekte-iç itilmiş), peak (pik-değer bakımından yükselme), royalty (röyalti-telif hakkı), projector (projektör-yansıtım aygıtı) and onomatopoeic (onomatopeik-yansımalı).

Table 3. The impact factor of the terms associated with their languages

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)

Y: Yandex; B: Bing

The mean scores of the impact factors of the terms associated with their disciplines were also illustrated in Table 4. As can be seen from the table, the majority of mean scores of the impact factors are negative in disciplines from anatomy to zoology. On the other hand, the Turkish synonyms of the terms in disciplines such informatics and history in A1, military and mineralogy in A2. Moreover, the search results of both the Yandex and Bing seem to be consistent in most disciplines and languages as given in Table 3 and 4.

4. 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 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

We 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.

Table 4. The impact factor of the terms associated with their disciplines

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	-	-	-	-	-	-	-	-
<i>grammar</i>	42	0,13	0,68	-0,13	0,52	0,79	0,23	-	2,39	-	-	-	0,00
<i>history</i>	3	0,24	1,50	-	-	-	-	-	-	-	-	-	-
<i>informatics</i>	21	0,50	0,80	-1,23	-1,27	-1,97	-4,17	-	-1,12	-	-	-	-
<i>law</i>	14	-0,69	0,74	-1,04	-0,04	1,64	-1,07	-	-0,96	-	-	-	0,00
<i>linguistics</i>	12	-1,07	-1,50	-0,18	0,98	-	-	-	-	-	-	-	-
<i>literature</i>	20	-1,15	-0,18	0,13	-0,31	-0,25	0,09	-	-	-	-	-	-
<i>logic</i>	6	1,09	-2,40	-3,16	0,57	-2,00	-4,35	-	-0,54	-	-	-	-
<i>marine</i>	8	0,63	-1,59	-0,67	-0,76	-1,19	-2,09	-	-	-	-	-	-
<i>mathematics</i>	24	-1,04	-0,04	-0,78	-0,53	-1,67	-0,09	-1,16	-1,16	-	-	-	-
<i>medicine</i>	90	-0,31	0,03	-0,83	-1,02	-0,81	-0,97	-	-1,23	-	-	-0,09	0,00
<i>meteorology</i>	3	-1,16	-1,56	-1,32	-0,63	-	-	-	-	-	-	-	-
<i>military</i>	7	-1,53	-1,01	1,83	-1,68	-1,67	-0,54	-	-	-	-	-	-
<i>mineralogy</i>	9	-0,46	-0,47	1,64	3,10	-	1,05	-	-	-	-	-	-
<i>music</i>	22	0,85	1,29	0,89	0,65	0,77	3,01	-	0,85	-	-	-	-
<i>philosophy</i>	142	-0,33	-0,10	-0,67	-0,04	-1,07	-2,67	-	-0,64	-	-	-	-
<i>physics</i>	90	-0,64	0,03	-0,76	-0,96	-1,01	-1,41	-0,71	-0,71	-	-	-	-
<i>physiology</i>	3	0,44	-1,71	-	-	-	-	-	-	-	-	-	-
<i>religion</i>	4	-0,87	0,18	-0,78	-1,25	-	-	-	-	-	-	-	-
<i>society</i>	51	-0,18	-0,53	-1,27	0,45	1,11	0,29	-	-	-	-	-	-
<i>sport</i>	74	0,11	0,27	-0,84	-0,60	-1,21	-1,16	-1,44	-0,72	-	-	-	-
<i>sprit</i>	49	-1,04	-0,56	-0,78	-1,80	-0,19	-0,77	-	-1,07	-	-	-	-
<i>technic</i>	8	-0,36	-1,17	-4,76	-0,02	-0,58	-1,79	1,02	1,39	-	-	-	-
<i>technology</i>	1	-	4,43	0,56	-	-	-	-	-	-	-	-	-
<i>television</i>	15	-0,20	-0,90	-0,03	0,14	-1,34	-4,82	-	-	-	-	-	-
<i>theater</i>	7	-0,28	-0,95	0,12	-0,59	1,67	-	-1,43	1,69	-	-	-	-
<i>trade</i>	26	-0,95	-1,09	-0,16	-0,33	-1,52	0,14	-	-1,21	-	-	-	-
<i>zoology</i>	9	-0,70	1,25	0,68	-1,60	-	0,71	-0,50	-	-	-	-	-
overall	2265	-0,22	-0,07	-0,59	-0,55	-0,68	-0,69	-0,78	-0,69	0,00	0,00	-0,08	0,00

References

- [1] Tekten, T., Türk Dilindeki Yabancı Kelimelerin Türkçe Öğretimine Etkileri. 1st International Symposium on Sustainable Development, 2009, Sarajevo, Bosnia and Herzegovina.
- [2] Akaslan, D., Salur, M. U., Tenekeci, M. E. & Gümüşçü, A. "Designing a Relational Database to Alleviate the Impacts of Foreign Words on the Turkish Language". International Journal of Social Sciences and Education Research, 2015, pp. 630-641.
- [3] Akaslan, D., Salur, M. E., Gümüşçü, A. & Tenekeci, M. E., Measuring the Prevalence of Foreign Originated Words on the National Newspapers using Search Engines "10th International Computer & Instructional Technologies Symposium (ICITS), 16-18 May 2016, Rize, Turkey
- [4] Akaslan, D., Tenekeci, M. E. & Gümüşçü & Salur, M. U A., Designing a relational database to alleviate the impacts of foreign words on the Turkish language, International Conference on Social Sciences and Education Research (ICSER), 29-31 October 2015, Antalya, Turkey
- [5] Brin, S. & Page, L. "Reprint of: The anatomy of a large-scale hypertextual web search engine". Computer Networks, 2012, pp. 3825–3833.