Open Access **Baghdad Science Journal** **P-ISSN: 2078-8665**
Published Online First: May 2022 **2022, 19(6): 1369-1377** **E-ISSN: 2411-7986**

DOI: https://dx.doi.org/10.21123/bsj.2022.6008

Using Graph Mining Method in Analyzing Turkish Loanwords Derived from Arabic Language

***Abbood Kirebut Jassim*[1*]** ***Muneam Jabbar Hamzah*[1]** ***Ahmed Hussein Aliwy*[2]**

[1]Department of Computer of Science, College of Science for Women, University of Baghdad, Baghdad, Iraq.
[2]Department of Computer of Science, Faculty Computer Science and mathematics, University of Kufa, Najaf, Iraq.
[*]Corresponding author: abboodkj_comp@csw.uobaghdad.edu.iq
E-mail addresses: muneamjh_comp@csw.uobaghdad.edu.iq , ahmedh.almajidy@uokufa.edu.iq

Received 11/1/2021, Accepted 28/11/2021, Published Online First 20/5/2022, Published 1/12/2022

Abstract:

Loanwords are the words transferred from one language to another, which become essential part of the borrowing language. The loanwords have come from the source language to the recipient language because of many reasons. Detecting these loanwords is complicated task due to that there are no standard specifications for transferring words between languages and hence low accuracy. This work tries to enhance this accuracy of detecting loanwords between Turkish and Arabic language as a case study. In this paper, the proposed system contributes to find all possible loanwords using any set of characters either alphabetically or randomly arranged. Then, it processes the distortion in the pronunciation, and solves the problem of the missing letters in Turkish language relative to Arabic language. A graph mining technique was introduced, for identifying the Turkish loanwords from Arabic language, which is used for the first time for this purpose. Also, the problem of letters differences, in the two languages, is solved by using a reference language (English) to unify the style of writing. The proposed system was tested using 1256 words that manually annotated. The obtained results showed that the f-measure is 0.99 which is high value for such system. Also, all these contributions lead to decrease time and effort to identify the loanwords in efficient and accurate way. Moreover, researchers do not need to have knowledge in the recipient and the source languages. In addition, this method can be generalized to any two languages using the same steps followed in obtaining Turkish loanwords from Arabic.

Keywords: Arabic language, Data mining, Graph mining, Loanwords, Turkish language.

Introduction:

Data mining is the task of extracting useful information from data that has been performed by analysts. There are two primary goals of data mining which are prediction and description [1, 2]. Predictive data mining produces the model of the system described by a given data set, whereas descriptive data mining produces new and nontrivial information based on an available data set [3].

In general, a graph is a group of nodes which are almost represented by circles; the link between each two nodes, which is represented by a line, is called an edge [4]. That means the graph is represented by a set of connected nodes by edges. For example, the routers or computers (nodes) are connected by wires or wireless (links) [5]. These nodes and links overall are represented by a tree [6]. The field of extracting useful information by traversing a tree looking for specified nodes is called the graph mining [7]. Graph mining has become an important topic of research recently because of numerous applications and a wide variety of data-mining problems in computational biology, chemical data analysis, drug discovery, and communication networking. Also, graph mining can be used in loanwords from other nonnative language [8].

Loan words are words coming from a lending language to a receptor language [9]. The language of the original word is called “donor”, “lending”, “source”, or “borrowing” language. Whereas, a language receiving words from a lending language is called “receptor” or “recipient” language [10]. Taking some words from some languages is beneficial to compensate some missing meanings in the recipient language [10]. This phenomenon contributed in a very important way to the fields of bilingualism [11]. The main factor for choosing the Turkish language in this

work is that Turkish language has adopted many words from Arabic language during the ottoman invasion.

Ottoman is the old name of Turkish republic. They invaded and occupied Iraq in 1534. Iraq was under their control until 1623. During that era, Turkish troops adopted many words and transferred those words to their country after they left Iraq. Because of this invasion, many Arabic words found their way into Turkish language. These adopted words from Arabic into Turkish are called loan words [12].

Loan words in Turkish are the words which are taken from Arabic language and integrated in Turkish. Some of these words are transferred from the lending language (Arabic) to the borrowing language (Turkish) without any changes in the pronunciation. For Example; "تمام" is an Arabic word loaned to Turkish; this word has the same pronunciation in the source (Arabic) and recipient (Turkish) languages which is "tamam". Most of the loan words in Turkish have been subjected to slight changes. For example, the word "كتاب" is a Turkish loanword from Arabic, but the pronunciation of this word is a little bit different in the two languages; it pronounces in Turkish "kitap", but pronounces in Arabic "kitab". There are many other examples but one is taken in the two situations for explanation. That means, a phonological changes have been happened to the Turkish loanwords for adaptation to be fitted into their native language. In general, borrowing words from some languages is to fill the lexical gaps in the language's dialect.

In this paper, the reasons that led to the great interaction between the Arabic and Turkish languages were explained, which led to the large number of Turkish words borrowed from the Arabic language. The researches on this topic have been reviewed. Graph mining was used to find loanwords by generating and filtering all possible words. The changes in words were also processed to reflect the mutation that occurred when words switched between two different languages. It can be emphasized that this method can be generalized to find loanwords between any two different languages.

Finding the origin of words is one of the most difficult tasks for linguists, and it is the main motivation of the proposed work. Also, the existing works in this field are very limited with low accuracy. The main contributions, for our research, are by (i) using graph mining techniques for detection loanwords, and (ii) dealing with Turkish and Arabic language, it is used for the first time and (iii) building an efficient system for detecting the loanwords. This paper has two main contributions. First, detection loanwords by using graph mining techniques, and dealing with the Turkish and Arabic languages. Second, it is used for the first time to detect loanwords in efficient way.

Related Works:

There are few works for automatic detection of loanwords but none of them dealt with Turkish-Arabic loanwords. In this section almost all the related works with any pair of languages are presented.

Buhmaid S [9] stated that many words from English penetrated the Hadhrami Arabic. He used specific phonological, morphological and semantic features for detection of loanwords. Salman et. al [13] showed that "borrowing can be achieved when some words are imported from one language to the lexicon of another language". They discussed the adaption process in the level of phonology and morphology. Loanword adaptations represent phonetically minimal transformations". Using the original form is not always; sometimes some changes happen to the borrowed words in order to be compatible with the structure and rules of the borrowing language.

Peperkamp et. al [14] showed that there are three problems in loanword adaption which are learnability, phonetically driven adaption, and unfaithful perception.

Peperkamp S [15] stated that some of the loanwords are not compatible with the native phonology, and the adaption process is subject to a minimal transformation. Rao [16] discussed loanwords between English and other eleven languages. He showed that borrowing words into English language is very important because the target language needs to fill gaps in their lexicon. Also, he states that most of the words find their way into English from other languages such as Arabic, Greek, Russian, Spanish, French, Latin, etc. Farazandeh-pour et. al [17] described a mechanism of adaption related to Persian words coming from German origins.

Mi *et al.* [18] proposed a recurrent neural network (RNN) based framework to identify loanwords (Chinese, Russian and Arabic loanwords) in Uyghur. They also suggested two features: inverse language model and collocation feature to optimize the output of loanword identification model. They stated that the model achieved significant improvements in loanwords detection task.

Koo [19] stated, that the loanwords coming from other languages to Korean can be identified by using unsupervised classifier. The results showed that the F-score of the classifier is 94.77. Also, he showed that the method can also be applied to other languages that have the same phoneme such as Japanese language.

Miller et. al [20] proposed a method to find loanwords in mono-lingual texts by using an automated frame work exploiting the phonological and phonotactic clues. The method depends on the use of Support Vector Machines, Markov models, and recurrent neural networks. The results show that using phonological and phonotactic clues derived from monolingual language are inadequate to identify the loanwords.

Aboh et.al [21] used edit-distance measures and a sound-class based method to measure the phonological similarity. They showed that this measuring can be neglected because words coming from the source language to a borrowing language are subject to some changes to fit the phonology of the borrowing language.

All these and other works, according to our knowledge, did not deal with Turkish loanwords from Arabic origin. Also, graph mining was not used for loanwords detection task.

Extracting Loanwords Method

Our proposed methodology is used for answering two questions: (i) Is a given Turkish word a loanword from Arabic language? And (ii) what are the Turkish loanwords for a set of character and a specific range of word length? For answering these questions, two stages should be done in our methodology. These stages are (i) transformation of Turkish and Arabic dictionaries into reference words, (ii) identification of Turkish loanwords from Arabic origin.

Transformation of Turkish and Arabic Dictionaries into Reference Words

Basically, when dealing with the Turkish and Arabic languages, there are two main problems have been faced; the two languages have different scripts, and Arabic language is rich in synonyms which produce dictionary gap between these languages [22].

For solving Arabic richness problem, compared to Turkish language, two dictionaries are taken; Turkish-English and Arabic-English dictionaries. English language is the intermediate language because it is very simple compared to Arabic language and it has many to one mapping where many Arabic synonyms have one meaning in English language. Also, the same situation for Turkish language with English language therefore the dictionary gap will be solved without using semantic relation of the words.

The other main problem is the difference between Turkish and Arabic languages in scripts. Therefore, a reference scripts should be used. This is can be done by converting the Turkish words into a reference language (English) depending on the pronunciation. In this case, each character in Turkish word is converted into its equivalent character in English ignoring the vowel letters to avoid some problems related to the accent differences and also to get correct matching results as shown in (Fig. 1). On the other hand, Arabic words in another data base need to be converted into English scripts in the same way with Turkish words as illustrated in (Fig. 2). This transformation will unify the scripts of equivalent words in Turkish and Arabic languages based on their pronunciation. Some letters have more than one pronunciation therefore many words will have more than one reference word, i.e., each word of this type will have a list of reference words based on different candidate pronunciations of this word.

Algorithm 1 shows the transformation of Turkish-English and Arabic-English dictionaries into reference words. The input to the algorithm is two dictionaries in the form <Turkish word, meaning in English> and <Arabic word, meaning in English> and the output will be in the forms < Turkish word, meaning in English and **list of reference words**> and <Arabic word, meaning in English and **list of reference words** >. It is easy to see that any Turkish word, that has the same reference word and meaning of an Arabic word, it is a loanword.

Turkish word	Meaning in English	Reference words list ignoring vowel letters
Gazap	Anger	Gzp
Farah	Happiness	Frh
Takvim	Calender	Tkvm
Hakikat	Truth	Hkkt
Servet	Fortune	srvt
Cumhuriyet	Republic	jmhrt

Figure 1. Turkish Dictionary (real words with their meaning and reference words)

Arabic word	Meaning in English	Possible pronunciations	Reference words list ignoring vowel letters
غضب	Anger	Gazap	Gzp
فرح	Happiness	Farah	Frh
تقويم	Calendar	Takvim, Takuim	Tkvm, tkm
حقيقة	Truth	Hakike, hakikat, hakikad*	Hkk, Hkkt, Hkkd
ثروة	Fortune	Serue, seruet, serued, serve, Servet, served	Sr, srt ,srd, srv, srvt, srvd
جمهورية	Republic	Cumhuriye, Cumhuriyet, Cumhuriyed, Cumhvriye, Cumhvriyet, Cumhvriyed	Jmhr, jmhrt, jmhrd, jmhvr, jmhvrt, jmhvrd

* In some cases the character "t" is pronounced as "d" or therefore this pronunciation is taken in account.

Figure 2. Arabic Dictionary (real words with their meaning and reference words)

Identification of Turkish Loanwords from Arabic Origin:

The identification of Turkish loanword from Arabic origin is based on the equivalence of the reference words and their meaning. But in some situation, the loanwords suffered from small and large modifications. Therefore, there is not exact matching in reference words or meaning for such words. For this reason, an efficient technique should be used for detection or identification. It is clear that Turkish loanwords from Arabic language fall into the following types:

1. Loanwords have same pronunciation and meaning of the Arabic origin.
2. Loanwords have same pronunciation but they modified in meaning compared to the Arabic origin.
3. Loanwords have modification in pronunciation but they have same meaning of the Arabic origin.
4. Loanwords have modification in pronunciation and meaning compared to Arabic origin.

The modification in meaning or pronunciation of loanwords with Arabic origin may be large (complete) or small modification. Because two dictionaries are used, the modification in meaning can be detected if the modification is done using synonyms. This is the main reason for using two dictionaries. The small modification in pronunciation can be detected easily using graph theory as will be shown.

In the proposed system, the main objectives (tasks) are the identification of Turkish loanwords for (i) a specific word or (ii) for all words in range of specific length and consists of limited subset of characters. The first task is subtask of the second task therefore the second subtask will be explained in this section.

According to these objectives, any set of characters can be taken that construct real Turkish words in range of length n. It is done by selecting all the words, in the Turkish dictionary, that have length of or less than n as real words. A binary search is used for this purpose because the data base is ascendingly sorted, and the time complexity of binary search is O (log (n)) compared with sequential search which is O (n). These real words are compared with the character set for neglecting the words that have any character outside this character set.

Algorithm 1:-Transformation of words into reference language

Input: **TWR**: Raw Turkish-English Words dataset;
AWR: Raw Arabic-English words dataset
Output: TW: Turkish Words dataset with a list of reference words to each word
AW: Arabic Words dataset with a list of reference words to each word
Step1: initially TW and AW are empty dictionaries:
Step2: for each word w in **TWR**:

- **References= []**
- Delete vowel character from w:
- If w has more than one spell combination, record all of them to **References** list
- Add w with its meaning and **References** list to TW dictionary.

Step2: for each word w in **AWR**:

- **References= []**
- Delete vowel character from w:
- If w has more than one spell combination, record all of them to **References** list
- Add w with its meaning and **References** list to AW dictionary.

Step3: Return **TW** & **AW** dictionaries.

Some of these words are adopted from Arabic language. Therefore, the next step is to find Turkish words coming from Arabic origins. This is can be done by using graph mining technique. It starts by retrieving reference words list of a given Turkish word and then this reference words list will be used for construction directed graph that represent all the combinations of reference words as shown in (Fig. 3). Any Arabic word that has a reference word as a subgraph of this graph, it is candidate to be origin of the selected Turkish word. Any matching in the meaning for the candidate origin with the meaning of the selected Turkish word, it will be the Arabic origin of this Turkish word. Any matching in the meaning for the candidate origin with the meaning of the selected Turkish word means this Turkish word is from Arabic origin. Algorithm 2 shows the identification of Turkish loanwords from Arabic origin.

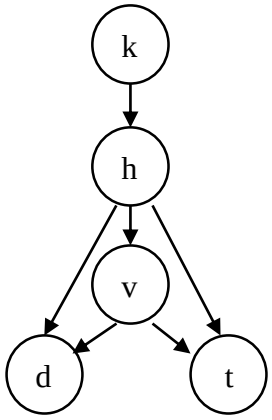

Figure 3. Directed graph for the reference words list of the possible pronunciations [kahue, kahve, kahuet, kahvet, kahued, kahved] of the word "قهوة" (coffee).

Algorithm 2:-Turkish-Arabic Loanwords Detection

Input: **ST**: Set of Turkish characters,
ML: the maximum length of generated substrings,
TW: Turkish-English words dataset with list of references for each Turkish word;
AW: Arabic-English words dataset with list of references for each Arabic word.
Output: Loanwords; list of pairs of Turkish word & its origin from Arabic language

Step1: Initially Loanwords = [ ]; RealWords=[ ]

Step2: Build undirected graph UG that represents each character in ST as node

Step3: for each word w in TW:

If length (w) <=ML and w is subgraph of UG:

- Add w to RealWords
- Also record its meaning & reference words list.

Step4: For each word w in RealWords do:

A- Produce graph that represent all reference words of the word w.

B-Search the equivalent reference word in AW that is subgraph of this graph.//the length of this subgraph should be same or less by one with the length of original graph.

C- If any of two reference words have the same meaning, add this word with the equivalent Arabic origin into Loanwords list.

Step5: Return Loanwords list.

Results and Discussion:

The proposed algorithm was tested using 1256 manually classified words where half of them (628) are loanwords and the other are non-loanwords. Also variant string lengths were used as input to the system and the results were checked manually.
This section has three parts; dataset and experimental setting, transformation of datasets to reference words, and loanwords detection.

Dataset and Experimental Setting

The proposed algorithm was implemented using VB.net programing language on laptop of Intel core i7 CPU, 8 G RAM, and windows 10 operating system. Two dictionaries were used in this work. The first one is Turkish-English dictionary in the form <Turkish Entry, English meaning (list)>. The Turkish entry is lemma of Turkish word. The second dictionary is Arabic-English dictionary in the form <Arabic Entry, English meaning (list)> where the

entries are in the form of lemmas of Arabic words. For testing the system 1256 manually classified words are taken. Half of these words are loanwords.

Transformation of Datasets into Reference Words

Arabic and Turkish words should be transformed into a reference language (English Language) as explained in the previous sections and shown in (Figs. 1, 2). The reference language is used because the two languages (Arabic and Turkish) have different styles in writing. This leads to impossibility of loanword matching in the two languages. The output of this step is two dictionaries in the form <Turkish Entry, English meaning (list), reference words (list)> and <Arabic Entry, English meaning (list), reference words (list)>. Each Turkish and Arabic word has list of references in the reference language. Reference words represent the different spelling of the Turkish entry or Arabic entry.

Loanwords Detection

Firstly, the system was tested using 1256 manually classified words. Only nine words of the loanwords are not recognized by the system as loanwords while all the non-loanwords are identified. This is means that the precision, recall and f-measure are 1, 0.98 and 0.99 respectively.

Then the system was tested for extracting all the loanwords that range in a specific length with a predefined character set. Many tests were done for different character sets and different word lengths. The entire outputs were checked manually. Two samples of these tests are shown in (Figs. 4, 5). The first step is entering any set of characters and the maximum length of the generated substrings. Then the button "Generate Substrings" needs to be pressed. This will extract all real words in this range by traversing the graph which represents the entered set of characters.

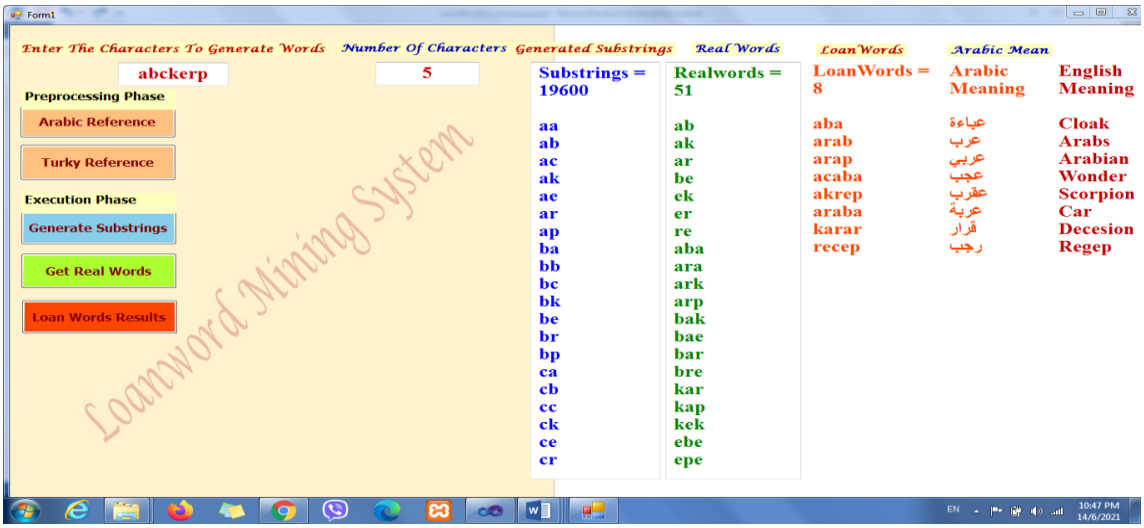

Figure 4. First sample of loanwords for the character set {a, b, c, k, e, r, p} with n=5

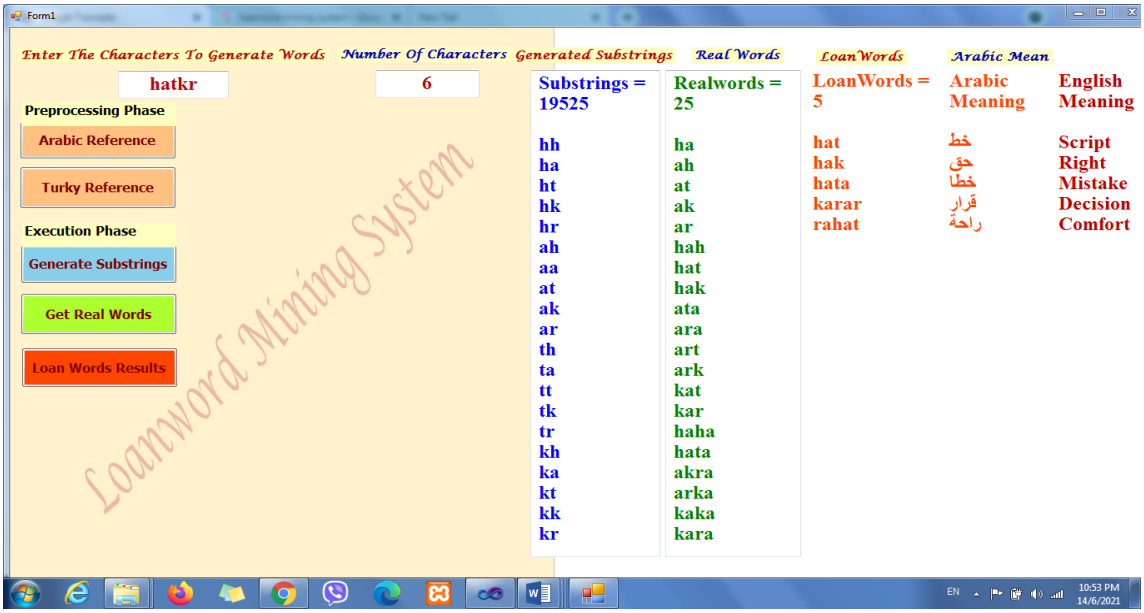

Figure 5. Second sample of loanwords for the character set {h, a, t, k, r} with n=6

Our methodology has few limits which can be summarized by: (i) it does not work directly on Turkish and Arabic encoding but it works on a reference language to bypass problems of encoding, and (ii) many comparisons will be done if a character set is taken with a specific range of word length.

Conclusion:

Detection of loanwords is an important task for linguists and history scientists. It is complicated and time consuming task. This paper proposed a novel methodology for detecting such words automatically between Turkish and Arabic words. The proposed system detects almost all loanwords in Turkish language from Arabic origin. This work faced many challenges such as the difference in writing scripts of the used two languages. Also the used two languages, Turkish and Arabic, are highly different in the morphology and word construction level. Therefore; the functions of calculating the distances in order to find the matching among terms or words are very difficult. Consequently, a reference language was employed to represent the terms or words in the two target languages.

The main motivation of the proposed work is to find the loanwords in Turkish language from Arabic language where the main contributions are (i) using graph mining techniques for detection loanwords, and (ii) dealing with Turkish and Arabic language, it was used for the first time, and (iii) building an efficient system for detecting the loanwords.

Furthermore, data mining has a good ability to deal with change in pronunciation, especially when words move among different languages through processing of vowels. Finally, it can be used with different languages despite the lack of full knowledge of the components of these languages. Manually testing of 1256 loanwords proved that our system works with high precision.

The suggestion for the future works are; (i) applying the proposed system on other languages, (ii) combining different methodologies for improving efficiency of the system, and using deep learning for detecting the small relations of words.

Authors' declaration:

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Baghdad.

Authors' contributions statement:

A. K. J. conceived of the presented idea. A. K. J. and M. J. H. developed the algorithm and programming execution. A. K. J. and A. H. A. verified the analytical methods. A. H. A. accomplished the investigation and data curation. All authors discussed the results and contributed to the final manuscript.

References:

1. Fernandes E, Holanda M, Victorino M, Borges V, Carvalho R, Van Erven G. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. J Bus Res.

2019 Jan 1;94: 335-43. https://doi.org/10.1016/j.jbusres.2018.02.012
2. Miao, Cai An, Tan Shi. Application of Data Mining Techniques on Tourist Expenses in Malaysia. Baghdad Sci.J. 2021. 18; 1: 737-745.
3. Kantardzic M. Data mining: concepts, models, methods, and algorithms. 2nd edition John Wiley & Sons; 2011 Aug 16.
4. Bacciu D, Micheli A, Podda M. Edge-based sequential graph generation with recurrent neural networks. Neurocomputing. 2020 Nov 27; 416: 177-89. https://doi.org/10.1016/j.neucom.2019.11.112.
5. Yuan W, He K, Guan D, Zhou L, Li C. Graph kernel based link prediction for signed social networks. Inf Fusion. 2019 Mar 1; 46: 1-0. https://doi.org/10.1016/j.inffus.2018.04.004
6. Priya A, Sinha K, Darshani MP, Sahana SK. A novel multimedia encryption and decryption technique using binary tree traversal. InProceeding of the Second International Conference on Microelectronics, Computing & Communication Systems (MCCS 2017) Springer, Singapore. 2019: 163-178. https://doi.org/10.1007/978-981-10-8234-4_15
7. Fournier-Viger P, He G, Cheng C, Li J, Zhou M, Lin JC, et al. A survey of pattern mining in dynamic graphs. Wiley Interdisciplinary Reviews: KDD. 2020 Nov;10(6): e1372. https://doi.org/10.1002/widm.1372
8. Yan X, Han J, Discovery of frequent substructures.In: Cook DJ, Holder LB, Mining graph data. John Wiley & Sons; 2006 Dec 18. p 99-113.
9. Bahumaid S. Lexical borrowing: The case of English loanwords in Hadhrami Arabic. IJLL. 2015 Dec;2; 6:13-24.
10. Pulcini V, Furiassi C, Rodríguez González F. The lexical influence of English on European languages. The anglicization of European lexis. John Benjamins Publishing Company. 2012. p.1-24. https://doi.org/10.1075/z.174.03pul
11. Hock HH, Joseph BD. Language history, language change, and language relationship: An introduction to historical and comparative linguistics.. Walter de Gruyter GmbH & Co KG; 3rd ed. 2019 Sep 2. https://doi.org/10.1515/9783110613285.
12. Metz HC, From autonomy to occupation: Ismail, Taqfiq, and the Urabi revolt. Egypt H. A Country Study. In GPO for the Library of Congress: Washington, DC, USA 1990. P 35-37.
13. Salman YM, Mansour MS. English Loanwords in Iraqi Arabic with Reference to Computer, Internet and Mobile Phone Jargon.'. Cihan Univ. Erbil Scij. 2017; 1: 271-94. https://doi.org/10.24086/cuesj.v1n1a14
14. Peperkamp S, Dupoux E. Loanword adaptations: Three problems for phonology and a psycholinguistic solution. Unpublished manuscript, Laboratoire de Sciences Cognitives et Pscyholinguistique, Paris & Universite de Paris. 2001; 8: 1-2.
15. Peperkamp S A. psycholinguistic theory of loanword adaptations. Annual Meeting of the Berkeley Linguistics Society 2004 Jun 25; 30; 1: 341-352.
16. Rao CS. The Significance of the Words Borrowed Into English Language. J res scholars prof Engl lang teach. 2018;2(6): 1-9.
17. Farazandeh-pour F, Kord Zafaranlu Kambuziya A. German Loanwords Adaptation in Persian: Optimality Approach. Int j humanit. 2013 Oct 10; 20; 4:23-40.
18. Mi C, Yang Y, Zhou X, Wang L, Li X, Jiang T. Recurrent neural network based loanwords identification in Uyghur. In Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation: Oral Papers. Paclic 30 Proceedings; 2016 Oct :209-217.
19. Koo H. An unsupervised method for identifying loanwords in Korean. Lang. Resour. Eval. 2015 Jun;49; 2: 355-73. https://doi.org/10.1007/s10579-015-9296-5
20. Miller JE, Tresoldi T, Zariquiey R, Beltrán Castañón CA, Morozova N, List JM. Using lexical language models to detect borrowings in monolingual wordlists. Plos one. 2020 Dec 9; 15. 12: e0242709. https://doi.org/10.1371/journal.pone.0242709.
21. Zhang L, Manni F, Fabri R, Nerbonnei J. Detecting loan words computationally. Variation Rolls the Dice. A Worldwide Collage in Honour of Salikoko S. Mufwene. John Benjamins Publishing Company. 2021 Oct; 15. https://doi.org/10.1075/coll.59.11zha.
22- Farghaly A, Shaalan K. Arabic natural language processing: Challenges and solutions. ACM Transactions on Asian Language Information Processing. 2009; 8(4): 1-22.

Open Access **Baghdad Science Journal** **P-ISSN: 2078-8665**
Published Online First: May 2022 **2022, 19(6): 1369-1377** **E-ISSN: 2411-7986**

استخدام طريقة تنقيب المخططات في تحليل الكلمات التركية المستعارة والمشتقة من اللغة العربية

عبود خريبط جاسم[1] منعم جبار حمزة[1] احمد حسين عليوي[2]

[1] قسم علوم الحاسوب، كلية العلوم للبنات، جامعة بغداد، بغداد، العراق.
[2] قسم علوم الحاسوب كلية الحاسوب والرياضيات جامعة الكوفة النجف العراق

الخلاصة:

الكلمات المستعارة هي الكلمات التي يتم نقلها من لغة إلى أخرى وتصبح جزءًا أساسيًا من لغة الاستعارة. جاءت الكلمات المستعارة من لغة المصدر إلى لغة المستلم لأسباب عديدة. على سبيل المثال لا الحصر الغزوات أو المهن أو التجارة. ان ايجاد هذه الكلمات المستعارة بين اللغات عملية صعبة ومعقدة نظرا لانه لايوجد معايير ثابتة لتحويل الكلمات بين اللغات وبالتالي تكون الدقة قليلة. في هذا البحث تم تحسين دقة ايجاد الكلمات التركية المستعارة من اللغة العربية. وكذلك سوف يساهم هذا البحث بايجاد كل الكلمات المستعارة باستخدام اي مجموعة من الحرووف سواءا كانت مرتبة او غير مرتبة ابجديا. عالج هذا البحث مشكلة التشويه في النطق وقام بايجاد الحلول للحروف المفقودة في اللغة التركية والموجودة في اللغة العربية. تقدم هذه الورقة طريقة مقترحة لتحديد الكلمات التركية المستعارة من اللغة العربية اعتمادًا على تقنيات التنقيب في المخططات والتي استخدمت لاول مرة لهذا الغرض. فقد تم حل مشاكل الاختلاف في الحروف بين اللغتين باستخدام لغة مرجعية وهي اللغة الانكليزية لتوحيد نمط وشكل الحروف. لقد تم اختبار هذا النظام المقترح باستخدام 1256 كلمة. النتائج التي تم الحصول عليها تبين ان الدقة في تحديد الكلمات المستعارة كانت 0,99 والتي تعتبر قيمة عالية جدا. كل هذه المساهمات تؤدي إلى تقليل الوقت والجهد لتحديد الكلمات المستعارة بطريقة فعالة ودقيقة. كما أن الباحث لا يحتاج إلى معرفة باللغة المستعيرة واللغة المأخوذ منها. علاوة على ذلك ، يمكن تعميم هذه الطريقة على أي لغتين باستخدام نفس الخطوات المتبعة في الحصول على الكلمات المستعارة التركية من العربية.

الكلمات المفتاحية: اللغة العربية ، التنقيب عن البيانات ، التنقيب في المخطط ، الكلمات المستعارة ، اللغة التركية