



Review Article

Processing of Semantic Ambiguity Based on Words Ontology

A Abou Shousha¹, Samir Hamada^{2*}, Salwa Hamada³ and Mohammad Alshibli²

¹Al Forsan International School, Riyadh, Saudi Arabia

²Farmingdale State College, Farmingdale, New York, USA

³Department of Informatics, Electronic Research Institute, Egypt

Received: 07 September, 2020

Accepted: 04 November, 2020

Published: 05 November, 2020

*Corresponding author: Dr. Samir Hamada, Farmingdale State College, Farmingdale, New York, USA, E-mail: hamadas@farmingdale.edu

Keywords: Semantic; Polysemy; Ontology ambiguity

<https://www.peertechz.com>



Abstract

This research provides an automatic treatment for the phenomenon of semantic Polysemy based on the ontology (meaning, accompaniment, and translation) of words and its effect on the applications of automatic processing of the Arabic language.

Polysemy may have negative effect on these applications. This research works on finding the solutions that helps improving the level of the automatic treatment of the Arabic language at all levels through applying a descriptive analysis for a sample of 50 words with their different derivatives. This descriptive analysis suggests to the user the most probable meaning of a suitable context through analyzing the context and relying on verbal structure and collocations and then determine the morphological analysis and appropriate translation of the English language as well as the availability of some statistical data such as the probability that this meaning is the appropriate sense of other meanings, the number of possible meanings and other statistical data.

This study clarifies that the main reason of the semantic polysemy phenomenon in the used texts is the absence of diacritics.

It was found that applying the proposed methodology in this paper on the ontological corpus helps to identify the exact sentences intended meaning by more than 80% accuracy.

As a result, this automatic processing will give the benefit to the searching sites like Google, and also in facilitating the teaching of significance; especially in the field of metaphor "فراغتمسالا" in the Arabic language for non-native Arabic speakers. Moreover, it will help in analyzing the Arabic texts and translating and many other applications of Arabic language computing.

Introduction

One of the important axes that underlie the process of automatic processing of Arabic written texts is the connotation of words.

The connotation of words is one of the most important problems that happens in the automatic processing of the Arabic language. Since there can be no deep processing of Arabic texts without sufficient information on the meaning of the words that make up these texts.

Automated semantic processing of Arabic language requires information on various aspects of the language. This information includes information about words, various

indications, what may be used with the word, what is not permissible, words that are approachable, and what is related to the accuracy of the meaning of the word.

For most automatic systems, the meaning of the word is the major problem. Since the single word has more than one meaning, the sensitivity of the context in determining the meaning of the word, the difference of the connotation in different cultures, those factors together make the automatic treatment of the connotation involve paradoxes that can't be described correctly.

Related work

Many users in the Arab countries either native Arabic



speakers or non-native speakers use the web with Arabic contexts. Many problems come to the users when dealing with Arabic texts and contexts. Al-Zoghby, et al. [1] have conducted a survey to figure out the problem, its causes, and an attempts to give solutions. The survey included the semantic web applications regarding the Arabic language in the field of ontology, Holy Quran and Islamic knowledge, and Arabic semantic search engines [1].

In the aim of solving semantic ambiguity to improve semantic web based ontology matching, Gracia, et al. [2] have introduced techniques from web sense disambiguation through testing two techniques; which exploit the ontological context of the matched and anchor terms, and the information provided by WordNet, can be used to filter out mappings resulting from the incorrect anchoring of ambiguous [2].

Hossam Ishkewy, et al. [3] have introduce a lexical ontology for the Arabic language. This lexical ontology gathers Arabic words into sets of synonym's called synsets, and records a number of relationships between words such as synonym, antonym, hypernym, hyponym, meronym, holonym and association relations. The ontology contains 26,195 words organized in 13,328 synsets. It has been developed and contrasted against AWN which is the most common available Arabic lexical ontology [3-7].

The importance of this study

We are interested in the semantic level of the language that is responsible for giving the sentence the cognitive and logical value necessary to complete the process of understanding.

This study is considered of great importance for many reasons such as; it is concerned with the semantic level for the language that gives the sentence the cognitive and logical value that is necessary to the process of understanding, it also studies the phenomenon of polysemy and its positive and negative effects on the automatic treatment for the Arabic language and the automatic understanding and the technologies that can be added to support the language and Arab culture.

Moreover, our study works on investing the semantic searching in the applied field such as the automatic translation, and providing automatic dictionaries for vocabulary, structures, and contexts.

The purpose of the study

This study aims at studying the phenomenon of semantic polysemy and its effect on the automatic treatment through several axes such as; the effect of semantic polysemy in the construction of the automatic dictionary, the effect of semantic polysemy in the automatic analysis of texts, and the effect of semantic polysemy in automatic translation. All that can be studied through the applied study of some of the words of this phenomenon.

Study methodology

The study is based on the analysis of this phenomenon and the study of its effect on the automatic treatment of the Arabic

language on the analytical descriptive approach, while not neglecting the other research methods such as comparative, historical and statistical when required.

Semantic multiplicity and automatic analysis of Arabic texts

Information retrieval, lexical analysis to study Automated analysis of Arabic texts means a range of linguistic, statistical and machine learning techniques that contribute in the extraction of information content. Text analysis thus includes the commonality of words, language pattern recognition, information extraction, data mining techniques including the analysis of relations between word. The purpose of all these tasks is to transfer Arabic text to data that can be analyzed with Arabic Language application.

Analysis levels

Statistical level using concordance 3.3 available at

<http://www.concordancesoftware.co.uk/>

In order to reach statistics that highlight the frequency of each of the words of the sample words in the texts of the code in question, and then examine the contexts in which these words are received in order to identify the frequency of each meaning of those words and identification.

Morphological analysis: Which gives us the (morph, root,...). That was done using a Morphological Analyzer tool. AraFlex because it has an important feature which connect results with their meaning. This program is available at <http://lexanalysis.com/araflex/input.html>.

Machine translation:

- Choosing some sentences (contexts) in which the meaning of this expression appears multiple through the code in the field of research.
- Entering these contexts into automatic translation through the Google site to translate them into English.
- Analyzing the site's output from translating these contexts and researching the accuracy of the English interviews chosen by the translator for each context to reveal the effect of multiple meanings of those words on the translation process.

The main results of the study

After studying the phenomenon of semantic polysemy and its effect on the automatic processing of the Arabic language and after conducting the applied study on contemporary Arabic. The important results summarized as follows:

First:

The study of the phenomenon of semantic polysemy is an influential addition to the automatic processing of the Arabic language. The importance of this study can be addressed in the following major points:

It helps in creating the desired automatic dictionary as it enriches the linguistic material of the dictionary. Also, it helps in avoiding the presence of errors in the examples and the confusion in meanings.

- It makes great benefit from the existence of compositions and word collocations in defining the required meaning and distinguishing between the multiple meanings of a single word and their types in terms of whether the meaning is real, metaphorical, functional or terminology. Also, defining the morphological, grammatical and semantic linguistic features of each meaning of the word.
- It provides the assistance in the automatic analysis of the Arabic text, as it helps in defining the intended meaning through the context, depending on the structures in which the word or linguistic accompaniment falls within the context in which the word is expressed and by meaning helps to determine the proper morphological analysis of the word in context.

Automatic translators assist in determining the proper translation of the term by determining the meaning intended through the context, based on the development of the databases of the automation and its supply of the possible meanings of the term and the appropriate translation for each of these meanings.

Second:

For example, the word “display” when examined through contexts or when it is revealed in the modern Arabic dictionary found that it is more likely than the vocalization “اليفشيتل” of it (ضُرْع / اضرْع / اضرْع / اضرْع / اضرْع), and since the word (ضُرْع), which is the subject of the search is more than likely, the absence of the vocalization of this word in the text opened the space for the existence of other possibilities already mentioned.

Third:

The study revealed the effect of the phenomenon of semantic polysemy in the automatic processing of the Arabic Language Lexicon and dictionary through three main axes:

1. By examining the words of the research sample through several old and contemporary Arabic lexicon and dictionary. Then examining the same words in the modern Arabic lexicon as a model of the Arabic lexicon identifying word's meaning and order.

The effect of the phenomenon of semantic polysemy in the Arabic Lexicography processing in general and the automatic dictionary, in particular, has emerged as follows:

- The phenomenon of semantic polysemy led to multiple results of the word and the multiple meanings of the single result, which enriches the linguistic material lexicon in case of attention to this phenomenon and also requires the makers of the dictionary the need to distinguish between the results and some of them as

well as the distinction between the meanings of the same result or the same word, The search for the linguistic characteristics of each word and each result and each meaning, which is a burden on the manufacturers of the dictionary and helps to get out if it is achieved in the best possible picture and can be developed.

- The phenomenon of multilingualism has moved the function of the lexicon from mere presentation of the meaning of the word to the expression of the meaning of words through contexts (i.e. from the reality of language use) by relying on a linguistic corpus as a reliable language in the lexicography.
- The phenomenon of multi-semantic has helped to determine the method of mentioning the meanings of the word or not, as well as in the order of the meanings of one word, where it revealed the existence of a disorder in the meanings of the meanings and order in the dictionaries of paper as evidenced by the tests, which helped the modern Arabic dictionary to determine the method of mentioning the meanings contained In the linguistic corpus on which they depend, and the order of the meanings of one word is statistically significant.
- The phenomenon of semantic polysemy revealed the importance of the fact that this lexicon can be updated both on the basis of the material on which it relies and on the lexical definition of the words. On the one hand, this is done by constantly updating the texts of the corpus on which it is based. Or whenever necessary. As for the lexical definition of words, it is necessary to update the lexicon subject, taking into account the development or use of words as a term in science.
- The phenomenon of semantic polysemy revealed some weaknesses in building the automatic Arabic dictionary, the most important of which are:
- Sometimes the lexicon does not address the existence of the phenomenon of semantic polysemy in some words such as phenomenon, street, live where the dictionary only mentions one meaning only.
- Examples were appropriate in most of the examples of the meanings contained in them. However, the failure to take into account the phenomenon of semantic polysemy sometimes led to the lexicon in errors at the semantic level, sometimes at the morphological level, at other times.
- The examination of the words (living, alive), (accurate, precise) showed the absence of the approach of dealing with the feminine word when it is possible to have other non-masculine meanings. In the first case, the lexicon only mentioned one meaning of a living word, and did not indicate that it could be feminine, in the second case, the lexicon transferred the meanings of a precise word and inserted it as it is under a precise word after mentioning the meaning of the word precise.

2. The Effect of Semantic Phenomenon in Automated Analysis of Arabic Text

The phenomenon of semantic ambiguity has influenced the automatic analysis of the Arabic text through the automatic indexing of texts and the morphological analysis:

- The phenomenon of semantic ambiguity has affected the indexing mechanism negatively and led to overlapping results. It may have been due to the development of the indexer for a foreign language and the lack of information about the nature of the Arabic word, as it deals with the form only, which is incompatible with the nature of the Arabic language. In particular, with the absence of the composition of the contemporary Arabic text.
- The phenomenon of semantic ambiguity revealed the importance of the development of automatic indexing so that its role in addition to providing statistics on the percentage of the receipt of the word in the text, in providing statistics on the proportion of the receipt of every sense of the word by examining the contexts collected by the indexer automated as well as limiting the words associated with each sense of The meanings of the word, as well as monitoring the structures and idioms that enter the word and the proportion of each of them.
- The phenomenon of semantic ambiguity revealed the absence of a semantic connection to morphological analysis, which leads to the multiplicity of the results of polysemy of the word in many morphological analysts such as the morphological analyst Hebron, which is unacceptable, especially that the process of morphological analysis in these programs is through the text And the consequences of the existence of the results of a wrong analysis where the word in the context can only tolerate a single exchange analysis.

The phenomenon of ambiguity revealed the inability of the morphological analyst to distinguish the results in some cases where there are multiple meanings of the term with the agreement of morphological analysis as expressed in many words such as the term (limit حد) as shown in the following Table 1.

Where the third and fourth results were as follows:

Limit: A name meaning stop, has been returned the root (حد).

Limit: A name in the sense of range or end, and has been returned to the root (حد).

Where the results agreed in the morphological analysis and differed in meaning and were supposed to be a single result and recall all meanings.

The results of the AraFlex morphological analysis process can be used to distinguish between Homonymy “كثرة المعاني”

3	Hadd حَدَّ	Hadd حَدَّ	H d ح د	Hadd noun stop/halt
4	Hadd حَدَّ	Hadd حَدَّ	H d ح د	Hadd noun extent/limit

Table 1: Morphological analysis of the word حد.

” and Polysemy “من عمل اددعت” “ Multiplicity of meaning” cases, taking into account that both are included in the semantics section.

Fourth:

The study revealed the effect of the phenomenon of semantic ambiguity in the automatic processing of the Arabic language through three main axes:

The study did not limit the effect of semantic ambiguity in the automatic processing of the Arabic language. The researchers tried to benefit from the positive effects and work to limit the negative effects in an attempt to lay the foundations for dealing with the phenomenon of semantic ambiguity during the automatic processing of the Arabic language. A database of semantic ambiguity as well as a conceptualization of a program based on this proposed rule to deal with the phenomenon of semantic ambiguity during the automatic processing of the Arabic language.

Finally, it can be said that the phenomenon of semantic ambiguity is a positive factor when attention to and taking into account in the processes of automated processing of the Arabic language and is a negative factor and cause errors when not paying attention to them and not to take them into account.

3. The effect of semantic ambiguity in machine translation from Arabic:

Through the results of tracking the research sample across multiple contexts in search of its impact in the translation process we conclude the following:

- The phenomenon of semantic multiplicity in the occurrence of the translator in many of the errors revealed by the analysis and these errors resulted from the translator or automated translation system did not realize according to its databases that these words are multi- ambiguous sense or realized that the words are multi- ambiguous sense, but the errors occurred in The stage of choosing the appropriate answer, which necessitates the development of the bilateral dictionary on which the program depends on appropriate interviews for each sense of the word multi- ambiguous sense.
- The phenomenon of semantic ambiguity revealed that the site did not rely on language translations (linguistic



clues) sufficiently, which was able to resolve the matter and guide it to the appropriate choice in many contexts,

- The effect of semantic phenomena in the process of automatic translation, although it seems negative, but it can be used as it forces us to search for new mechanisms added to the translation systems to be able to distinguish between each sense of the word sense and then work to provide The databases on which the system depends, with a large amount of data, which help to distinguish between these meanings.
- It is possible to rely on the phenomenon of semantic polysemy as a measure of the quality of translation or not since it represents an advanced level of ambiguity. If the translation system is able to distinguish the meanings of the word multi-significance, it is undoubtedly a sign of the quality of translation, otherwise it can be measured at the level of this translation, It is possible to develop a statistical scale for this by counting the number of words in the translated text and then counting the percentage of meanings that the system could distinguish, and the percentage of meanings that it could not distinguish.

In the light of the study of the phenomenon of semantic ambiguity, we can see the future of machine translation and what it should be in the future where the process of automatic translation can be improved through the following:

- Do not deal with Arabic words as mere words that have direct interviews in other languages or what is known as literal translation.
- Develop databases for translation programs and sites by building a database of semantic expressions, with verbal expressions at different levels and appropriate translation for each case by carefully collating and verifying each word for use in translation.
- Make lists of the structures in which the words are entered, whether these words are multi-literal or not, and translate these compositions (such as effect, live footage, crown) into other structures that directly affect the translation process and cause errors if Translated into literal translation, as well as idiomatic expressions whose meaning is not explained by the mere interpretation of each of their words, which cannot be literally translated from one language to another, such as “hit the wall, hit him with an eye”.
- Preparation of lists of terms used in terms of words such as (Prophetic Hadith, multiplication table, organism, liquid blood, conditional cessation) to other terms that are translated professionally in the sciences in which they are used.

Our study

Our study proposed database to deal with the phenomenon of semantic multiplicity.

Limitations and research sample

The research in this study is limited to a random sample of the words that fall under the section of semantic multiplicity, to be representative of this phenomenon, and bear the same characteristics.

This method is preferred since it provide us with information that is as accurate as comprehensive inventory information, the difficulty to limit all semantic variations in one study, and the time that is taken to deal with the samples is short compared with the time needed for the comprehensive inventory.

If the researchers are in the process of selecting the sample, they fully understand that there is a major condition that governs their ability to generalize their results to the rest of the words of this phenomenon, namely representation. The following conditions were taken into account in selecting the sample:

- The availability of all the characteristics and characteristics of semantic diversity in the sample, so that it is a miniature model of this phenomenon, and then we can say: What is true on this model is true to the rest of the words that fall under this phenomenon.
- The fact that the words of the sample have a contemporary character (i.e., the multiple meanings of these words are used in the current era, even in varying degrees, i.e., some are not negligible in terms of use) so that the results of the research have value in contemporary language use.
- Taking into account the diversity of nouns and verbs in simple words so that the results are more accurate and more applicable.

Test the validity of the research sample:

In order to verify the validity of the sample and ensure its validity to research, you have taken the following steps:

- Follow the words of the sample in a number of Arabic dictionaries.
- Track the words of the sample in a number of contexts of the blog field of research.

The results of this trace resulted in confirmation that each word of the sample belongs or enters under the semantic section.

The scope of the study

In order to make the study more useful and realistic, I chose a “Contemporary Arabic Dictionary” to be a field of applied study in this research. This thesis was compiled by the researcher: Al-Moataz Bellah Al-Saeed Taha in his Master’s thesis titled “ Dar Al-Uloom, Cairo University in 2008 under the supervision of Dr. Mohamed Hassan Abdel Aziz, Dr. Salwa Hamada, which is characterized by its contemporary language,



where the contemporary meanings emerge, which is what researchers see what serves the study of the issue of semantic multiplicity and highlighting its impact in the analysis, It is characterized by its large size The words of the blog are almost six and a half million words, which enrich the study in order to reach accurate results.

In the previous sections, we tried to uncover the effect of the phenomenon of semantic ambiguity in the automatic processing of the Arabic language, and then in the previous study, we examined the most important influences in the phenomenon of semantic multiplicity. In this light we try to present the idea of establishing a database to rely on in developing a program that can deal with words we will present the general mechanism for building the proposed database and then the mechanism of utilization in each of the axes of the proposed analysis, which works on the service of the semantic word.

The program used to build the proposed database

We build the database on Microsoft Access, one of the Microsoft Office programs, which is used to build databases. These rules can also be used to develop programs using software environments by linking them to databases.

First: General mechanism for building the proposed database:

Before we begin to outline the general mechanism for building the proposed database, we must define the objective or objectives of building this database, which we include in three basic paths related to the user:

- The first: is that the user wants lexical information about the pronunciation in general, such as the extent of its use, the number of its meanings, the proportion of the use of each meaning, the type of each meaning of the term.
- It is, therefore, necessary to provide this information in a schedule for the automatic dictionary.
- Second: the user wants to analyze the word through the context contained in a specific context in the text of the corpus, such as meaning through the context, appropriate translation of the term in this context.
- It is, therefore, necessary to provide this information in some tables, such as the meanings table, translation, and the exchange table.
- Third: The user wants the morphological analysis of a particular term either a general analysis or an appropriate analysis of the term in a given context.
- It is, therefore, necessary to provide this information in a table of morphological analysis so that it is linked to semantic analysis to avoid errors in the analysis process.

The general mechanism for constructing the proposed

database includes a set of tables which in turn contain a set of data that helps or provides the required information in any of the previous tracks. These tables are related to relationships that are determined by the type of data in each Table of Tables the following is a detailed breakdown of the proposed database.

Evaluation of Results

The results are compared with a number of models analyzed manually without using computer to determine the accuracy of the program through the number of the total of the results analyzed and the comparison of the number of valid results and the number of incorrect results and where the accuracy of the program can be measured by dividing the number of correct results on the total number of models that have been They were analyzed using the computer once and analyzed again without using the computer.

Conclusion

The research discussed the treatment of the phenomenon of semantic ambiguity automatically relying on a text corpus aimed at building an ontology of words that were selected sample of the research in advance a proposed model for the grammar of these words as well as a software application based on the proposed database, and this application is aimed at dealing with the ambiguous words during processing The first is the lexical definition, the second is the semantic analysis and the definition of meaning through the context, including automatic translation, and the third is the morphological analysis. The research presented a program to illustrate the possibility of automated processing. It also presented photographs of findings and compared them with the manual results of a linguist who produced about 80% accuracy.

The research has reached some important results, including The absence of the composition often from the contemporary Arabic text corpus led to an increase in the phenomenon of semantic ambiguity, and that the study of the phenomenon of semantic ambiguity is an effective addition to the field of improving the quality of automated processing systems for the Arabic language, The positive and negative effects of the phenomenon of semantic ambiguity in the automatic processing of the Arabic language.

The research proposals include: Providing a database and a software application to deal with multi-lingual words during the automatic processing of the Arabic language.

Thanks, and appreciation

The researchers are pleased to thank Sakhr Software team for agreeing to submit some of its Arabic lexicon materials to the research process to serve this research as a model of the contemporary Arabic automatic dictionary which helped in accomplishing this research. Al-Saeed Taha for agreeing to rely on a contemporary Arabic corpus, which the researcher collected in his study of the Master in 2008, a field of study in this work.



References

1. Zoghby A, Sharaf Eldin A, Hamza TT (2013) Arabic Semantic Web Applications – A Survey. Journal of Emerging Technologies in Web Intelligence 5. [Link: https://bit.ly/38c3ZEq](https://bit.ly/38c3ZEq)
2. Gracia J, Lopez V, d' Aquin M, Sabou M (2007) Solving semantic ambiguity to improve Semantic Web based ontology matching, Conference: 2nd Ontology Matching. Workshop (OM'07), at 6th International Semantic Web Conference (ISWC'07) At: Busan, Korea. [Link: https://bit.ly/2GquhqX](https://bit.ly/2GquhqX)
3. Ishkewy H, Hany HM, Farahat H (2014) Azhary: An Arabic Lexical Ontology, International Journal of Web & Semantic Technology (IJWesT) 5. [Link: https://bit.ly/3mVJShJ](https://bit.ly/3mVJShJ)
4. Keats, the Odes of 1819. [Link: https://bit.ly/32d9w9Z](https://bit.ly/32d9w9Z)
5. A Corpus of Historical Arabic Dictionary of Arabic Language Computerization (2010) Introduction to the Faculty of Dar al-Ulum, Cairo University, Ph.D., by Moataz Bellah Al-Saeed Taha, under the supervision of Mohamed Hassan Abdel Aziz, Mohsen Abdel Razek Rashwan.
6. (20 Feb. 2019). Arabic Morphological Analyzer. [Link: https://bit.ly/32bcbrjv](https://bit.ly/32bcbrjv)
7. Aziz MHA, El-Sayed Hamada S (2008) A Corpus that was compiled by the researcher Moataz Bellah Al-Saeed Taha in his master's thesis titled "The Corpus of Contemporary Arabic Dictionary of Computerized Language Processing.

Discover a bigger Impact and Visibility of your article publication with Peertechz Publications

Highlights

- ❖ Signatory publisher of ORCID
- ❖ Signatory Publisher of DORA (San Francisco Declaration on Research Assessment)
- ❖ Articles archived in worlds' renowned service providers such as Portico, CNKI, AGRIS, TDNet, Base (Bielefeld University Library), CrossRef, Scilit, J-Gate etc.
- ❖ Journals indexed in ICMJE, SHERPA/ROMEO, Google Scholar etc.
- ❖ OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)
- ❖ Dedicated Editorial Board for every journal
- ❖ Accurate and rapid peer-review process
- ❖ Increased citations of published articles through promotions
- ❖ Reduced timeline for article publication

Submit your articles and experience a new surge in publication services
(<https://www.peertechz.com/submission>).

Peertechz journals wishes everlasting success in your every endeavours.

Copyright: © 2020 Shousha AA, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Shousha AA, Hamada S, Hamada S, Alshibli M (2020) Processing of Semantic Ambiguity Based on Words Ontology. Trends Comput Sci Inf Technol 5(1): 070-076. DOI: <https://dx.doi.org/10.17352/tcsit.000027>