

Antonyms in Stem-Based Information Retrieval

Shapur Reza Berenjian

Faculty Member with Regional Information Center for Science and Technology (RiCeST)

Sh_berenjian@yahoo.com

Abstract: This paper introduces the outline of Persian antonyms in keyword-based storage and retrieval of information. In this regard, a definition is provided of antonyms and their typology. Subsequently, the factors creating antonymy in words are addressed in the article. Then we explain the Automatic Antonym Recall Software (AARS) in information retrieval (IR). When a keyword is entered into the software, it will provide the antonym for the given query. If the keyword or its antonym is not meant by the searcher, he may change or modify the keyword and send a new query. In case of alternate spellings, the searcher may alter the spelling and send a new query. This would help prevent the deletion of relevant information and documents and retrieval of non-relevant information. The present findings may be used in teaching Persian to language learners and beginners.

[S.H. Berenjian. **Antonyms in Stem-Based Information Retrieval**. *Stem Cell*. *Stem Cell* 2015;6(2):6-10]. (ISSN 1545-4570). <http://www.sciencepub.net>. 2

Key words: Antonyms, Persian language, Language teaching.

1. Introduction

An antonym is word with a meaning opposing another word (Dehkhoda, 13). Antonyms are words or groups having opposite meanings. It is more difficult to study antonyms than synonyms. Antonym elements have opposite meanings in the chain of speech. Palmer believes that two words are synonyms when they share the same antonym (Safavi, 1981).

In terms of interdependency, antonyms are not only different in terms of pronunciation and meaning but also have opposing meanings (Ahmadi Givi & Anvari, 2006). A word semantically opposite to another word is an antonym (Richard, 1993). Two words sharing all features except for one are called antonyms. This exception feature should be able to divide the phenomenon into two distinct situations. Thus, big and small are antonyms because both words implicate the size. However, each word denotes an opposing side of the size. One should note that 'man' and 'girl' are not antonyms because they vary in more than one feature (Falk, 1992).

Khodaparasti writes, "antonymy is the opposite of synonymy and is defined as follows":

1. Antonyms are cross words with opposing meanings.
2. The antonym of every word is a word conjuring up an opposing meaning in the mind.
3. An antonym is a semantically opposing word.

The above definitions are not incorrect; however, they are simplistic and may not be used in examining the relations among words. A more comprehensive and dependable definition of antonymy is provided below:

When any of the following relations hold true for two words of the same lexical category and some common semantic features, they are antonyms:

A) Interdependency: will exist between two words when the presence of one depends on the existence of the other. That is, it is irrational to expect the exclusive existence of one of them. The sets 'to sell # to buy' and 'husband # wife' are interdependent pairs because one needs to sell something before another could buy it, and no wife exists without husband.

B) Complementarity: is a relation where the acknowledgement of any word requires the rejection of the other. This is a bipolar and reciprocal relation. A clear example is the set of 'dead and alive'.

C) Opposition: is a relation where the acknowledgement of any element requires the negation of the other but not necessarily the other way round.

In other words, interdependency is the demonstrable contrast, complementarity is the negative contrast and opposition is the negative-positive contrast. In computational linguistics coding, these relations are written as follows:

1. Interdependency: $a \diamond b / ((a \diamond 0) \text{ and } (b \diamond 0))$
2. Complementarity: $a \diamond b / ((a \diamond 0) \text{ and } (b \diamond 1))$ or $((a \diamond 1) \text{ and } (b \diamond 0))$
3. Opposition: $a \diamond b / (((a \diamond 0) \text{ and } (b \diamond 0)) \text{ or } ((b \diamond 1)) \text{ or } ((a \diamond 1) \text{ and } (b \diamond 0)) \text{ or } (b \diamond 1))$

In the above relations, a and b represent the two elements of the antonym sets. The symbol \diamond represents opposition and negation, 0 is the lack of existence and 1 represents existence. 'And' and 'or' are logical operators and have the same meaning as and/or in spoken language, respectively (Khodaparasti, 1997).

Thus, antonyms are two words with all the same features except for one. This exception feature should

split a given phenomenon into two separate situations. Thus, big and small are antonyms since both mark the size. However, each one denotes an opposite pole of the size. One should note that man and girl are not antonyms because they differ in more than one feature (Modaresi, 2008).

2. Types of antonyms

According to Khodaparasti, one should distinguish between grammatical antonyms and dictionary antonyms.

Grammatical antonyms

In grammatical antonyms, there is a non-lexical prefix in the structure of at least one of the words. In fact, the key characteristic of such antonyms is an attached prefix that gives them a compound structure in the general sense of the word. In this regard, negative prefixes such as 'bi', 'gheir', 'la' and 'na' accompany grammatical antonyms. Because the morphology of these antonyms follows grammatical rules and processes, they adopt characteristics different from those of dictionary antonyms. One should bear in mind that grammatical antonyms often depend on the formal structure so that antonyms are formed based on morphological properties. However, semantic components are responsible for the development of dictionary antonyms. In Persian language, there are two types of grammatical antonyms:

- 1) Antonyms comprising two compounds
- 2) Antonyms comprising a compound and a monomorphemic word

The two words in grammatical antonyms usually share the same stem or root. The antonyms comprising two compounds fall into two groups based on the presence or absence of an infix:

A) If neither of the words has an infix in its structure, a grammatical prefix such as 'bi', 'gheir', 'la' and 'na' precedes one of the words; e.g. Biarzesh (worthless)#Arzeshmand (worthy).

B) Since the stem is the same in either word, the semantic properties of the prefix and suffix will be the contrasting factor. A word contains an infix that functions as the contrasting agent. Example: Tajzieh Pazir (degradable) # Tajzieh na Pazir (non-degradable). However, one should note that the number of such antonyms is not considerable.

In the antonyms comprising a compound and a monomorphemic word, only one word will have a prefix. The distribution of the subsets of this category is based on the commonness and non-commonness of the antonym stem although the words often share the same stem. In the subset A, the antonyms share the same stem. Thus, the opposition is created in antonyms by a negative prefix. Example: Shad (happy) # Na shad (unhappy).

One, however, should note that every word

sharing the same stem and a prefix may not necessary oppose another derivative, which normally lacks the prefix, of the same stem semantically. Example: Morovat (manliness) # Bi Morovat (merciless)

The above set is not an antonym because adding a prefix may change the lexical category of the word per se, and thus the word is delisted from the inventory of possible antonyms. For example, 'Morovat' is a noun. Adding the negative prefix 'bi' renders the word into an adjective. The underlying principle in the antonymy relationship is the sameness of their lexical category.

In the subset B, the words do not share the same stem. Still, they usually share the same root and belong to different inflectional typologies in Arabic. Example: Motaghi (pious) # Bi Taghva (impious).

Dictionary antonyms

The content and concept is more important than morphology in dictionary antonyms. Such antonyms may contain prefixes, yet their prefixes are lexical not functional. That is, they contain prefixes that are considered as content words with specific semantic loading. Therefore, some linguists contend that such words are not prefixes indeed but morphemes – dictionary morphemes that may appear in different lexical combinations like any ordinary word (by compounds, we refers to all types of compound words). The most important morphemes in Persian dictionary antonyms include 'bad', 'pak', 'por', 'pas', 'past', 'pish', 'tar', 'tond', 'tohi', 'khord', 'khoshk', 'khoob', 'khosh', 'dorost', 'dorosh', 'dour', 'dir', 'rast', 'zood', 'sabok', 'sepid', 'sakht', 'sost', 'sefid', 'sangan', 'siah', 'sia', 'ghalat', 'kam', and 'geran'.

A glance at the above words shows that the majority of them are highly-frequent and commonly-used adjectives or adverbs. Moreover, all of them are among the basic Persian words. Besides, the above list contains antonym pairs. In terms of structure, dictionary antonyms may be divided into three distinct groups:

- A) Compound dictionary antonyms
- B) Compound and monomorphemic dictionary antonyms
- C) Monomorphemic dictionary antonyms

Subset A is further subdivided into three categories:

1. Prefixed
2. Suffixed
3. Mixed-type

There are two types of prefixed antonyms: 1) same-headed: such as Khosh Saligheh (tasteful) # Bad Saligheh (tasteless); Kam Mayeh (sparse) # Por Mayeh (abundant) (in these antonym pairs, the prefix creates the opposite meaning); and 2) hetero-headed: such as Sakht Ahd (faithful) # Sost Peiman (perfidious), and Khosh Nama (appealing) # Bad

Manzar (unappealing). Headedness here refers to the head word shared by either word in an antonym pair.

Suffixed antonyms also have two subsets: 1) similar: such as Malal Angiz (tedious) # Shoor Angiz (sensational), Solh Joo (pacifist) # Jang Joo (belligerent); and 2) dissimilar. Hence, the subsets are categorized based on the similarity or dissimilarity of the antonym head words. It is notable that verb roots, particularly present forms of verb roots, constitute an important component of compound suffixed dictionary antonyms, which can be considered as a characteristic peculiar to this category.

Examples of mixed-type antonyms include Doshman Navaz (pro-enemy) # Doshman Setiz (anti-enemy) that are same-headed mixed dictionary antonyms with dissimilar or antithetic suffixes. Normally, the head words are identical while dissimilar suffixes with verb roots oppose each other.

Examples of hetero-headed mixed dictionary antonyms include Farah Bakhsh (pleasurable) # Gham Afza (gloomy), Rooh Angiz (exulting) # Jan Godaz (heart-rending), and Setizeh Gar (aggressive) # Solh Joo (pacifist). This category of antonyms also contains mixed antonyms constituting a pair of low-frequency words that have no verb roots. Example: Sahl-ul-Vosool (easily-accessed) # Sa'b-ul-Hosool (far-reached).

The second group of dictionary antonyms involves pairs in which a member is monomorphemic, and the other is compound. The majority of monomorphemic words are borrowed Arabic words; however, compound words have Persian roots. Example: Mostabed (despotic) # Democrat Manesh (democratic), and Jahel (ignorant) # Daneshmand (pundit).

The third group of dictionary antonyms constitutes monomorphemic words such as Zesht (ugly) # Ziba (pretty), and Sard (cold) # Garm (warm). In terms of semantics, these antonyms are categorized based on the following features:

A) Quality: Zibaie (beauty) # Zeshti (ugliness), Khoob (good) # Bad (bad)

B) Quantity: Aghaliat (minority) # Aksariat (majority), Kam (little) # Ziad (much)

C) Attitude: Rasti (honesty) # Doroogh (untruth), Dorost (true) # Ghalat (false)

D) Emotions and feelings: Eshgh (love) # Nefrat (hatred), Mohabat (affection) # Doshmani (enmity)

E) Actions: Raftan (to go) # Amadan (to come), Ayab (to come) # Zahab (to go)

F) Changes: Poyaie (dynamicity) # Istaie (stagnation), Rokood (depression) # Ronagh (boom, prosperity)

G) Relations: 1) spatial: Vorood (entrance) # Khorooj (exit), Bala (above) # Paiin (under); and 2)

temporal: Ghabl (before) # Ba'd (after), Dir (late) # Zood (soon) (Khodaparasti, 1997)

Linguists enumerate at least ten types of semantic antonyms. Still, there are three main types of semantic antonyms encompassing a few subsets. These are binary, inverse and scaled antonyms (Modaresi, 2008).

A. Binary antonyms

When the semantic relationship between two words is such that one negates the other, the antonyms are said to be binary. Example: Zنده (alive) # Mordeh (dead), Moanath (female) # Mozakkar (male), and Mojarrad (single) # Mota'ahel (married). In a pair of binary antonyms, there are no variable degrees of antonymy relationship. For example, one may not distinguish a more dead or more alive animate being (Modaresi, 2008).

B. Inverse antonyms

In inverse antonyms, the antonymy relationship between two words holds true the other way round. Example: Bala (above) # Paiin (under). Thus, if x is above y, then y is under x.

In Persian language, inverse antonymy relationship exists between all pairs of compound verbs with the same stem where the light verb in one is 'Dadan' (to give) and in the other is 'Gereftan' (to get). There is inverse antonymy between the comparative forms of opposite adjectives. Example: Bozorgtar (larger) # Koochehtar (smaller), and Saritar (faster) # Ahestehtar (slower) (Modaresi, 2008).

C. Scaled antonyms

Scaled antonymy exists between two words when they fall on the either ends of a continuum of interrelated concepts. Example: Kootah Tarin (shortest) # Boland Tarin (tallest), and Ziba Tarin (most beautiful) # Zesht Tarin (ugliest). Scaled antonymy exists between the superlative forms of opposite adjectives. As the name suggests, the antonyms fall on a scaled continuum with varying degrees of antonymy relationship while the intervening degrees reflect a spectrum of scaled features. For example, one can imagine different degrees of shortness between the shortest and tallest (Modaresi, 2008).

Unilateral and bilateral antonyms

Antonymy relationship is unilateral when only one of the words exerts a one-sided impact. For example, in the antonym pairs Shisheh (glass) # Sang (stone) or Sang (stone) # Saboo (pitch), the stone may break the glass and pitch while the latter two cannot break the stone. However, in the pair Rooz (day) # Shab (night), both elements exert the same impact on each other (Modaresi, 2008).

Practical antonyms

In practical antonyms, the two words are not antonyms literally. However, they may be used against each other in their figurative meaning. For

example, Sorkh (red) is used as the facial color of healthy people (Sorkh Rooy (red-skinned)) while Zard (yellow) connotes a lean and pale person (Zard Rooy (yellow-skinned)).

What creates antonymy in words?

One should consider the following points in studying antonyms:

A) Only can the words with a single meaning have absolute antonyms. In polysemic words, every antonym opposes a given meaning of the word. Thus, a polysemic word can potentially have several antonyms. For example, the word 'Azad' (free) can stand opposite to Zendani (prisoner), Asir (captive), Bardeh (slave), Dargir (entangled), and Mashghool (involved, engaged).

B) The literal and figurative meanings of every word have distinct antonyms. Interestingly, the antonym of a figurative concept can have a real, literal meaning and vice versa.

C) In antonyms, gradation makes sense. For example, when Soozan (burning) stands on the one end of the continuum, Yakh Zadeh (frozen) can stand on the opposite end with many concepts in between such as Jooshan (simmering), Dagh (hot), Garm (warm), Velarm (lukewarm), Khonak (cool), Sard (cold), etc. (Khodaparasti, 2008).

3. Statement of the problem

The lexical relations in WordNet exist among the words of the same lexical category such as synonymy, antonymy, hyponymy and metonymy or of different lexical category including feature and derivation (Davari Ardakani & Fakoorian, 2013).

One of the most commonly used sections of a website is the search engine. It is highly important to use website search engines as they decrease the search time. Search in databases is crucial since most of the leading websites use and organize their content in databanks. Different languages have tended to align themselves with computer sciences due to the increasing advancement of such sciences in the world and their permeation into the third world countries. In addition to language, script should also be aligned with computer technologies. Computer scientists and linguists need to modify Persian language and script in order to align it with computer technologies.

4. A review of literature

One of the effective techniques of teaching vocabulary is to use antonyms. Thus, the majority of English course books introduces and includes antonyms. Both Persian dictionaries and extracurricular textbooks list Persian antonyms. Various antonym dictionaries have already been published in Iran as follows:

1. Persian Synonym & Antonym Thesaurus by Khodaparasti (1997)
2. Recently, Seyed Ali Pur Hosseini, a scholar

with Central Library of Astan Qods Razavi, has compiled 'A Dictionary of Persian Synonyms and Antonyms' in 6 volumes sponsored by Faragostar Mohaghegh Research Institute. Mr. Pur Hosseini has compiled the dictionary in order to help accelerate information retrieval (IR). In the preface, the author indicates that the dictionary may have other uses than information storage and retrieval such as developing glossaries, selecting headlines, building up vocabulary in writers and presenters and so on. The dictionary contains 24921 entries, 41826 synonyms and 7781 antonyms. In sum, the dictionary contains 250,000 words. It is hoped that this invaluable work be published as soon as possible (aqlibrary, 1393).

3. Jafar Sadeghi published the 'Dictionary of Analogous Words, Synonyms and Antonyms in Persian Language' that includes the vocabulary, colloquial expressions, new and common words as well as the works of contemporary Persian writers. This dictionary is to supplement the 'Jamal Zadeh Dictionary of Persian Slang'. It is meant to serve as a reference for scholars of Persian texts and be a specific Persian thesaurus of the common words frequently used by prose writers. Some of the characteristics of this dictionary are as follows:

- a) Displaying the pronunciation of words and phrases
- b) The word in parenthesis in front of the entry indicates alternate spelling.
- c) In the words with alternate pronunciation, the contemporary pronunciation is illustrated.
- d) The silent morpheme 'ha' at the end of the words is considered to be a diacritic.
- e) Every entry is followed by its denotation (meaning) in contemporary Persian.
- f) Use of examples
- g) Information is provided on the origin of non-Persian and non-Arabic words.

However, to the best of the author's knowledge, no software has already been developed for keyword-based automatic antonym recall in information retrieval in Iran.

5. Research methodology

In this study, all entries in the Khodaparasti's Persian Synonym & Antonym Thesaurus were extracted and typed along with their antonyms into the computer by a typist. Then a software developer used RICEST software to develop and operationalize the intended program.

First, a page was opened to enter the entries vertically. The entries were entered word by word in the beginning. Afterwards, they were entered as multiple words (e.g. Khahar (sister) # Baradar\ Dadash (brother)).

Subsequently, a template of search window was developed (see Figure 1) and installed on RICEST.

When a word is typed into the search engine, it provides an antonym as the output.

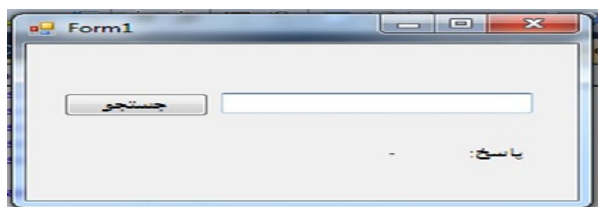


Figure 1. Template of search window

6. Programming language

C# in the form of visual studio was used to develop the software. C# was used as the programming language due to its consistency with RICEST. In this software, the operator may enter a word one at a time and see the search result with the word's antonym.

Advantages of Automatic Antonym Recall Software (AARS)

AARS has the following advantages:

- AARS can be used to increase search words.
- AARS can be used in information storage and retrieval systems.
- AARS can be used in language teaching.
- AARS can be used to compile Thesauri.
- AARS can help increase the vocabulary repertoire of authors, poets, orators and presenters.
- AARS can be used to add diversity to speech and writing.

Research population

The population of the study consisted of all antonym entries in Khodaparasti's Persian Synonym & Antonym Thesaurus.

Conclusion and implication

The present findings showed that computer software may be used in building up vocabulary in writers, poets, presenters and orators. Besides, computer could be used in teaching Persian language and compiling thesauri and dictionaries.

Recommendations for further studies

It is recommended that future studies develop the present software so that:

- a) The operator can obtain several antonyms of the same entry rather than a single antonym.
- b) The operator can obtain both the antonym and synonym of a single entry.
- c) The software may be developed in different

programming languages in order to be installable on smart phones and tablets.

- d) The software may be developed for other languages such as English and French.

Correspondence to:

Shapur Reza Berenjian

Faculty Member with Regional Information Center for Science and Technology (RICEST)

Sh_berenjian@yahoo.com

References

1. Ahmadi Givi, H., & Anvari, H. (2009). Persian grammar 1. Tehran: Fatemi Institute.
2. Abu El-Kahir, I. (2006). Effects of stop words eliminating for Arabic information retrieval: A comparative study. *International Journal of Computing and Information Sciences*, Vol. 4 (3).
3. Davari Ardakani, N., & Fakoorian, N. (1993). Studying morphological semantics and its application in vocabulary networks. *The Second National Conference on Teaching Persian Language and Linguistics*, Shiraz, Iran.
4. Dekhoda, A.A. (1992). *Thesaurus*. Tehran: Tehran University Publications.
5. Falk, J.S. (1992). *Linguistics and language: A survey of basic concepts and applications*. Kh. Gholam Alizadeh (Trans.). Mashhad: Astan Qods Razavi.
6. Jafar Sadeghi, H. (2011). *Dictionary of Analogous Words, Synonyms and Antonyms in Persian Language*. Tehran: Alami Institute Publications.
7. Khodaparasti, F. (1997). *Persian Synonym & Antonym Thesaurus*. Shiraz: Danesh Nameh Fars Publications.
8. Modaresi, F. (2008). *From syllable to sentence*. Tehran: Chapar Publications.
9. Richards, J.C., Platt, J.T., & Weber, H. (1993). *Longman dictionary of applied linguistics*. H. Vosooghi, & S.A.A. Mir Hosseini (Trans.). Tehran: Center for Translation and Publication.
10. Safavi, K. (2007). *An introduction to semantics*. Tehran: Pezhvak Keivan Publications.
11. Voorhees, E. (1994). Query expansion using lexical-semantic relations. *Proceedings of the Seventeenth Annual International Conference on Research and Development in Information Retrieval*. 61-69.