

Automated Synonym Dictionary Generation Tool for Turkish (ASDICT)

Özlem Aktaş*

Çağdaş Can Birant**

Belgin Aksu***

Yalçın Çebi****

Abstract

In this paper, an Automated Synonym Dictionary Generation Tool for Turkish (ASDICT) was briefly described and the development process of the algorithms was given in detail.

By applying the ASDICT onto the data of Contemporary Turkish Dictionary published by Turkish Linguistic Association (TDK: Türk Dil Kurumu), a synonym database was obtained. The synonym dictionary generation process was carried out by applying four processes. As a result of these processes, the definite synonyms were classified as *Definite Synonym* (D_n) and put into the *Synonym List* (*SL*). Some words, which could not be classified as D_n , were classified as *Ambiguity* and stored in a file called *Ambiguity File* (*AF*) to be checked out by supervised methods to build a more reliable synonym database.

The synonym database for Contemporary Turkish Dictionary, which is called "*Definite Synonyms Database* (*DSDB*)", was built by applying ASDICT, and it is currently available on the official web site of TDK (TDK 2009).

Keywords

Synonyms, automated, dictionary, Turkish, natural language processing

* Assist. Prof. Dr., Computer Engineering Department, Dokuz Eylül University, Tinaztepe Campus Buca – Izmir / Turkey
ozlem@cs.deu.edu.tr

** Assist., Computer Engineering Department, Dokuz Eylül University, Tinaztepe Campus Buca – Izmir / Turkey
cagdas@cs.deu.edu.tr

*** TDK Atatürk Blv. Nu.: 217, Kavaklıdere – Ankara / Turkey
belgin.aksu@tdk.gov.tr

**** Prof. Dr., Computer Engineering Department, Dokuz Eylül University, Tinaztepe Campus Buca – Izmir / Turkey
yalcin@cs.deu.edu.tr

1. Introduction

A *dictionary* is a collection of words (also called headwords) in a specific language, often listed alphabetically, with definitions, etymologies, phonetics, pronunciations, and other information (Agnes 1999). Dictionaries are commonly printed as the form of a book, but nowadays they can also be used as online via the Internet.

Specialized dictionaries are sometimes found in specialized areas, such as idioms, proverbs, synonyms, acronyms, antonyms, etc. Some concept of assets may meet the same or different words in a language. These words, the structure of which are different but have the same meanings, are called *synonyms*. Synonyms may have identical or very similar meanings, which are used for communicating easily by expressing thoughts appropriately. Synonym dictionaries have, at least, the following advantages for both general users and researchers:

- Automated data extraction from a large body of text corpus,
- Query clustering applications to help the search engines which use a question-answer structure (Wen et. al. 2002),
- Automatic indexing procedures to help to assign each word-stem to a concept class (Salton 1971),
- Automatic machine translation studies (Edmonds 1999),
- Automatic author recognition through lexical choice (Reiter et. al. 2002).

Apart from their general use, they have the following advantages for researchers in the field of linguistics specifically:

- Defining verbs' conceptual structures and event types in order to provide more complete verb frames for syntactic parser software (Chu-Ren et. al. 2005),
- Producing more lexically cohesive texts for authors from various fields (Donnelly 1994),
- Revealing the interactional relationship between syntax and semantics (Chief et. al. 2000),
- Helping foreign language teachers in teaching vocabulary and helping learners of a second language in using appropriate vocabulary according to the situation (Martin 1984).

In this paper, the Automated Synonym Dictionary Generation Tool (ASDICT) developed for Turkish was described in details. As data source, The Contemporary Turkish Dictionary, in which word synonym and antonyms are given in different forms and prepared by Turkish Linguistic Association (TDK), was used.

2. Automated Synonym Dictionary Generation Tool for Turkish (ASDICT)

2.1. The structure of the data used for ASDICT: The basic data source used in this study was The Contemporary Turkish Dictionary, which includes more than 70,000 words, and was published by the Turkish Linguistic Association (TDK). In this dictionary database, word synonym and antonyms are given in different forms. The dictionary database was taken from the Turkish Linguistic Association, the unnecessary fields and tables were removed from the database and the database structure is simplified as given in Table 1.

Table 1. *The Structure of the Source Data*

Head Word	Meaning
Bayrak	<p>(<i>Flag</i>) Bir milletin, belli bir topluluğun veya bir kuruluşun simgesi olarak kullanılan, renk ve biçimle özelleştirilmiş, genellikle dikdörtgen biçiminde kumaş, sancak. <i>A usually rectangular piece of fabric of distinctive design that is used as a symbol as of a nation, a particular society or an institution, pennon.</i> Öncü <i>Pioneer</i> Simge, sembol <i>Ensign, symbol</i> Baklagillerde diğerlerinden daha üstte bulunan, daha büyük olan ve çoğunlukla başka bir renkte ve yuvarlakça olan taç yaprağı. <i>Any of the petals of pulse that is positioned above other petals, bigger than other petals and that differs in colour.</i> Atletizmdeki bayrak yarışında dört sporcunun elden ele geçirdiği kısa, yuvarlak sopa. <i>The short, rounded baton that the athletes pass from hand to hand in relay races in athletics.</i> Gerektiğinde indirilip kaldırılan, açılıp kapatılan kol. <i>A handle that is lowered or raised / opened and closed when needed.</i></p>
Çözümleme	<p>(<i>Analysis</i>) Çözümlemek işi. <i>The act of analysing.</i> Bir cümledeki kelimelerin hangi kelime türünden olduklarını veya özne, tümleç, yüklem görevlerinden hangisinde bulduklarını belirtme, tahlil.</p>

Defining the part-of-speech and the syntactic functions (subject, complement, predicate) of the words in a sentence, test.

Kelimenin kökünü bulup eklerini gösterme.

Defining the root of the word, and showing the affixes.

Bir metni belirli yöntemlere bağlı kalarak gözden geçirme, tahlil.

Reviewing any text abiding by particular methods, test.

Herhangi bir konunun, bir nesnenin düşüncede veya gerçeklikte kurucu parçalarına ayrılmak yoluyla yapısının, işleyişinin ve gelişim yasalarının ortaya konması işlemi.

The process of revealing the structure, functioning and the developmental rules of any issue or physical object in theory or in practice by separating it to its constituents.

Bir maddenin birleşimindeki yalın cisimlerin niteliğini veya niceliğini anlamak için yapılan işlem, tahlil, analiz.

The dissolution performed in order to understand the quality and quantity of the pure forms in the composition of any substance, test, assay.

Bir sayıyı onluk ve birliklerine ayırıp yazma.

Rewriting any number in its decimals and units.

Beyaz

(White)

Ak, kara, siyah karşıtı.

Hoar, opposite of dark, black.

Bu renkte olan.

Things which are white.

Beyaz ırktan olan kimse.

A person belonging to white race.

Baskıda normal karalıktaki görünen harf çeşidi.

A kind of letter visible in normal darkness at press.

Beyaz zehir.

White poison/heroin.

Kara

(Dark)

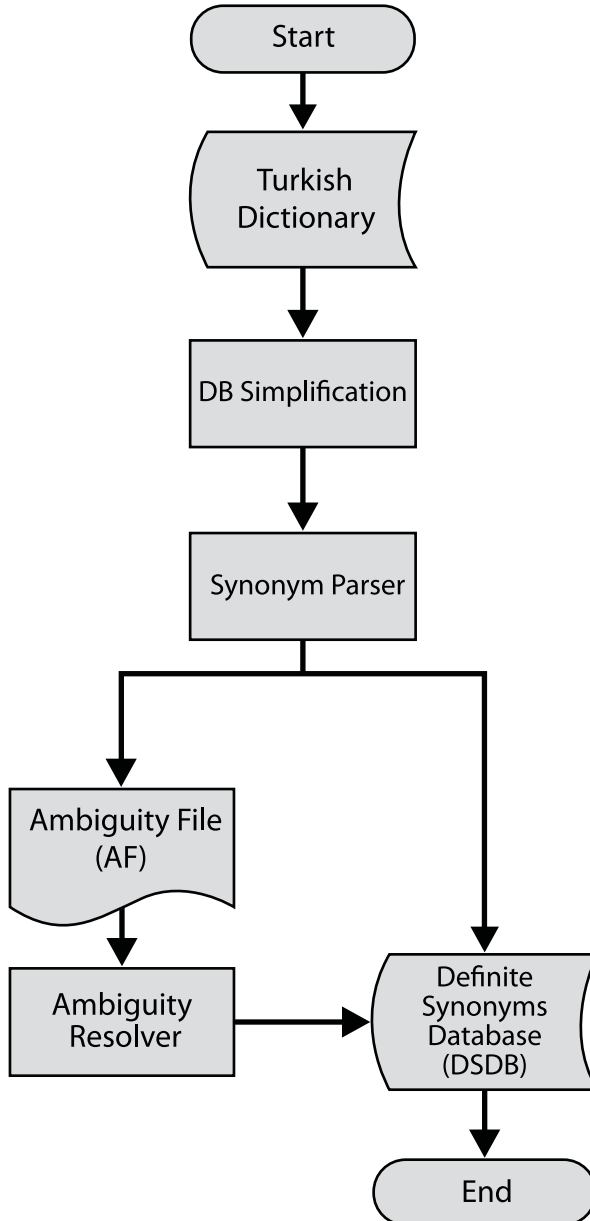
En koyu renk, siyah, ak, beyaz karşıtı

The darkest colour, black, opposite of hoar, white.

As given in Table 1, some words have more than one meaning, and some meanings include both synonyms and antonyms. In general, synonyms are located as separate words in the meaning part, but in some cases, they are located at the end of the descriptive sentence. Therefore, some ambiguities occurred during the synonym extraction process.

2.2 General flow of ASDICT algorithm: In this study, depending on both the characteristics of Turkish and the structure of the Turkish Dictionary, an Automated Synonym Dictionary Generation Tool (ASDICT) was developed. The general workflow diagram of the tool is given in Figure 1.

Figure 1. *General workflow diagram.*

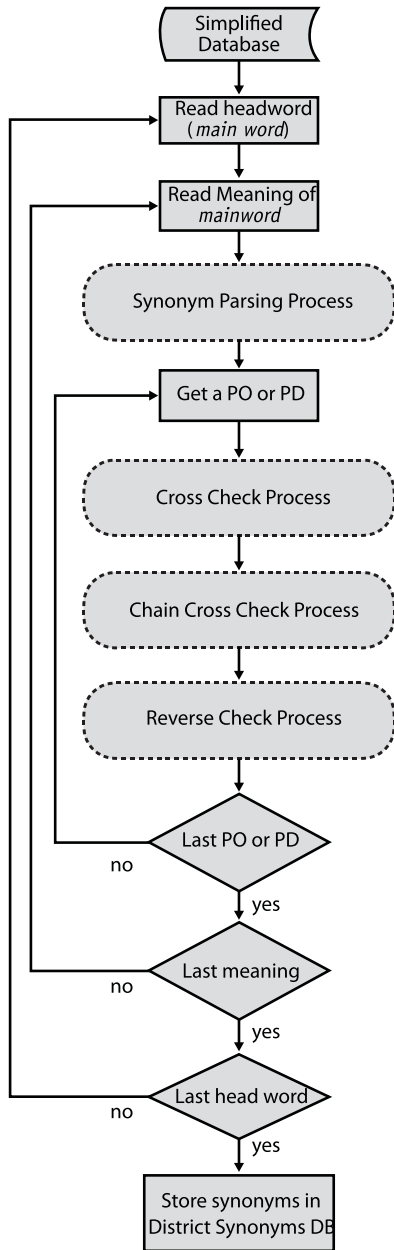


In order to extract synonyms from the database and put them in an order, a synonym parser should be used. By using the simplified dictionary database, including head words and meanings of each word, the *Synonym Parser (SP)* module extracts the synonyms and words causing ambiguities.

The words which are definitely classified as *synonyms* are stored in another database for further processes and the words or phrases classified as *ambiguities* are stored in a file for resolving process which is carried out by supervised techniques.

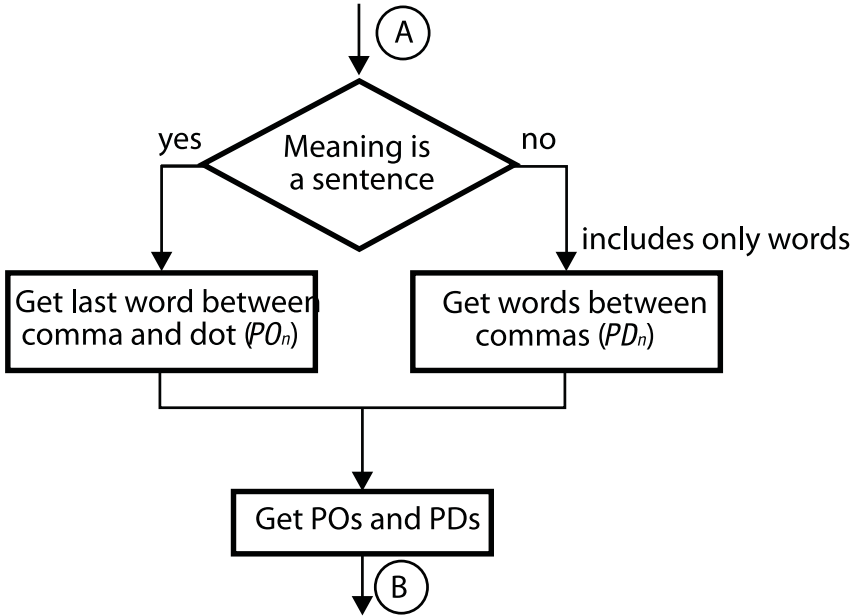
In the dictionary database simplification step, the meanings of the head-words were taken one by one from different tables in the current Turkish dictionary database, and a new simplified database was generated. After generating the simplified database, the synonym parsing process is applied by *Synonym Parser (SP)* module. The flowchart of the SP module is given in Figure 2.

Figure 2. *General workflow diagram of Synonym Parser (SP) module.*



2.3. Description of ASDICT: In the first step of the process, which can be defined as synonym parsing, meanings of head words are taken one by one and examined (Figure 3).

Figure 3. Workflow diagram of Synonym Parsing Process.



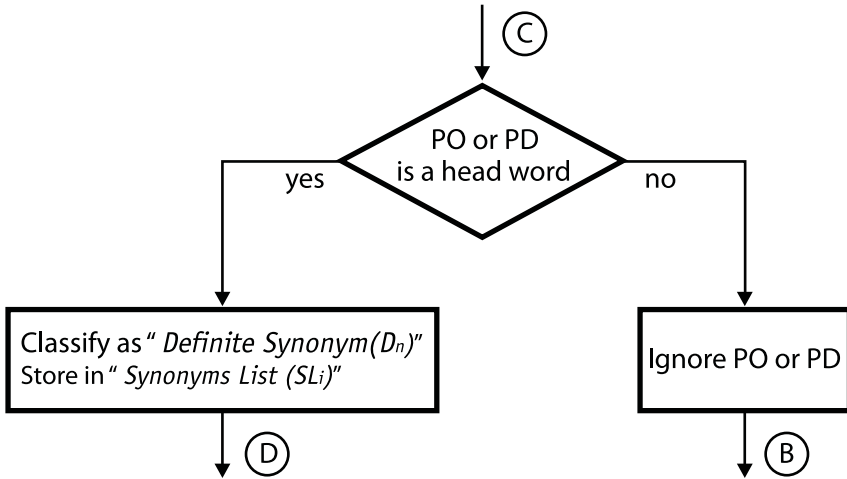
The head word, which its synonym is to be searched, is denoted as *main word*. Each meaning of the main word is parsed and controlled if it is a whole sentence, a single word or a sequence of single words separated by commas (“,”). A word in the meaning of the main word is accepted as a *pre-definite* (PD_n) synonym, when it is located in the meaning part alone, or a single word between commas. Otherwise, it is accepted as a sentence and an additional parsing process is required.

In this process, the sentence is examined from the point of punctuation marks. The part of the sentence between the last comma and the end of sentence mark (“.”, “!”, “...”) is taken as a phrase. This part is considered as a *possible* (PO_n) synonym and will be controlled in the further steps.

For the data given in Table 1, the words “öncü (pioneer)”, “simgе (ensign)”, “sembol (symbol)” were selected as pre-definite synonyms for the main word “bayrak (flag)”, and the word “sancak (pennon)” was selected as *possible synonym*.

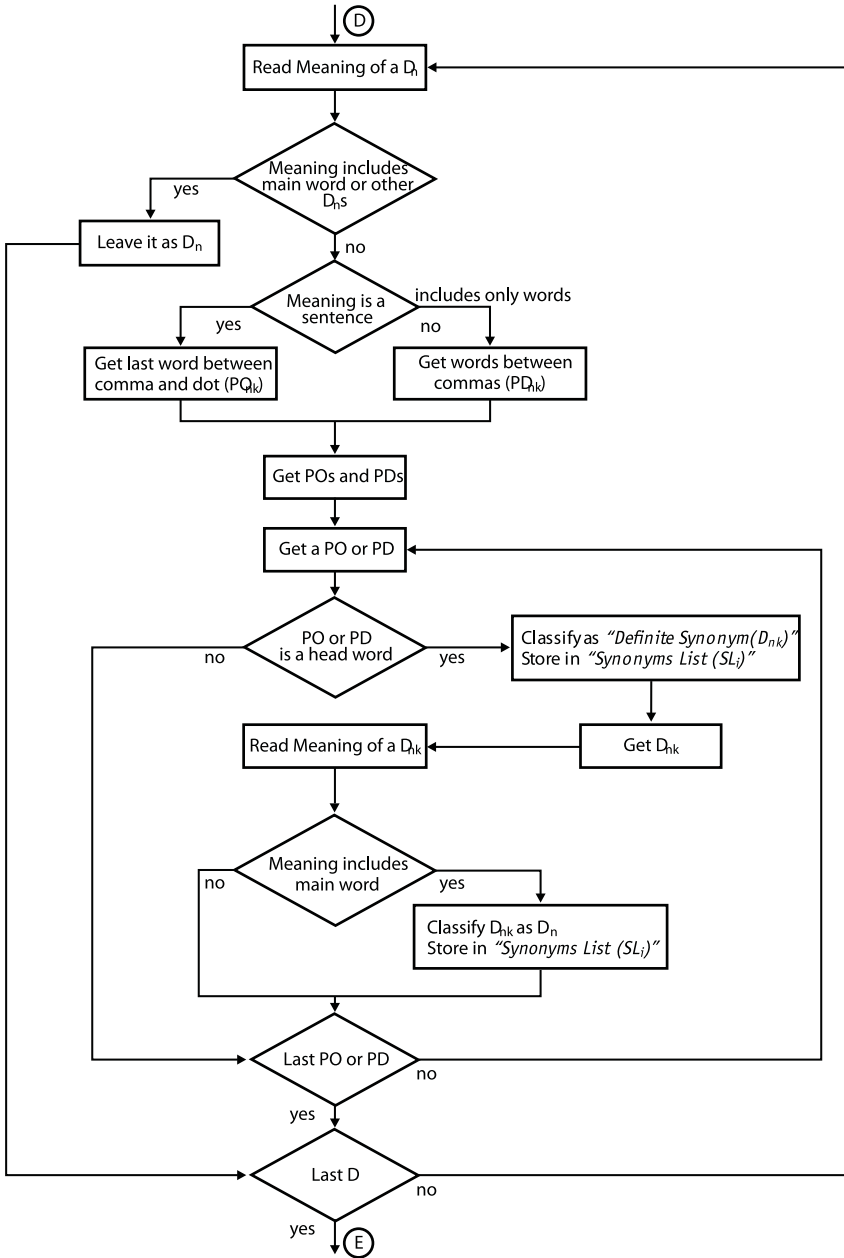
After the synonym parsing process, all PD_n and PO_n synonyms are cross-checked with the head words in the dictionary (Figure 4). If it is found as a head word in the dictionary, then it is classified as *Definite* (D_n) synonym and put into the *Synonym List* (SL_i), otherwise it is discarded.

Figure 4. Workflow diagram of “Cross Check Process”.



After a PD_n or PO_n synonym is classified as D_n synonym, the synonym chain cross check process is applied (Figure 5). In this process, meaning of each D_n is checked as if it includes the main word or any of other D_n s. When the meaning of the D_n does not include either main word or any of the D_n s in the SL_i , the secondary checking process is carried out. In this process, at first, the PD_{nk} and PO_{nk} synonyms for each D_n in the SL_i are found and a secondary level synonym list was created (SL_{ij}). After that, the chain cross check process is applied for the SL_{ij} . If any of the PD_{nk} or PO_{nk} synonyms of the D_n includes only the main word in their meanings as a PD_{nkm} and PO_{nkm} synonym, then PD_{nk} or PO_{nk} , of which includes main word, is classified as D_n .

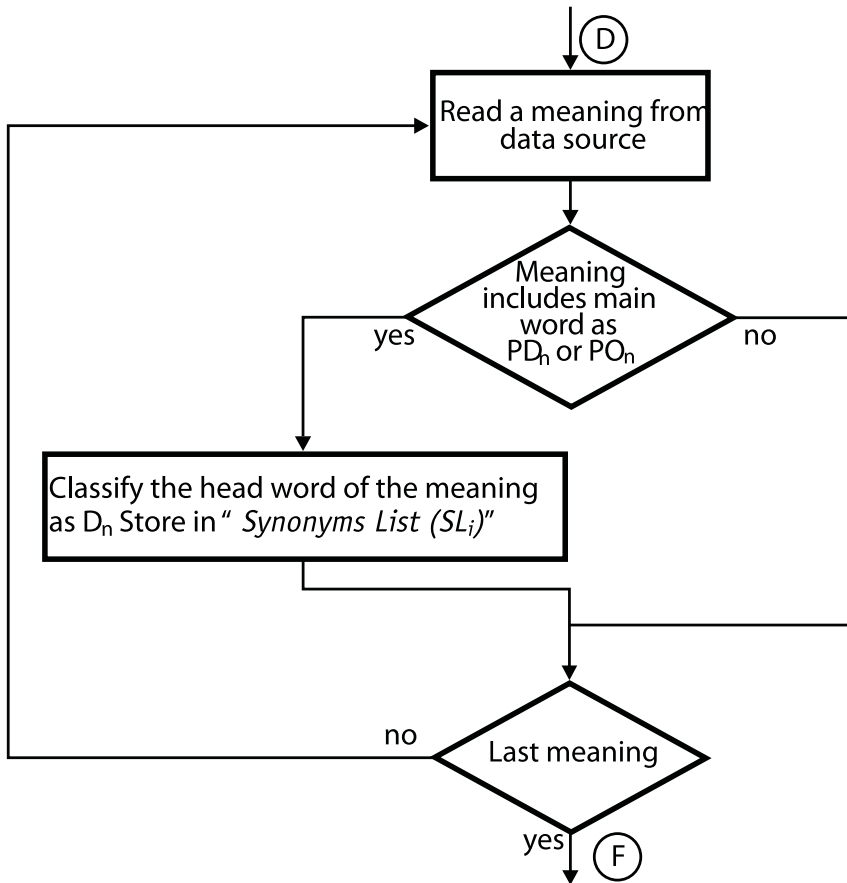
Figure 5. Workflow diagram of Chain Cross Check Process.



The *chain cross check process* continues until no more words are found to be classified as D_n . In some cases, the entire dictionary is required to be checked to find the synonyms of the main word.

In the last step, the *reverse check process* is carried out by checking all meanings in the dictionary as if they include main word as a synonym (Figure 6).

Figure 6. Workflow diagram of *Reverse Check Process*.



When the main word is found, then the head word of that meaning is classified as D_n . The notations, which are used in *Synonym Parser (SP)* module, are simply given in the following.

The processes given above can be explained on a sample word. The selected main word for this process is “adım (step)” which has six different

meanings in Turkish. The meanings of the main word and the words obtained during the processes are given in Table 2.

Table 2. *The Meanings of Main Words, Pre-definite and Possible Synonyms*

Head Word	Meaning
Adım (<i>Step</i>)	İki dış arasındaki aralık <i>The interval between two threads.</i>
	Yürümek için yapılan ayak atışlarının her biri <i>An advance or movement made by raising the foot and bringing it down elsewhere for walking</i>
	Bir ayak atışıyla alınan ve uzunluğu yaklaşık 75 cm olan mesafe <i>The space passed over in one step, approximately 75 cm</i>
Girişim, hamle	<i>Attempt, move</i>
	Bir gösterge ucunun eş olarak ayrılmış yaylardan biri boyunca aldığı yol <i>The distance covered by a probe over one of the springs separated evenly</i>
	Bir yarışın belirli uzaklığı kapsayan bölümlerinden her biri, etap <i>Each of the parts of a race that covers a particular distance, lap</i>
Girişim (<i>Attempt</i>)	Bir işe girişme, teşebbüs <i>Set to work, undertaking</i>
	İki veya daha çok dalga hareketinin, aynı noktaya aynı anda gelmesiyle birbirini yok edebilmesi veya kuvvetlendirebilmesi olayı <i>The event where two or more wave movements arrive at the same point in the same time causing an increase or decrease in the amplitude</i>
Hamle (<i>Move</i>)	Atak <i>Attack</i>
	İleri atılma, atılım, saldırı <i>Rush ahead, exploit, assault</i>
	Saldırı, savlet <i>Assault, onslaught</i>
	Satrançta ve damada taş sürme işi <i>Moving pieces in chess or checkers</i>

Etap (*Lap*)

Aşama

Stage

Adım

Step

Teşebbüs (*Undertaking*)

Girişim, girişme

Attempt, attempting

Atak (*Attack*)

Geveze

Talkative

Düşüncesizce her işe atılan, cüretkâr

One who bursts in upon without thinking, courageous

Çevik, hareketli

Agile, dynamic

Atılım

Exploit

Saldırın, saldırış, hücum, hamle, akın

Offence, assault, assail, move, raid

Gol atmak veya sayı kazanmak amacıyla yapılan akın, ofans

Moves performed in order to score, offence

Atılım (*Exploit*)

Atılma işi

The act of exploiting

İleri atılma

Rush ahead

Hızla ilerleme, hamle, savlet

Forge ahead, move, onslaught

Sayı kazanmak amacıyla yapılan atılış, hücum

Moves performed in order to score, attack

Herhangi bir konuda ilerleme çabası, hamle

Any attempt to progress in any issue, move

Saldırış (*Assault*)

Saldırma işi

The act of attacking

Savlet (*Onslaught*)

Şiddetli saldırı

Fierce attack

Aşama (*Stage*)

Bir yarışın belirli uzaklığı kapsayan bölümlerinden her biri, etap

Each of the parts of a race that covers a particular distance, lap

Önem veya değer bakımından gitgide yükselen bir sıra basamakların her biri, rütbe, mertebe, paye

Each of the steps that increase in terms of importance or value, rank, degree, position

Vanılması istenen bir amaca doğru geçilmesi gerekli dönemlerden her biri, evre, basamak, adım, merhale

Each of the periods to pass in order to achieve a desired goal, phase, echelon, step, gradation

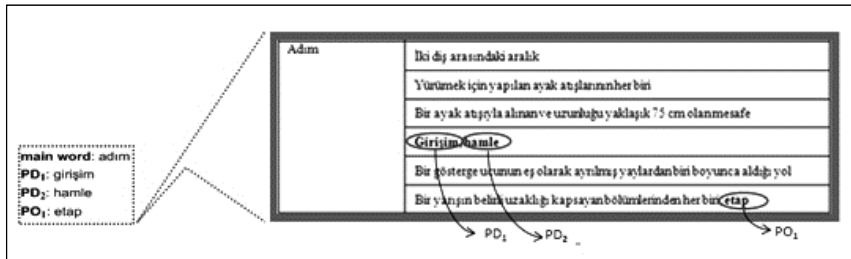
Kadem (*Foot*)

Ayak, adım

Foot, step

In the first step of the general process, which is called as *Synonym Parsing Process*, the words “girişim” and “hamle” and “etap” were found as synonyms. The first two of them are defined as *Pre-Definite (PD_n)* and the last is defined as *Possible (PO_n)* synonym (Figure 7).

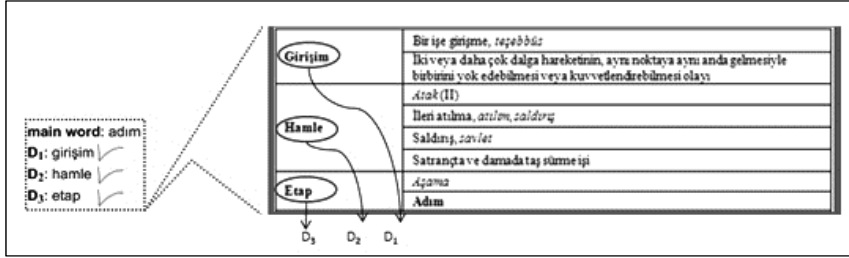
Figure 7. Synonym parsing process for “adım” main word.



After the *Synonym Parsing Process*, the words “girişim (attempt)”, “hamle (move)” and “etap (lap)” were checked if they exist in the dictionary as head words. All of these words were found in the dictionary, classified as D_1 , D_2 , and D_3 synonyms respectively, and put into the SL_1 .

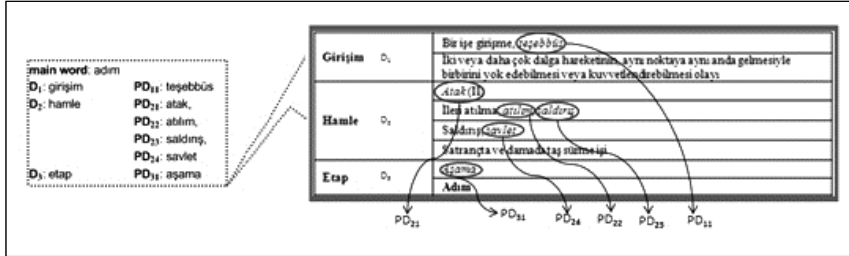
In the *Synonym Cross Checking Process*, the meanings of $D_{n,s}$ did not include other $D_{n,s}$, and the main word “adım (step)” was found only in the meaning of the word “etap (lap)” (Figure 8).

Figure 8. Synonym cross checking process for “adım” main word.



In the *Chain Cross Checking Process* step, firstly, the PD_{nk} and PO_{nk} synonyms of the D_n are found as; “teşebbüs (undertaking)” for “girişim (attempt)”; “atak (attack)”, “atılım (exploit)”, “saldırı (assault)”, “savlet (onslaught)” for “hamle (move)”; “aşama (stage)” for “etap (lap)”. After applying the cross checking process onto these words, they are put into the list SL_{ij} as *Probable (PR_{nk})* synonyms (Figure 9).

Figure 9. First step of chain cross checking process for “adım” main word.



Since only the word “aşama (stage)” includes the main word “adım (step)” as PD₃₁₇ synonym in its meaning part, it was selected as D₄ for the main word “adım (step)”. This step is repeated recursively until no new PR is found in all sub steps.

For this sample, next sub step will be carried out for the words “girişim (attempt)”, “girişme (attempting)”, “geveze (talkative)”, “cüretkâr (courageous)”, “çevik (agile)”, “hareketli (dynamic)”, “atılım (exploit)”, “saldırı (offense)”, “saldırı (assault)”, “hücum (assail)”, “hamle (move)”, “akın (raid)”, “ofans (offence)”, “savlet (onslaught)”, “etap (lap)”, “rütbe (rank)”, “mertebe (degree)”, “paye (position)”, “evre (phase)”, “basamak (echelon)”, “merhale (gradation)” (Figure 10).

Figure 10. Second step of chain cross checking process for main word, “adım”.

PD ₂₁₁ : tepebbüs-	PD ₂₁₁ : girişim-	PD ₂₁₂ : girişim-
PD ₂₁₂ : atak	PD ₂₁₃ : gevize-	PD ₂₁₃ : cüretkar-
	PD ₂₁₄ : çevik-	PD ₂₁₄ : hareketsiz-
	PD ₂₁₅ : atılım-	PD ₂₁₅ : saldırı-
	PD ₂₁₆ : saldırı-	PD ₂₁₆ : hücum-
	PD ₂₁₇ : hamle-	PD ₂₁₇ : akın
PD ₂₂₁ : atılım-	PD ₂₂₂ : hamle-	PD ₂₂₂ : savlet-
PD ₂₂₃ : saldırı-		
PD ₂₂₄ : savlet-		
PD ₂₃₁ : aşama ✓	PD ₂₃₂ : etap-	PD ₂₃₂ : rübe-
	PD ₂₃₃ : meritebe-	PD ₂₃₃ : paye-
	PD ₂₃₄ : evre-	PD ₂₃₄ : basamak-
	PD ₂₃₅ : adım-	PD ₂₃₅ : merhale-

Tepebbüs	Girişim, girişim
	Gevize
Atak	Düptencesizce herşe atılan, cüretkar
	Çevik, hareketsiz
	Atılım
	Saldırı, saldırı, hücum, hamle, akın
	Geçil atılmak veya sayı kazanmak amacıyla yapılan akın, ofans
Atılım	Atılma işi
	İlim atılma
	Hızla ilerleme, hamle, savlet
	Sayı kazanmak amacıyla yapılan atılıp, hücum
Saldırı	Herhangi bir konuda ilerleme çabası, hamle
Savlet	Saldırma işi
	İddetli saldırı
Aşama	Bu aşamın belirlenmesi için yapılacak kapağı yan b-düzenlerinden her bir, etap
	Çnem veya değer bakımından girgede yükselen bu sıra basamaklarının her bir, rübe, meritebe, paye
	Yarıdanma internetlik aşamaca doğru gelişimini gerektiren dönemlerden her bir, evre, basamak, adım, merhale

As the last step, the *Reverse Check Process* is carried out on the dictionary and only the word “kadem (foot)” was detected which has the main word “adım (step)” in its meaning part. Therefore, this word was also selected as a synonym of the main word, and added to the SL₁ as D₅ (Figure 11).

Figure 11. Reverse check process for main word, “adım”.

main word: adım	Kadem	Ayak adım
head word: kadem		
PD ₁ : ayak		
PD ₂ : adım ✓		
D ₅ : kadem		

The definite synonyms of “adım (step)” main word, which are stored in *Definite Synonyms Database (DSDB)*, are “girişim (attempt)”, “hamle (move)”, “etap (lap)”, “aşama (stage)” and “kadem (foot)”.

Both the definite synonyms found by ASDICT, and the definite synonyms determined by supervised techniques are stored into the *Definite Synonym Database (DSDB)* in the form as shown in Table 3.

Table 3. Sample Data in DSDB

Head Word	Synonyms
Çözümleme	tahlil / analiz
çözümlmek	parçalanmak / dağılmak / çökmek
bayrak	sembol / simge / sancak / öncü / alem

2.4. Ambiguities and solving methods: Like synonyms, the antonyms of the head word are written in the full meaning between commas. Therefore, during synonym search process, some ambiguities are appeared such as in the “beyaz (white)” headword which has five different meanings as given in Table 1.

In the first meaning of the headword “beyaz”, the definition is given as “Ak, kara, siyah karıştı”. The words between commas, “ak” and “kara” are accepted as PDs. Also, after applying the method mentioned above onto these words, both of them are determined as *Definite Synonyms (D_n)* of the word “beyaz”.

However, even “ak” is synonym of “beyaz”, “kara” is antonym. Since the word “ak” is located in the meaning part of “kara” and is determined as its definite synonym, this problem cannot be solved during cross check process. Meanings, which include such problems (ambiguities), has also an additional phrase as “..... karıştı (opposite of ...)” at the end. In most cases, this problem should be taken into consideration by taking the preceding word just before that phrase. The PDs or POs, which has such ambiguities, are stored in a file named as *Ambiguity File (AF)* for further supervised resolving processes.

In Turkish, the idioms may include auxiliary verbs such as “olmak, etmek, eylemek”, which turn the status of the preceding word to “verb” when considered together. Therefore, this kind of idioms may be accepted as synonyms of the head words and can be determined by SP. For example, the idiom “analiz etmek (to analyze)” is determined as synonym of head word “çözümlemek (to examine)”.

In the dictionary, there are some head words written as phrases, such as “açıktan tayin (direct appointing)”. In such cases, the synonym can be found in two steps. In the first step, all words in the phrase are taken one by one and synonyms of each word are searched in the dictionary by using the method described above. When the synonym of a word cannot be found, then the word is taken as is. In the second step, the synonyms are combined together forming word phrases. For example, the synonym of “açıktan tayin” is found as “açıktan atama”, since synonym of “tayin” is “atama” in Turkish. Since the structure of Turkish is suitable, this method gives adequate results.

The synonym phrases, which are found as PDs or POs of any main words and cannot be resolved either definite synonyms or not, are stored in the AF to be resolved in the *Ambiguity Solver (AS)* module.

In the AS module, at first, the data in the ambiguity file, some samples of which are given in Table 4, is verified as if they are *Definite Synonyms (D)* or not.

Table 4. *Sample Data in AF*

Bayrak çözümlene tahlil / analiz sembol / simge / sancak / öncü / alem.
Çözölmek parçalanmak / dağılmak / çökmek.
Çözümlene tahlil / analiz.
Çözölmek analiz etmek / tahlil etmek.
Darmadağın tarumar / darmadağınık darmadağın etmek karıştırmak / dağıtmak.

The supervised process is carried out by the experts in Turkish Linguistic Association (TDK) and Linguistic Department of Literature Faculty in Dokuz Eylül University. After the verification process, the synonyms, which are denoted as definite synonym (D) of any head word, are stored into the DSDB.

2. Conclusion

The synonyms dictionaries are required and commonly used especially in the field of linguistics. The studies on the computerized analysis of Turkish were begun in the 1990s and till now a synonym dictionary has not been created. Therefore, with the collaboration of Turkish Linguistic Association (TDK:Türk Dil Kurumu), an Automated Synonym Dictionary Generation Tool for Turkish (ASDICT) was developed.

In this study, ASDICT was described in general. The data used throughout this study was taken from Contemporary Turkish Dictionary, which is published by TDK. By using this dictionary data, a synonym database for Turkish was developed and has currently been published by TDK in the official web site at www.tdk.org.tr (Figure 12).

Figure 12. Synonyms of “bayrak” in the “Online Synonyms Dictionary of TDK”.



Since the dictionary database was not suitable and included unnecessary data, a simplification and filtering process was carried out on it by removing unnecessary fields and tables. Although the synonyms in the dictionary were correctly determined by ASDICT, in some cases, depending on both the nature of Turkish and structure of the dictionary, some words were mis-determined. For example, some words have more than one meaning, and some meanings include both synonyms and antonyms in the data source, resulting ambiguities in synonyms extraction process. In general, synonyms are located as separate words in the meaning part such as “bayrak: öncü”, but in some cases, they are located at the end of the descriptive sentence after the “,” (comma) mark such as “bayrak: Bir milletin, belli bir topluluğun veya bir kuruluşun simgesi olarak kullanılan, renk ve biçimle özelleştirilmiş, genellikle dikdörtgen biçiminde kumaş, sancak”. However, the antonyms are located in the same way, which also causes ambiguity, such as “beyaz: Ak, kara, siyah karşıtı”. In

order to generate a reliable synonym dictionary and overcome these kinds of ambiguities, supervised methods are required. For the synonym dictionary, all ambiguities were overhauled and finalized by the experts in Turkish Linguistic Association (TDK) and Dokuz Eylül University, College of Social Science and Literature Linguistic Department. The ambiguities, which arised from the non-standardized syntax of the Contemporary Turkish Dictionary, can be resolved by replacing the punctuation mark “,” (comma) with “;” (semicolon) such as “beyaz : Ak; kara, siyah karşıtı”. After this process, more synonyms can be correctly extracted from the dictionary; therefore the usage of supervised methods can be minimized.

Acknowledgment

The authors wish to thank Turkish Linguistic Association and Dokuz Eylül University Linguistic Department of the Literature Faculty for their support and help in this project.

References

- Agnes, Michael E. (1999). *Webster's New World College Dictionary*. 978-0028631189, Fourth Edition.
- Chief, Lian-Cheng, (2000). What can near synonyms tell us?, *Computational linguistics and Chinese Language Processing* (5/1): 47-60.
- Chu-Ren, Huang and Jia-Fei Hong (2005). “Deriving Conceptual Structures Form Sense: A Study of Near Synonymous Sensation Verbs”, *Journal of Chinese Language and Computing (JCLC)*. Volume 15. No 3. Singapore.
- Donnelly, Colleen (1994). *Linguistics for Writers*, Albany, State University of New York Press.
- Edmonds, Philip (1999). *Semantic Representations of Near-Synonyms for Automatic Lexical Choice*, Ph.D. thesis, Computer Science Department University of Toronto, Toronto.
- Martin, Marilyn (1984). *Advanced Vocabulary Teaching: The Problem of Synonyms*, *The Modern Language Journal* (68/2): 130-137.
- Reiter, Ehud and Somayajulu Sripada (2002). “The SMART Retrieval System - Experiments in Automatic Document Processing”, *Computational Linguistics* 28/4, 447-485
- Salton, Gerard (1971). *The SMART retrieval system: Experiments processing*, Englewood Cliffs, NJ: Prentice-Hall Inc.
- Turkish Language Association (Türk Dil Kurumu: TDK), tdk.org.tr/esveyakin/ (25.10.2009)
- Wen, Ji-Rong, Jian-Yun Nie and Hong-Jiang Zhang (2002). “Query clustering using user logs, *ACM Transactions on Information Systems*. 20/1 59-81.

Türkçe için Otomatik Eşanlamlılar Sözlüğü Oluşturma Aracı (ASDICT)

Özlem Aktaş^{*}

Çağdaş Can Birant^{**}

Belgin Aksu^{***}

Yalçın Çebi^{****}

Öz

Bu makalede, gerçekleştirimi yapılan Türkçe için Otomatik Eşanlamlılar Sözlüğü Oluşturma Aracı (ASDICT) ve algoritmaların geliştirme süreçleri detaylı bir şekilde anlatılmıştır. ASDICT, Türk Dil Kurumu (TDK) tarafından yayımlanan Güncel Türkçe Sözlük veritabanı üzerinde uygulanarak bir eşanlamlılar veritabanı elde edilmiştir. Eşanlamlılar sözlüğü oluşturma süreci, dört ayrı işlem uygulanarak gerçekleştirilmiştir. Bu işlemlerin sonucunda kesin olarak belirlenmiş olan eşanlamlı kelimeler *Kesin Eşanlamlı (Definite Synonym (Dn))* olarak sınıflandırılmış ve *Eşanlamlılar Listesi'ne (Synonym List (SLi))* kaydedilmiştir. Kesin Eşanlamlı olarak sınıflandırılmayan bazı kelimeler, *Belirsiz (Ambiguity)* olarak sınıflandırılarak, daha güvenilir bir eşanlamlılar veritabanı oluşturabilmek amacıyla, denetimli yöntemlerle kontrol edilip belirlenmek üzere *Belirsizlik Listesine (Ambiguity File (AF))* kaydedilmiştir. İşlemlerin sonucunda, Kesin Eşanlamlılar Veritabanı (Definite Synonyms Database (DSDB)) olarak adlandırılan eşanlamlı kelimeleri içeren veritabanı oluşturulmuş ve Türk Dil Kurumu'nun resmi web sitesinde kullanıma açılmıştır (TDK 2009).

Anahtar Kelimeler

Eşanlamlı, otomatik, sözlük, Türkçe, doğal dil işleme

* Yrd. Doç. Dr., Bilgisayar Mühendisliği Bölümü, Dokuz Eylül Üniversitesi – İzmir / Türkiye
ozlem@cs.deu.edu.tr

** Araş. Gör., Bilgisayar Mühendisliği Bölümü, Dokuz Eylül Üniversitesi – İzmir / Türkiye
cagdas@cs.deu.edu.tr

*** Uzman, Türk Dil Kurumu – Ankara / Türkiye
belgin.aksu@tdk.gov.tr

**** Prof. Dr., Bilgisayar Mühendisliği Bölümü, Dokuz Eylül Üniversitesi – İzmir / Türkiye
yalcin@cs.deu.edu.tr

Автоматический словарь синонимов турецкого языка (ASDICT)

Озлем Акташ^{*}

Чагдаш Жан Бирант^{**}

Белгин Аксу^{***}

Ялчын Чеби^{****}

Аннотация

В данной статье подробно описан созданный инструмент автоматизированного создания словаря синонимов турецкого языка (ASDICT) и процесс разработки его алгоритмов. Путем применения ASDICT к базе данных современного турецкого словаря, изданного Ассоциацией турецкого языка (TDK: Türk Dil Kurumu), была получена база данных синонимов турецкого языка. Процесс создания словаря синонимов состоял из четырех этапов. В результате этого процесса, точно определенные как синонимы слова были классифицированы как определенно синонимы (Dn) и сохранены в списке синонимов (SLi). Некоторые слова, которые не были точно определены как синонимы, были классифицированы как неопределенные и сохранены в списке неопределенных (AF), который в целях создания более надежной базы данных синонимов будет рассмотрен и проконтролирован проверенными опытными методами.

В результате этого процесса была создана база данных синонимов, названная «База данных точных синонимов турецкого языка» (Definite Synonyms Database (DSDB)), которая в настоящее время доступна на официальном сайте Ассоциации турецкого языка (TDK 2009).

Ключевые слова

синонимы, автоматизированный, словарь, турецкий, естественная языковая обработка

* Док., кафедра электронной инженерии, Докуз Еулул университета – Измир / Турция
ozlem@cs.deu.edu.tr

** Науч. сотр., кафедра электронной инженерии, Докуз Еулул университета – Измир / Турция
cagdas@cs.deu.edu.tr

*** Спец-т, Турецком языке ассоциация – Анкара / Турция
belgin.aksu@tdk.gov.tr

**** Проф. докт., кафедра электронной инженерии, Докуз Еулул университета – Измир / Турция
yalcin@cs.deu.edu.tr