

Turkish Natural Language Processing Studies

Yılmaz İnce, E

Department of Computer Technologies, Isparta University of Applied Sciences, Isparta, Turkey

ABSTRACT

Natural language processing which is an engineering field which is concerned with the design and implementation of computer systems whose main function is to analyze, understand, interpret and produce a natural language. Literature review was conducted using Turkish natural language studies documentation method. Documentation on the current method of examination to be reliable scientific research and thesis studies on natural language processing in Turkey were examined by scanning pages of the thesis of Higher Education. Evaluated in terms of the subjects of the study samples obtained as a result of the literature review; Morphological analysis studies, syntactic analysis studies, semantic analysis studies and problem analysis criteria were determined and presented.

KEYWORDS: *natural language processing; morphological analysis; syntactic analysis; semantic analysis; problem analysis*

How to cite this paper: Yılmaz İnce, E "Turkish Natural Language Processing Studies" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-1, December 2019, pp.1122-1128, URL: www.ijtsrd.com/papers/ijtsrd29831.pdf



Copyright © 2019 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



1. INTRODUCTION

Natural language processing which is defined as an engineering field which is concerned with the design and implementation of computer systems whose main function is to analyze, understand, interpret and produce a natural language. Applications are developed by using methods such as morphological, syntactic, and semantic analysis with natural language processing, and rules are created for problem situations and solutions are produced by analyzing problems (Nabiyev, 2012). Morphological, syntactic and semantic analysis is used to control and correct spelling errors by using natural language processing methods, computerized translation, information extraction, information retrieval, development of question and answer systems, summary subtraction and automatic exam scoring.

Morphological analysis examines the suffixes, the types of suffixes and the lineage of words, which are the structural features of language. Syntactic analysis explores the hierarchical harmony of the sequence of words and the elements of the words that make up the sentence. Semantic analysis enables the matching of discrete words to appropriate objects in the database using a corpus of corpus. In addition, semantic analysis is concerned with the development of appropriate models for the integration of discrete words.

2. METHOD

Literature review was conducted using Turkish natural language studies documentation method (Yıldırım & Şimşek, 2000). Documentation on the current method of examination to be reliable scientific research and thesis studies on natural language processing in Turkey were examined by scanning

pages of the thesis of Higher Education. Evaluated in terms of the subjects of the study samples obtained as a result of the literature review;

- Morphological analysis studies
- Syntactic analysis studies
- Semantic analysis studies
- Problem analysis studies.

3. TURKISH NATURAL LANGUAGE PROCESSING STUDIES

3.1. Morphological analysis studies

Güngör (1995) conducted a morphology analysis in Turkish. In this study, additional structure and order were examined and the morphological structure of Turkish was modeled by computer in extended transition network formation. In addition, the software developed in the study can make spell checking and correction. Commitment decomposition is the method that enables the analysis of that sentence by detecting binary relations between words within a sentence.

Eryigit (2006) examined and modeled the loyalty decomposition of Turkish. In this study, different types of parsers have been developed and the performance of the parsers and models have been compared and it has been shown that the classifier-based discriminatory discriminator-based discriminatory discriminator has the best results.

Alkım (2006), designed for Turkish as a shell software for the morphological analysis and dictionary design in natural language processing for Turkish, supported by interfaces that allow to change and improve the rules of dictionary and

morphology. This software can be used for different languages.

Karadeniz (2007), in his research aimed to determine the correct one of the results produced by the morphological analyzer for Turkish. In this study, the uncertainty distributions of the Turkish words were extracted and then the words were clustered according to their uncertainty qualities. Uncertainties are eliminated by writing rules for each type of uncertainty. The performance of the study was tested and it was found that the performance was 82.6%.

Shylov (2008), two-sided morphological machine translator application for Turkmen and Turkish. In the research, morphological analysis and morphological productive practice was conducted for Turkmen and Turkish. Root dictionaries were created for Turkish and Turkmen.

Kışla (2009), in his research, one of the main problems in the field of natural language processing, morphological analysis and word type detection problems presented for the original solutions. For the morphological analysis, which is theoretically known as NPcomplete, the complexity for additive languages has been put forward by a simplified method considering the grammatical features of Turkish. The method, which uses statistical and rule-based approaches together, provides a single result as a result of the analysis and ensures the elimination of uncertainty. In addition, the proposed method uses a closed and limited dictionary, which is an important feature that differs from other methods.

Yilmaz (2009), developed the morphological parser to add the results of the simplification of the words, the full morphological analysis of verbs, the numerical value of the date, time, such as analysis of words, abbreviations and special names can be modified for the production and additions, especially for Turkish users such as tagging has features. In addition, different features have been realized with the text shredder that brings the text to be analyzed into the analysis format, comparison with other analyzers and word derivation modules that produce words from the additional series.

In study, Savaşçı (2010) determined the degree of completeness of Turkish and English with statistical data, and the degree of completeness of Turkish was 0.56 and the degree of completeness of English was calculated as 0.12. Therefore, the difficulty level of natural language processing in Turkish is higher than in the English language.

Aktaş (2010) in order to determine the morphological features of languages, a corp that can represent the features of language is required. In this study, Natural Language Processing methods for Turkish have been developed by using a rule-based approach and an infrastructure called Rule-Based Automatic Corpus Generation has been created to realize the methods.

Yalçınkaya (2018) identifies and eliminates the errors that may occur in the software development life cycle at an early stage and conducts review activities within the requirements management process. The automatic review of the Turkish software requirements and the successful elimination of errors are directly proportional to the scope, accuracy and

success of the language processing algorithms. For this purpose, a new Formal Automatic Requirement Review tool has been developed for Turkish software projects.

3.2. Syntactic analysis studies

Hallaç (2007) developed TurPOS, a rule-based Word Types Identifier developed for Turkish, aiming to assign the most appropriate word class to each word in a given Turkish document. TurPOS uses a collection produced by the morphological analysis application as the input document. The system also uses a rule file containing Turkish grammar rules.

In study, Agun (2008) presented a graph-based automatic learning model for learning the syntactic features in Turkish. In this study, the graph model designed by using a collection was trained and the correct syntactic tags for a given sentence were extracted through this model. In the design of the model, the probability based graph model, Hidden Markov Models and graph theory were used. In the present study, unlike other probability based labeling algorithms and statistical natural language processing studies, a probability based diagram model was developed in which Turkish morphological features can be used.

Eker (2009) studied syntactic decomposition, one of the main problems of natural language processing. The two most common parsing, heap structure parsing and independence parsing. In the study, parsers were evaluated with textual requirements and the chunk structure parser got the highest score.

Şentürk (2009), in his study, tried to make a rule-based find-and-replace function for the Turkish language. The most important reason for this study is that the writing programs are inadequate for the Turkish language. In this research, finite state machines were created and correct analyzes were made for all root and attachment types. It is ensured that the suffixes of the found words are analyzed correctly and the correct replacements are made with these resolved suffixes.

Kutlu (2010) developed a noun phrase extraction system for Turkish texts. In the study, a weighted restrictive dependent parser was used to show the relationship between sentence components and to find noun phrases.

Tahiroğlu (2010) created a web-based compilation dictionary to be used in computer-assisted dictionary studies in his research. MySQL is the database of the compiled dictionary and PHP is encoded in the web interface. It is thought that the developed software can provide data for future lexicology research and dictionary writing during the preparation of online, shared dictionaries.

A rule-based system has been developed to identify phrases in Turkish and a software that uses the system has been prepared. It is provided to define Turkish word structures for determination of word types with the developed software (Turna, 2011).

Zafer (2011) developed a general sentence analyzer for Turkish languages. The developed analyzer is based on context-independent grammar rules, morphological analysis and validity rules. In this study, the grammar of Turkish and

Turkmen was examined in terms of computer and the system was implemented for Turkish. Later, it was adapted to Turkmen with partial changes.

Kazkılınç (2012) the subject of the news text, predicate, place and time to specify the phrase in the text, the text has been tagged. For this purpose, the most dominant subject, predicate, place and time were selected from the sentences in the text. Thanks to these tag information obtained in the study, the subject of the text is represented and can be used as a tag in the semantic network and can be used to reach the desired data in search engines.

3.3. Semantic analysis studies

Çakıroğlu (2001) In this study, a model was developed to realize and solve simple arithmetic problems by computer. When the semantic analysis was performed, the information base used was arranged as semantic network. In the solution of problems; The meanings of the sentences were determined through semantic networks and necessary calculations were made in accordance with these meanings. In order to perform semantic analysis in the processing of Turkish as a natural language by computer, a software has been developed and the performance level of this software has been determined.

Orhan (2006), the most appropriate algorithms and features that will enable the clarification of the meaning of the words in the Turkish text has been discussed. For Turkish, words and meaning classes that can be used in word clarification studies have been created, handwriting of texts to be used in algorithms has been realized and a conceptual dictionary has been prepared and a significant contribution has been made to the researches in this field.

Bahadır (2007) conducted the analysis and symbolization of arithmetic mathematics problems at the level of primary school students in his research. Modules of the application developed; morphological analysis of words, determination of the suffixes of the words, determination of the pattern of questions, solving the problem, expressing the solution in the form of equations, symbolizing the problem and the solution.

Birant (2008), developed by Aktaş (2006) and accuracy rate is 99% based on the end-of-sentence algorithm and some of the changes used in this study has made a standardized structure. In this research, the XML structure is used in the files that are presented to the user as the result of the targeted morphological study program and the rules of interposition between the attachments during morphological parsing. The names of the attachment types to be included in the XML structures have been systematically shortened within the framework of a set of rules, allowing the labeling of roots and attachments.

Özdemir (2009) worked on the process of meaning clarification by selecting the appropriate features and algorithms for the words with ambiguity in Turkish texts. In the research, due to the lack of marked texts on which the studies will be conducted, data consisting of sentences containing the selected sample words were collected first. Afterwards, the features that distinguish the meaning for the selected words were determined and Instructional Learning algorithms were applied to the data and the results obtained with the evaluation methods were evaluated. NaiveBayes, Kstar, SimpleCart and Bagging algorithms were used to

clarify the meaning of the test processes. In addition, features that could be effective in the study were determined for the words. It has been demonstrated how effective these selected algorithms and features are.

Adalı (2009) worked on the automatic processing of Turkish documents and extracting information from these documents. While traditional information extraction systems treat input text as sequential words, they generally focus on semantic features, and the clues provided by the document structure are utilized with the help of the proposed architecture. However, in order to ensure document consistency, the relationships between the assets in the document are tested and the extracted assets are compared with the data actually used. Proposed integrated architecture; Morphological analysis module for Turkish includes document structure analysis module, field ontology, inference ontology.

Kalender (2010) suggests a system that automatically generates semantic labels for documents in its research. In this study, UNIPedia is to provide a knowledge base containing contemporary (current words on the web) references. UNIPedia associates various ontological knowledge bases with WordNet concepts. Wikipedia and OpenCyc knowledge bases, which contain up-to-date and reliable information, are mapped to WordNet concepts. Rule-based heuristics using ontological and statistical properties of concepts were used to relate knowledge bases.

Per (2011), in his research, put forward a concept extraction system for Turkish. Due to the fact that Turkish characters do not fit into the computer language and the complex structure of the Turkish add-on, a preliminary processing step is required first. As a result of the preprocessing, only the names of the words that were separated from the affixes were used. Although the system produced more concepts than it had to produce, it found the concepts of documents with 51 percent success. Considering the fact that the concepts do not appear exactly in the documents in terms of structure and the complex structure of Turkish, this result can be considered quite successful.

Numerous keyword extraction and text summarization algorithms in the field of natural language processing, some of which we discussed in the study (Güvenç, 2016). To understand the methods we started with a research on automatic text summarization. The results show that our summarizing algorithms give the best results on news texts and provide less optimal results for short stories.

Kazmı (2017) proposed a new methodology based on solution set programming, focusing on fixed coded rules in our Inspire system and predicting the Interpretable Semantic Similarities with a rule-based approach. It has shown that the improvements made in the research have obtained similar results with the latest technology systems in learning big data sets.

In the study, Aktaş (2017) created a Wordnet ontology that contains network terms separated from its attachments, and the connection between the two terms in the ontology was calculated. The most effective way to name terms is natural language processing. It is aimed to search computer network terms in an ontological dictionary and to add the words that

are not in the dictionary to the ontology automatically. The terms in the close-up and up-down relationship are connected with each other as graph structure. All terms were linked to each other and dictionary ontology was created. Firstly, concept map was used to make correlation and then it was automated with natural language processing algorithm. The algorithm has been prepared by looking at the ontology nodes of the 10 words in the dictionary on the right and left of the word that are not in the paragraph and are associated with the meaning relation close to the top node.

Ergüven (2018) focuses on word representations. In word representations, previous methods rely on the counting of statistics between the word and the words it accompanies, while the current methods are learning-based. In the thesis, the relationship between these two approaches was investigated. Both approaches used context as normalization factor. The word representation of a multi-meaning word includes more than one meaning. To overcome this problem, a method has been developed that provides a separate representation for each meaning of the multi-meaning word.

3.4. Problem analysis studies

Amasyalı (2003) conducted a natural language question answering system for Turkish. The system first translates the question that the user asks in natural language to the search engine query and selects possible answer sentences from the search engine's results page or pages in the links. The first five sentences that score the highest possible scores according to various criteria are communicated to the user. The system was evaluated with 524 questions and was able to answer approximately 43% of the questions when the search engine's results page was used, and 60% when the pages in the links of the results page were used.

An expensive and slow method of manual translation with fast and high error rate in the study of computerized transcription methods are combined (Akman, 2004). Patterns containing the hypotheses about conversations to be written and the output of a speech recognition engine were transformed into letter-based, requisite, weight-finite status receivers, trained in a text collection that corresponded to speech data in terms of content, and combined with a letter-based statistical language model.

Eş (2005) studied the subject of filtering that can automatically sort e-mails into legal or unwanted categories. In this research, it has been tried to obtain effective results in electronic mail filtering business by using some machine learning methods and some ideas of obtaining information in order to automatically classify electronic mail as legal or unwanted.

Yıldırım (2005) was made meaningful by using the statistical natural language processing methods of sms messages. In this study, a solution method has been presented with N-Gram method which is one of the most important statistical natural language processing methods.

Tülek (2007) conducted a text summarization study for Turkish. Considering the structure of Turkish in the software, different statistical methods for summarizing a text were introduced and implemented with software and their suitability to Turkish was discussed. As it is necessary in all other Turkish information retrieval systems, the effect of

different staging algorithms on summarizing performance is examined in order to observe the Turkish additive structure. In order to have higher performances, morphological parser which uses possible stem and attachment combinations of words is used in the implemented staging algorithms. These categorized words are examined through different summarizing methods and the sentences to be included in the summary are determined for each method. Then the results produced by these methods were combined to form the final summary.

Delibaş (2008) found the roots of the words in the sentences in the text entered, parsing the suffixes, testing the accuracy of the word, suggesting words to the misspelled words and adding non-Turkish foreign words to the dictionary on the control of Turkish spelling errors by natural language processing.

Yılmaz (2008) aimed at high school students and university candidates and developed a natural language interface prototype for the semantic network that processes information about Turkish Universities. When asked questions in sentences about the developed software, the answers are reflected on the screen and a program has been prepared to assist the students' university choice period.

Aksoy (2008), in his research using natural language processing and genetic algorithm has created a three-dimensional scene. In this study, an application software that produces 18-dimensional scene interpretation by using genetic algorithm application, which takes 18 relations between objects and objects given within the framework of uses the settlement rules defended in the study for conformity assessment.

Kopru (2008) presented a unique approach that integrates automatic voice recognition and automatic translation systems for voice translation purposes. The presented method is unique in that it includes the first rule-based automatic translation system capable of processing audio data in word Web search engines are used to find documents containing the information sought. However, in many cases the user needs more specific information than a set of documents. Question answer systems address this problem. Question answer systems return explicit answers rather than a set of documents as the answer to a question. In his study, Er (2009) developed a software with a pattern matching approach for answering questions in Turkish singular answers.

Görmez (2009) realized the Turkish text-to-speech system software with machine learning algorithms. The system has two different audio databases consisting of units recorded directly and interrupted from a continuous conversation, and the software allows the units to be disconnected manually and automatically from the conversation. The signal characteristics of the audio signal such as the number of crossings from zero and the energy of the sound are used for automatic cutting.

Soysal (2010) obtained structured information from Turkish radiology reports using natural language processing and a field ontology methods. Field ontology was used during information extraction, asset identification and relationship

extraction stages. Since Turkish is a morphologically rich language in the natural language processing processes of the study, the system was used as a morphological analyzer and these morphological features were utilized in inference rules. Radiology departments use techniques that visualize patients' bodies and these images are examined by doctors and put into plain text reports. For medical information systems, it is important to extract information from these plain text reports.

Hadımlı (2011) proposes two methods that can be used in the processing of Turkish radiology reports, one of which is rule-based and one is database-based. Unlike previous medical natural language processing studies, neither of these methods uses a medical dictionary or a medical ontology. Knowledge extraction is done at the level of determining the medically related phrases and their relationships in the given sentence. The aim is to determine the reference performance that Turkish features can offer for medical information extraction and access in the absence of other factors.

Albayrak (2011) investigated the relationship between the use of Turkish and psychological status. In this study, Turkish articles were collected from depressed, depressive, anxious and non - anxious individuals. Writings were analyzed according to the use of features such as word, modals, person pronouns, verbs and nouns which are used most by each word, each diagnostic group. A program was developed to test the differences in the use of interpersonal vocabulary. The results of the test show that the use of words in Turkish gives many clues about the psychological state.

Çelikkaya (2015) aims to demonstrate the effect of advanced natural language processing methods on these applications and to develop a high-performance personal assistant application that will work with Turkish natural language input on smartphones and tablets. A virtual assistant is an application that contains almost every component of artificial intelligence. The hybrid model developed was 98.30% successful on our test data. Then the parameters that are meaningful for the specified service are extracted and the information is returned from the 3rd party sources for the required services. The success of the whole system was determined as 70.79% on average.

Gökdeniz (2016) aims to exclude the relationships between brain parts from published articles using natural language processing techniques. With a linguistic approach, the sentences containing the relationships depending on the patterns were selected, and then the related brain sections and their relations with each other were extracted by using the loyalty separator and element separator on these sentences. In this study, a connection graph showing the connections of brain sections with each other is determined by determining the direction of these relationships. After evaluating the system we have developed on the collection of Whitetext project, the same methods are used to draw and analyze the connection graph of PVT brain section. PVT is an important brain part that is believed to have an impact on a wide range of functions such as arousal, stimulation, drug seeking behavior and attention. As the results of our study show, PVT may be a new research focus on brain section behavior evaluation.

Yılmaz İnce (2016) developed the software for evaluating the written exams on the web using natural language processing

methods. In the software, Turkish WordNet is used for semantic relations and Zemberek is used for natural language processing infrastructure. The developed software uses a hybrid model that includes cosine similarity and integrated latent semantic analysis methods. In order to measure the accuracy of the software, a special case study was carried out in the Distance Education Program of the Computer Engineering Department of the Faculty of Engineering at Süleyman Demirel University. According to the results of the study, it was found that the software can be used for Turkish automatic exam scoring with a success rate of 92%.

Emotion analysis aims to classify emotions as positive and negative by analyzing emotions and thoughts about a subject in the texts (Yelmen, 2016). Attribute selection is often used today to improve classification performance and success. Different methods are used in this selection and the aim is to select the most important attributes by disabling irrelevant attributes that affect the success of the classification in the data set. In this way, the success rate can be increased. This thesis focuses on the selection of attributes from Turkish texts written in daily spoken language and uses support vector machines, artificial neural networks and centroid based classification algorithms on detailed preprocessed data. Gini Index, Knowledge Gain and Genetic Algorithm were used as hybrids with 3 different classification algorithms on tweets belonging to the followers of 3 different GSM operators. 100% success was achieved for 3 different GSM operators especially when the support vector machines were used as hybrids with the genetic algorithm which has an important place in size reduction and works intuitively.

Kaya (2018) author recognition process. 120 Turkish books have been studied by 20 Turkish authors. Character n-gram was used as the author's stylometry feature and the classification process was done with Naive Bayes classifier method. First of all, 120 Turkish books were found and converted to txt format. Bi-gram, tri-gram and quadri-gram properties of the authors were calculated by calculating the frequency from the educational books, and the most common 200 were stylometric vector spaces of the author. At this point, our system is ready for author recognition process. Author recognition performance of N-gram vector spaces was measured. As a result of the observations, the bi-gram vector space failed. In addition, tri-gram and quadri-gram gave good results. The best performance yielded quadri-gram with 82% performance. At the end of the thesis, all results, complexity matrix are given.

Keklik (2018) proposes a new rule-based approach for automatic question generation. The proposed approach focuses on the analysis of both the syntactic and semantic structure of a sentence. Although the main purpose of the designed system was to produce questions from sentences, the results of the automatic evaluation showed that the system performed well on paragraphs that required comprehension skills. As for human evaluations, the designed system outperformed all other systems and produced the most natural (human-like) questions.

Ayata (2018) investigated the efficiency of using machine learning and statistical natural language processing techniques in solving emotion analysis problem for Twitter data. The thesis covers Turkish tweet emotion analysis, sector based emotion analysis, English tweet emotion

analysis and predicting political orientation. A framework for sector-specific Turkish tweet emotion analysis was proposed with the combination of machine learning and statistical natural language processing techniques, and the proposed framework was applied to the finance, retail, telecommunications and sports sectors. In the political orientation analysis, it was estimated which voters 'tendencies according to Twitter messages belong to the 'democratic' or 'republican' classes.

Bozyiğit (2019) presents an automatic concept definition model that converts Turkish requirements into a Unified Modeling Language (BMD) class diagram to facilitate the work of the people in the software team and to reduce the cost of software projects. Natural Language Processing techniques were used in the study and a new set of rules containing twenty-six rules was created to find Object-Oriented design elements from requirements. Since there is no data set available to other researchers in the online repositories, a well-defined data set containing twenty software requirements was created in Turkish and made publicly available on GitHub for use by other researchers.

Ay (2019) was written using Python 3.7, a program to model natural language processing and production processes. In the text given as input, the program performs a preprocessing on the text with text processing techniques. It then determines which type of production is cell type, u type, workshop type or mass production type, and accordingly models the process by placing the existing machines. Then the modeled process can be saved / printed or the simulation process can be started via the modeled process. The aim of the study is to provide an initial basis for a program that will provide all the results with only audio narration without the need for any other program for modeling and realization of production processes.

4. CONCLUSION

Natural language processing; It is an engineering field which is concerned with the design and implementation of computer systems whose main function is to analyze, understand, interpret and produce a natural language. In this study, to be valid and reliable documentation of the method of scientific research thesis examined studies on natural language processing in Turkey were examined by scanning pages of the thesis of Higher Education. Evaluated in terms of the subjects of the study samples obtained as a result of the literature review; Morphological analysis studies, syntactic analysis studies, semantic analysis studies and problem analysis criteria were determined and presented.

REFERENCE

[1] Adalı, Ş., 2009. An Integrated Architecture for Information Extraction from Documents in Turkish. Istanbul Technical University, Institute of Science and Technology, Doctorate Thesis, 109, İstanbul.

[2] Agun, H.V., 2008. Doğal Dil İşlemede Çizgisel ve Olasılık Tabanlı bir Otomatik Öğrenme Uygulaması. Trakya Üniversitesi, Fen Bilimleri Enstitüsü, Master Tezi, 59, Edirne.

[3] Akman, Ç.K., 2004. Computer-Aided Transcription Tool. Boğaziçi University, Graduate School of Natural and Applied Sciences, Master Thesis, 75, İstanbul.

[4] Aksoy, E., 2008. HPSG Teorisinin ve Semantik Frame Teorisinin Bir Uygulaması Olarak, Sahne Betimleyen

Doğal Dil Cümlelerinden Görsel Yapılar İnşa Edilmesi. Trakya Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 91, Edirne.

- [5] Aktaş, Ö., 2006. Türkçe için Verimli bir Cümle Sonu Belirleme Yöntemi. Akademik Bilişim 2006 - Bilgi Teknolojileri Kongresi IV, Pamukkale University, Denizli, Türkiye.
- [6] Aktaş, Ö., 2010. Rule-based natural language processing methods for Turkish, Dokuz Eylül Üniversitesi / Fen Bilimleri Enstitüsü, 211, İzmir.
- [7] Aktaş, Y., 2017. Natural language processing based computer network terms using wordnet ontology creation. Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, 73, Isparta.
- [8] Albayrak, N.B., 2011. Doğal Dil İşleme Teknikleri Kullanarak Görüş ve Duygu Analizi. Fatih University, Institute of Sciences and Engineering, Master Thesis, 69, İstanbul.
- [9] Alkım, E., 2006. Türkçe için Doğal Dil İşlemede Biçimbirimsel Çözümleme ve Sözlük Tasarımı için Yeni Bir Yöntem. Dokuz Eylül University. Graduate School of Natural and Applied Sciences, Master Thesis, 81, İzmir.
- [10] Amasyalı, M.F., 2003. A Natural language processing application running on internet : A question answering system, Yıldız Teknik Üniversitesi / Fen Bilimleri Enstitüsü, 66, İstanbul.
- [11] Ay, H., 2019. Manufacturing process modeling with natural language processing. Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, 62, Eskişehir.
- [12] Ayata, D., 2018. Applying machine learning and natural language processing techniques to twitter sentiment classification for turkish and english. Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü, 70, İstanbul.
- [13] Bahadır, Ö., 2007. Aritmetik Problemlerin Çözülmesi ve Simgelenmesi. Istanbul Technical University, Institute of Science and Technology, Master Thesis, 66, İstanbul.
- [14] Birant, Ç.C., 2008. Root-Suffix Separation of Turkish Words. Dokuz Eylül University, Graduate School of Natural and Applied Sciences, Master Thesis", 56, İzmir.
- [15] Bozyiğit, F., 2019. Object oriented analysis and source code validation using natural language processing, Dokuz Eylül Üniversitesi / Fen Bilimleri Enstitüsü, 107, İzmir.
- [16] Çakıroğlu, Ü., 2001. Knowledge modelling by semantic networks in natural language processing. Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 106.
- [17] Çelikkaya, G., 2015. Development of a turkish mobile assistant software using natural language processing techniques, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 93, İstanbul.
- [18] Delibaş, A., 2008. Doğal Dil İşleme ile Türkçe Yazım Hatalarının Denetlenmesi. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 78, İstanbul.
- [19] Eker, Ö., 2009. Parser Evaluation Using Textual Entailments. Boğaziçi University, Master Thesis, 107, İstanbul.
- [20] Er, N.P., 2009. Turkish Factoid Question Answering Using Answer Pattern Matching. Bilkent University, The Institute of Engineering and Sciences, Master Thesis, 155, Ankara.
- [21] Eş, S., 2005. AComputational Analysis of a Language Structure in Natural Language Text Processing. Çankaya

- University, The Graduate School of Natural and Applied Sciences, Master Thesis, 52, Ankara
- [22] Gökdeniz, E., 2016. Natural language processing for mining neuroanatomical relations among brain regions. Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü, 81, İstanbul.
- [23] Görmez, Z., 2009. Implementation of a Text-to-Speech System with Machine Learning Algorithms in Turkish. Fatih University, Institute of Sciences and Engineering, Master Thesis, 63, İstanbul.
- [24] Güngör, T., 1995. Computer Processing of Turkish: Morphological and Lexical Investigation. Boğaziçi University, Doktora Tezi, 185, İstanbul.
- [25] Güvenç, B., 2016. Machine learning methods in natural language processing. Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü, 137, İstanbul.
- [26] Hadımlı, K., 2011. Processing Turkish Radiology Reports. Middle East Technical University, The Graduate School of Natural and Applied Sciences, Master Thesis, 86, Ankara.
- [27] Hallaç, Ü., 2007. Determination of Turkish Word Types. Dokuz Eylül University, Graduate School of Natural and Applied Sciences. Master Thesis, 51, İzmir.
- [28] Kalender, M., 2010. Automated Semantic Tagging of Text Documents. Boğaziçi University, Master Thesis, 107, İstanbul.
- [29] Karadeniz, Z.İ., 2007. Türkçe için Biçimbirimsel Belirsizlik Giderici. İstanbul Teknik Üniversitesi. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 67, İstanbul.
- [30] Kaya, S., 2018. Author-book recognition with natural language processing techniques. İstanbul Aydın Üniversitesi / Fen Bilimleri Enstitüsü, 89, İstanbul.
- [31] Kazkılınç, S., 2012. Türkçe Metinlerin Etiketlenmesi. İstanbul Teknik Üniversitesi. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 63, İstanbul.
- [32] Kazmı, M., 2017. Learning logic rules from text using statistical methods for natural language processing, Sabancı Üniversitesi / Mühendislik ve Fen Bilimleri Enstitüsü, 97.
- [33] Keklik, O., 2018. Automatic question generation using natural language processing techniques, İzmir Yüksek Teknoloji Enstitüsü, Mühendislik ve Fen Bilimleri Enstitüsü, 52, İzmir.
- [34] Kışla, T., 2009. Türkçe için Tümleşik bir Biçim Çözümleme ve Sözcük Türü Tespit Yöntemi. Ege Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, 185, İzmir.
- [35] Köprü, S., 2008. Coupling Speech Recognition and Rule-Based Machine Translation. The Middle East Technical University, The Institute of Informatics, Doktora Tezi, 130, Ankara.
- [36] Kutlu, M., 2010. Noun Phrase Chunker for Turkish Using Dependency Parser. Bilkent University, The Institute of Engineering and Sciences, Master Thesis, 124, Ankara.
- [37] Orhan, Z., 2006. Türkçe Metinlerdeki Anlam Belirsizliği Olan Sözcüklerin Bilgisayar Algoritmaları ile Anlam Belirginleştirilmesi. İstanbul Teknik Üniversitesi. Fen Bilimleri Enstitüsü, Doktora Tezi, 124, İstanbul.
- [38] Özdemir, V., 2009. Word Sense Disambiguation for Turkish Lexical Sample. Fatih University. Institute of Sciences and Engineering, Master Thesis, 63, İstanbul.
- [39] Per, M. U., 2011. Developing A Concept Extraction System for Turkish. The Middle East Technical University. The Institute of Informatics, Master Thesis, 59, Ankara.
- [40] Savaşçı, A., 2010. Türkçenin Bitişkenlik Derecesinin İstatistiksel Verilerle Belirlenmesi. Ege Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 58, İzmir.
- [41] Sevgili Ergüven, Ö., 2018. A systematic evaluation of semantic representations in natural language processing. İzmir Yüksek Teknoloji Enstitüsü, Mühendislik ve Fen Bilimleri Enstitüsü, 83, İzmir.
- [42] Shylov, M., 2008. Turkish and Turkmen Morphological Analyzer and Machine Translation Program. Fatih University, Institute of Sciences and Engineering, Master Thesis, 46, İstanbul.
- [43] Soysal, E., 2010. Ontology Based Information Extraction on Free Text Radiological Reports Using Natural Language Processing Approach. The Middle East Technical University, The Institute of Informatics, Doctorate Thesis, 110, Ankara.
- [44] Şentürk, F., 2009. Biçimbirimsel Bul ve Değiştir. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 129, İstanbul.
- [45] Tahiroğlu, B. T., 2010. Bilgisayar Destekli Sözlük Bilimi Çalışmalarında Derleme Sözlüğü Veri Tabanı Örneği. Çukurova Üniversitesi. Sosyal Bilimler Enstitüsü, Doktora Tezi, 207, Adana.
- [46] Turna, S. E., 2011. Sözcük Türlerinin Belirlenmesi için Türkçe Kelime Yapılarının Tanımlanması. Dokuz Eylül University, Graduate School of Natural and Applied Sciences, Master Thesis, 38, İzmir.
- [47] Tülek, M., 2007. Türkçe için Metin Özetleme. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 88, İstanbul.
- [48] Yalçınkaya, E., 2018. Natural language processing based formal automated review tool for Turkish software requirements. Türk Hava Kurumu Üniversitesi / Fen Bilimleri Enstitüsü, 117.
- [49] Yelmen, İ., 2016. Sentiment analysis with natural language processing methods on Turkish social media data, İstanbul Aydın Üniversitesi, Fen Bilimleri Enstitüsü, 91, İstanbul.
- [50] Yıldırım, Ö., 2005. Improved Handling of Sms Messages with statistical Natural Language Processing Techniques. Boğaziçi University, Master Thesis, 66, İstanbul.
- [51] Yıldırım, A., Şimşek, H., 2000. Sayfa 140. Sosyal Bilimlerde Araştırma Yöntemleri. Ankara. Seçkin Yayıncılık.
- [52] Yılmaz, E. Ç., 2008. A Turkish Natural Language Interface for the Semantic Web: A Case Study on Turkish Universities. Atılım University, Graduate School of Natural and Applied Sciences, Master Thesis, 42, Ankara.
- [53] Yılmaz, S., 2009. Türkçe için İyileştirilmiş Biçimbilimsel Çözümleyici. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 55, İstanbul.
- [54] Yılmaz İnce, E., 2016. Web Ortamındaki Yazılı Sınavların Doğal Dil İşleme Yöntemleri İle Değerlendirilmesi, Süleyman Demirel Üniversitesi Fen Bilimler Enstitüsü, Doktora Tezi, 117, Isparta.
- [55] Zafer, H. R., 2011. A Generic Syntactic Parser for Turkic Languages. Fatih University, Institute of Sciences and Engineering, Master Thesis, 59, İstanbul.