T.C.
EGE University
Institute of Science and Technology

Negation Detection in Turkish Medical Texts with Deep Learning

Zana SÖĞÜT

Advisor: Prof. Dr. Oğuz DİKENELLİ

Department of Science
Computer Engineering

İzmir
January 18, 2025

....................tarafından .... tezi olarak sunulan “Derin Öğrenme ile Türkçe Sağlık Metinleri Üzerinde Olumsuz Anlam Tespiti” başlıklı bu çalışma EÜ Lisansüstü Eğitim ve Öğretim Yönetmeliği ile EÜ Fen Bilimleri Enstitüsü Eğitim ve Öğretim Yönergesi’nin ilgili hükümleri uyarınca tarafımızdan değerlendirilerek savunmaya değer bulunmuş ve ....tarihinde yapılan tez savunma sınavında aday oybirliği/oyçokluğu ile başarılı bulunmuştur.

Jüri üyeleri		**İmza**
Jüri Başkanı	**:** Jüri Başkanının Ünvanı-Adı-Soyadı	
Raportör Üye	**:** Raportör Üyenin Ünvanı-Adı-Soyadı	
Üye	**:** Üyenin Ünvanı-Adı-Soyadı	

EGE UNIVERSITY INSTITUTE OF SCIENCE

DECLARATION OF COMPLIANCE WITH ETHICAL RULES

EÜ Lisansüstü Eğitim ve Öğretim Yönetmeliğinin ilgili hükümleri uyarınca Yüksek Lisans Tezi olarak sunduğum "Derin Öğrenme ile Türkçe Sağlık Metinleri Üzerinde Olumsuz Anlam Tespiti" başlıklı bu tezin kendi çalışmam olduğunu, sunduğum tüm sonuç, doküman, bilgi ve belgeleri bizzat ve bu tez çalışması kapsamında elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara atıf yaptığımı ve bunları kaynaklar listesinde usulüne uygun olarak verdiğimi, tez çalışması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını, bu tezin herhangi bir bölümünü bu üniversite veya diğer bir üniversitede başka bir tez çalışması içinde sunmadığımı, bu tezin planlanmasından yazımına kadar bütün safhalarda bilimsel etik kurallarına uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul edeceğimi beyan ederim.

01/01/2024

İmzası

Zana SÖĞÜT

ABSTRACT

Negation Detection in Turkish Medical Texts with Deep Learning Models

SÖĞÜT, Zana

MSc in Department of Computer Engineering
Supervisor: Prof. Dr. DİKENELLİ Oğuz
01.01.2024, 89 pages

Language allows us to convey the information we observe and interpret it in daily life to another recipient. Being used continuously and so frequently with different forms of expressions makes language a valuable data source for deep learning and machine learning models. Clinical notes provide valuable information for models to be used in tasks such as diagnosing, predicting drug risks and evaluating treatment efficacy. The language we use constitutes an important part of the health and clinical data that is generated on a daily basis and in large numbers. Understanding the meaning components of the language used in these notes, such as negation and ambiguity, is critical for the semantic awareness of corpora.

Besides being such an important resource, language also has processing difficulties. Due to the fact that sentences can be semantically dense and structurally long, use phonetic expressions such as gestures, icons, grammatical errors, etc., resources cannot be used efficiently in machine learning tasks. Negation is one of the most complex and important feature of the language. It can directly affect meaning by expressing the way, absence or degree of partial or total rejection of a proposition. Negation is one of the most complex and important components of language. It can directly affect meaning by expressing absence or degree of partial or total rejection of a proposition. In recent years, research on negation and its features has been studied in many languages. The large amount of text data generated in the field of healthcare daily. But an important part of it may not constitute a healthy digital data source for machine learning. In order to extract knowledge from the data, the data needs to be processed or enriched manually by trained experts or with auxiliary autonomous models.

There is currently a lack of comprehensive study on negation in Turkish health texts. Interdisciplinary research is needed to address the complexity of the subject and the challenges posed by the language resources used in the compilation. This study explores deep learning models that can enrich health data semantically with additional information. The dataset consists of 20 case reports published in 12 medical disciplines and includes annotation for negation, scope, and event labeling.

Different methods found in the literature and encountered during the research were applied in the labeling section and component models. Labeling methods conveys 3 different methods which were applied before model was put through a training process. The suitability degree of labeling methods discussed regarding to the scores. For the scope and event models, negation cue labels were vectorized and added to the end of the embedding vectors of tokens. The highest F1 score is 94.4% which taken by using first token tagging method in negation cue prediction task. 71.16% and 68.77% F1 score recorded as highest respectively. For the scope and event models, the highest metrics obtained with continuous tagging and same tagging methods. Effects of labeling techniques and linguistic challenges the results evaluated for the dataset.

Keywords: Negation, Negation Extraction, Negation Detection, Deep Learning, Transformers, Information Extraction, Medical Corpora.

FOREWORD

Ode to my grandmothers...

While our first conversations with my mentor, I was aware of the importance of language and my desire to do research on NLP. However, I remember that when I first heard the concept of negation from my mentor, I was very surprised and thought about what research on this topic could contribute. As our research progressed on the grammatical side, we realized that the importance of the concepts of negation, ambiguity and its component terms was critical. Over time, it became evident that these terms, which we might initially view with hesitation or negativity, hold more meaning than just bad or negative connotations.

Negation is different from the negative polarity of meaning, which usually indicates a bad situation or event. It is an expression that allows us to convey and understand our feelings and thoughts within a logical framework. Of course, seeing that these expressions can lead to good or bad results with a little bit of human use, as same with everything that is human made which lead more curiosity about these language properties.

The use of artificial intelligence models and decision-making mechanisms in practical areas of medical practice is increasing. However, in order to use artificial intelligence in practical areas, high criteria must be met. For this reason, it is important that NLP systems fully understand the meaning. When we came across articles examining the processes of understanding negation expressions in human and some animal species in adults and children, the importance of machine understanding of the semantic expressions of language was also supported. Indeed, although written and audio resources are widely used in medicine, the scope and focus of work to reduce complexity in most languages is still ongoing.

During the thesis, the fact that the need for research in this direction is still ongoing in many languages has been my motivation for the necessity of this research and what it can contribute. I believe that a better understanding of these features of language will contribute a lot to us and to large models in artificial intelligence research. The NLP group has been working for almost 2 years. I would like to specially thanks to my mentor Prof. Dr. Oğuz Dikenelli and everyone I have worked with in this process in which I have gained valuable achievements.

İZMİR
01/01/2024

Zana Söğüt

TABLES of CONTENT

Page

LIST OF FIGURES

TABLES

Figure **Page**

ABBREVIATIONS

Abbreviations	Explanation
AI	Artificial Intelligence
BERT	Bidirectional Encoder Transformer
BiLSTM	Bidirectional Long Short Term Memory
BIO	Begin-Inside-Out
BoW	Bag of Word
CNN	Convolutional Neural Network
CRF	Conditional Random Field
DANN	Domain Adversarial Neural Network
EPAM	Extra Propositional Aspect of Meaning
FFNN	Feed Forward Neural Network
FN	False Negative
FP	False Positive
GPT	Generative Pre-trained Transformers
HMM	Hidden Markov Model
IES	Information Extraction System

ABBREVIATIONS (cont.)

Abbreviations	Explanation
IR	Information Retrieval
$LSTM$	Long Short Term Memory
NER	Name Entity Recognition
NLP	Natural Language Processing
P	Precision
PoS	Part of Speech
R	Recall
RNN	Recurrent Neural Network
SCL	Structural Correspondence Learning
SVM	Support Vector Machine
$TF-IDF$	Term Frequency - Inverse Document Frequency
TN	True Negative
TP	True Positive
UDA	Unsupervised Domain Adaptation
$ULMFiT$	Universal Language Model Fine Tuning
WE	Word Embedding

1 INTRODUCTION

Language enables us to convey the information that we observe and sense in daily life to another receiver Buroker (1993). It also contributes, sometimes shape our thoughts and knowledge as we shape it, while our monologues and thinking processes. These are some strong reasons that make language an important part of our lives. Human language has been evolving with great momentum since 150.000-200.000 years Perreault ve Mathew (2012). It includes many features from biological, sociological and technological evaluations. These features are interacting with language over complex logical, morphological, semantic, phonetics and syntactic evaluations since the beginning of our journey. It is important to note that these features can be the contributors to the aesthetic and colorful differences between languages. They additionally make similarities of structures among languages which can also be linked to anatomical, neurological and psychological relevance, as (Pagel, 2017) states.

Texts contain a lot of information, propositions, actions and concepts. Text data can be complex and densely meaningful by linking various contextual expressions and phonetic structures. In addition, healthcare texts contain many medical terms and written ontological information describing the relationships between these terms. Therefore, these texts, which contain the language used in the transmission of healthcare knowledge, have become an important data source for deep learning and machine learning models. However, the complexity of language makes it difficult for these text data to be a meaningful and understandable data source for machines. This broad research area has been studied in many different scientific disciplines. Natural Language Processing (NLP) is an interdisciplinary field that uses methods from computer science, artificial intelligence and computational linguistics to capture and process the meaning in language (Hao ve ark., 2021). After recent advances, NLP has become one of the main tools to study in many health research fields. There is a huge amount of data produced in biomedical which is making NLP important tool for biomedical studies. (Huang ve Lu, 2015) underline the rapid growth of biomedical literature with more than 3000 articles published daily in 2015. If we add the fact that in 2009, HITECH provided $30 billion in incentives for hospital and physician practices to digitize records, we can easily understand the importance and growing interest in information extraction. (Shickel ve ark., 2018)

There are numerous studies on negation to improve context awareness of the AI models with NLP techniques. Negation processing has become a crucial requirement for the accurate performance of medical natural language processing applications. As an illustration, the study by (Mukherjee ve ark., 2017) highlighted that effective doctor-patient communication could be enhanced by avoiding using negation. This indicates that the use of negation increases confusion and it is critical for machines to be aware of negation if there is any. This hedge has become a very important input for health applications (Vincze ve ark., 2011), especially for Information Retrieval (IR) applications. (Morante ve Sporleder, 2012) discussed the traditional approaches in NLP which are focused on propositional aspects and underlined that the extra-propositional aspects are equally important to understand the meaning more accurately. He defines Extra-Propositional Aspects of Meaning (EPAM) to be actuality, uncertainty, opinions, beliefs, intentions or subjectivity and underlines the fact that modality and negation usually play a major role in these extra-propositional meaning of aspects. DI++ (Shi ve ark., 2022), which was implemented in collaboration with the industrial partner of the authors, is one of the best practical examples for the feature extraction systems. This tool is designed to identify the patient's condition based on an analysis of clinical notes. In the study, negation was employed to avoid identifying non-existent diseases or medical terms which were introduced with negation in the clinical report. This example shows the key role of negation for such real life tasks in medical science.

The primary focus of this thesis is to explore the crucial role of negation and understanding the extra-meaning aspects present in medical domain. In such documents, words are strongly related to health and biological and psychological knowledge. This study is a part of a bigger project which is led by NLP research team of Ege University. The project consists of three main parts which are collecting medical reports; pre-process and annotating these reports by using Python libraries and Inception Annotation tool; post processing and measuring the performance of the state-of-art models. Second part of the project is driven by the collective work of experts in Linguistic and Computer Engineering departments. First and third parts will be presented in this study and the outputs of the project will be extended in a bigger frame of approach later. This study focuses on Transformers models, known for their proficiency in encoding input sequences. Transformers represents diverse features and demonstrates enhanced efficiency compared to traditional sequence models in many research area. In addition to investigating the power of DL models, this thesis also incorporates the utilization of Name Entity Recognition (NER) techniques for the pre-processing stage. Employing NER, it is aimed to categorize words within relevant contexts by tagging words effectively based on the scope and event of negation.

In the next chapter, the study addresses several complex challenges primarily stemming from the complexities of sentences and their long sequence nature when describing health issues. Additionally, the frequent use of gerunds forms along with conjunctions poses significant difficulties in identifying the scope and event of negation. Throughout the challenges section, we will exemplify and analyze these intricate linguistic aspects to shed light on their significance in health-related contexts.

The concise overview of the thesis structure is as follows. Chapter 1 offers a brief history of negation and presents a literature review encompassing both linguistic and AI perspectives on negation. In Chapter 2, we delve into fundamental concepts that underpin the philosophy of negation and its application in the Turkish language. Additionally, this chapter includes a review of traditional approaches and introduces state-of-the-art deep learning models, which will be used to detect negation along with its complementary concepts. Challenges associated with negation identification in Turkish are discussed in this chapter as well. Chapter 3 describes the dataset project and corpus preparation processes. Chapter 4 describes the experimental process involving the dataset and models utilized in our research. In Chapter 5, experimental results and comparisons are shared. This last chapter also revisits the challenges encountered during the research by discussing the results. Chapter 5 also provides an encompassing discussion of the thesis work and suggests potential avenues for future research.

2 LITERATURE REVIEW & HISTORY

Negation is one of the core concepts in NLP. It is one of the fundamental knowledge feature for sentiment analysis, opinion mining, question-answering systems and accurate information extraction. Regarding the interdisciplinary concept in this study, we can approach negation history from Linguistics and NLP studies perspectives.

2.1 Negation Review in Linguistics

First known conceptual usage of negation mentioned by the philosophers Socrates and Plato in their dielectric dialogue during the age of Ancient Greek (Speranza ve Horn, 2010). Socrates, in his famous "Socratic Method", used negation to dissect arguments and reveal inconsistencies. Plato, in his dialogues, discussed negation concepts especially in “The Sophist" and rarely in others. In the Middle Ages, the study of grammar and rhetoric flourished. According to (Corrigan, 2023), the Areopagite analyzed negation in theological contexts in his study “On the Divine Names”. It is important to note that Areopagite’s approach on negation is mostly grammatical and focused on negative structures in the sentences. After this period, the next comprehensive study was conducted by the Port-Royal grammarians Lancelot and Arnauld in the 17th century. In their influential work "Grammaire générale et raisonnée" explored the syntax and semantics of negation. They identified negation as a crucial component of language, essential for conveying precise meanings and making logical distinctions. Their work, along with other valuable studies by Leibniz and others, lead the foundation of Cartesian grammar (Buroker, 1993).

Lately structuralist linguistics and pragmatic linguistics has been prepared as a basis for today’s negation concept. (Chomsky, 1957) presents transformational-generative grammar for negation analysis. As an important study for psychology and linguistics, (Giora ve ark., 2005) (Beukeboom ve ark., 2019) examined the "Negation Bias" phenomenon and its implications in communication. For 20th and 21th linguistic studies, (Horn, 1989) gives remarkable works for the history of negation. After this period, negation studies have been spread to many languages with different aspects and there are quite many long lists which are investigating negation in many aspects. All of these studies have enriched the content of negation with pragmatic, morphological, syntactic and lexical manners. Also the acquisition of language has started to be reviewed lately (Batet ve Grau, 1995) (İnci Kavak, 2019) (Tagliani ve ark., 2022).

The concept of negation has gained considerable attention by linguistic researchers in Turkish language, particularly after significant changes were implemented during the new republic. While the focus of negation in Turkish has traditionally centered on affixes, studies aimed to understand its various dimensions, both in linguistics and computational linguistics in recent years. Due to the referenced studies given here, one of the early studies can be attributed to an unpublished work by (Tura, 1981) titled "A Study of Negation in Turkish". This study directly addresses the negation and used in academic studies (Emeksiz Erk, 2010). Later on, (Erguvanlı, 1984) pointed out the affixes related to negation syntactically, semantically, and pragmatically. In the 2000s, (Kelepir, 2001) published thesis and investigated clausal structures and scope in Turkish syntactically. (Emeksiz Erk, 2010) provided a brief explanation of general negation rules concerning negation grammar and syntactic-semantic relations. Although it does not touch on scope or focus, it can still be considered a valuable study for Turkish in NLP. (Ince, 2011) discussed how negation understood by children, and in the same year, (Görgülü, 2017) studied negative polarity, contributing to the understanding of the interaction between negation and polarity-sensitive elements. (İnci Kavak, 2019) can be given as an example of morphological research, discussing negation under the manner of acquisition. (Seydi, 2020) researched existential negation, providing some insights into the scope of negation rules. As a final example, guide of the used dataset in this study can offer a guideline. This paper includes a comprehensive review and insights on negation, as well as methods for extracting features like Scope, Focus, and Event. For Turkish, negation has been well-defined in grammatical definition, but further discussions are needed on the linguistic rules for scope and focus ambiguity. Additionally, incorporating domain-specific rules could improve precision in detection task for these features. This approach would enable models to capture semantic, pragmatic and morphological insights more effectively.

2.2 Review of Negation in Dataset

Negation detection tasks can be viewed under two categories. The first one is dataset preparation which is heavily bound with linguistic topics. The second category is the rule-based systems, machine learning and deep learning methods and their review. This section briefly describes the historical background and evaluation of studies with other statistical results for both dataset and negation detection studies.

The model learns with data. For this reason, a dataset is one of the most important components of machine learning and deep learning. Rich and more structured data certainly needs a review by experts. (Emilie, 2018) By this, systems can have ability to generalized, inner diversity, quantity and quality. It also needs careful research with interdisciplinary fields if it is necessary. In their valuable review (Morante ve Sporleder, 2012), categorized the corpora of negation studies into two types:

❄ The BioScope (Vincze ve ark., 2008) and ConanDoyle negation shared task (Farkas ve ark., 2010) can be given as the main stone negation corpora in NLP research. Also in this category, the first division is made between event-based and scope-based annotation which is discussed in (Messikommer ve ark., 2021). After this discussion, the studies started to include this bridge between two approaches and support full features of negations in this category. (Please check Table 2.1)

❄ Second category can be viewed as two subgroup. Those that represent negation in a logical form by using quantifiers, predicates, and relations; and those that use string-level annotations to mark negation information such as cues, scope, negated events and focus. The Groningen Meaning Bank (Bos ve ark., 2017) and the DeepBank corpus (Flickinger ve ark., 2012), can be given as examples of the studies of the logical form.

A comprehensive literature review, (Mahany ve ark., 2022), highlights the growing interest in negation processing and presents example datasets developed for different languages such as English, Spanish, Swedish, Dutch, Japanese, Chinese, German, Italian, Dutch, Japanese and Italian. Some of these studies are limited to sign and scope detection.The authors have comprehensively presented a collection of datasets categorized according to the features and domains supported by their literature review.Table 2.1 lists the datasets and their feature properties.

BioInfer (Pyysalo ve ark., 2007) could be given as an example of a pioneer study in biomedical and also other categorical dataset studies. It is a corpus that contains 1100 sentences from abstracts of biomedical papers annotated with negation, entity, relationship, and dependency tags. The GENIA Corpus (Kim ve ark., 2008), initially annotated with part of speech and syntactic trees, was later expanded to include negated biological events and uncertainty. It provides 9372 sentences for the study of negation's role in the biomedical domain. The BioScope corpus covers negating and speculative cues and their scope in diverse biomedical texts, offering valuable insights into negation's impact on clinical radiology reports, scientific papers, and abstracts. The difference of BioScope is that it includes the scope feature which can be assigned as an improved conceptual term that can be used for the context feature studied to be identified by early rule-based negation detection systems (NegEx (Chapman ve ark., 2001), NegFinder (Mutalik ve ark., 2001) etc.).

Early work on annotating negation mainly targeted the biomedical domain but later expanded to product reviews, short stories, and dialogues (see Table 2.1). The CoNLL 2010 Shared Task (Farkas ve ark., 2010) focused on identifying speculation cues and scope in biological publications and Wikipedia articles. In the realm of product reviews, the Product Review Corpus (Councill ve ark., 2010) became the first to be annotated for negation, containing 2111 sentences from reviews extracted from Google Product Search. The PropBank Corpus (Blanco ve Moldovan, 2011) introduced the concept of negation focus, contributing valuable annotations to sentences outside the biomedical domain.

Additionally, the SFU Review Corpus (Jiménez-Zafra ve ark., 2017), consisting of 400 documents across various domains, was annotated for negating and speculative cues, as well as linguistic scope in sentences. The ConanDoyle-neg Corpus explored negation in Sherlock Holmes short stories, annotated with negating cues, scope, and event information. The Twitter Negation Corpus contains tweets annotated with negation cues and scope, offering insights into negation usage in social media.

Table 2.1: Negation Dataset

Year	Dataset	Language	Field	Content	Negation	Scope	Event	Focus	Available
2007	BioInfer	English	Biomedical	1100	✓	✓			✓
2008	GENIA	English	Biomedical	9372	✓		✓		✓
2008	BioScope	English	Biomedical	20,924	✓	✓			✓
2010	Stockholm EPR	Swedish	Clinical	6740	✓				
2010	CoNLL-2010	English	Wikipedia	40,289		✓			✓
2010	Product Reviews	English	Comments	2111	✓	✓			
2011	PropBank FOC	English	Stories	3779	✓			✓	✓
2012	SFU Review	English	Reviews	17,263	✓	✓			✓
2012	ConanDoyle-neg	English	Short Stories	4423	✓	✓	✓		✓
2014	hUnCertainty	Hungarian	Misc.	15,203					
2014	Review and News	Japanese	Review and News	2147	✓			✓	✓
2014	EMC	Dutch	Clinical	12,888	✓		✓		
2015	CNeSp	Chinese	Reviews and Financial Articles	16,841	✓	✓			✓
2015	Twitter Negation	English	Tweets	4000	✓	✓			
2016	EMR	Chinese	Biomedical	36,828	✓	✓			
2016	GNSC	German	Biomedical	2234	✓	✓	✓		
2016	BioArabic	Arabic	Biomedical	10,165	✓	✓			

Table 2.1 cont. next page

Table 2.1 cont. on next page

Year	Dataset	Language	Field	Content	Negation	Scope	Event	Focus	Available
2016	DT-Neg	English	Dialogs	1088	✓	✓		✓	✓
2017	IULA	Spanish	Biomedical	3194	✓	✓			✓
2017	UHU-HUVR	Spanish	Clinical	8412	✓	✓	✓		
2017	SFU Reviews	Spanish	Reviews	9455	✓	✓	✓		✓
2017	Fact-Ita Bank and Tweets	Italian	News, Tweets	1591	✓	✓	✓	✓	
2018	NegPar	English- Chinese	Short Stories	5520	✓	✓	✓		✓
2018	SFU SOCC	English	Reviews	1043	✓	✓		✓	✓
2019	ESSAI	French	Medical	6547	✓	✓			
2019	CAS	French	Medical	3811	✓	✓			
2020	REBEC	Portuguese	Clinical	3228	✓	✓		✓	
2020	Clinical reports	Portuguese	Clinical	9808	✓	✓		✓	
2020	NUBES	Spanish	Biomedical	29,682	✓	✓	✓		✓
2020	NewsComm	Spanish	Reviews	4980	✓	✓	✓	✓	✓
2021	T-MexNeg	Mexican	Tweets	13,704 tweets	✓	✓	✓		✓
2021	ArNeg	Arabic	Wikipedia	6000	✓	✓			
2024	MTNeg	Turkish	Medical	3000	✓	✓	✓	✓	✓

Furthermore, the DeepTutor Negation (DT-Neg) Corpus (Banjade ve Rus, 2016), focused on negation phenomena in dialogue-based systems, consisting of tutorial interactions between high-school students solving physics problems. The SFU Opinion and Comments Corpus (SOCC) (Kolhatkar ve ark., 2019) encompasses opinion articles and comments, annotated with negation cues, scope, focus, and "xscope" annotation. These annotated corpora have significantly contributed to our understanding of the scope of negation across different linguistic domains, enabling researchers to develop better NLP models for accurately processing and interpreting negation in various text types. Table 2.1 shows a review of dataset studies. (Çöltekin ve ark., 2022) made a review on some Turkish corpora. Although there are many studies working to produce structured dataset, this task still needs more attention and a number of studies in the negation perspective. In this review, the closest task topic for thesis study, (Vural, 2013) translated SentiStrength (Thelwall ve ark., 2010) algorithm to detect lexical negation while translating English dataset into Turkish and marked the lexical negation. Also for the NER task, agreeing with (Ozcelik ve Toraman, 2022) after a quiet research for the literature, there is a lack of comprehensive study in Turkish language. Existing studies are mainly taken from structural sources as news articles (Aras ve ark., 2021).

2.3 AI Methods

Beyond the dataset studies, as a next step, the produced rich texts are fed into models to learn how to extract information or features from these dataset. This leads to produce autonomous annotation tools or new inputs for bigger decision maker models. In negation tasks, examples are generally tried to find a relevance path by using grammar and regex rules and finds a narrow scope with the dependency parsing and path finding methods. Rule-based systems are generally able to find solid relation between negation cues and words affected by negation in a window size. Narrow scope problems are often linked to the lack of sequence tracking beyond short windows and the absence of memory architectures in models that can capture insights from long sentences.(Pröllochs ve ark., 2020) also point out that these rule-based systems can not cope with the implicit negation and domain-specific (biomedical, cultural etc.) characteristics of negation. Authors discussed that these lack of rule-based systems are partially overcome with the applied machine learning techniques. This effect of achievements led to research are spread out with feature selection like scope, event, focus for negation domain. (Morante ve Sporleder, 2012) shows a quite brief list for the studies that are researching these features in machine learning and deep learning methods. Also in their study (Mahany ve ark., 2022) showed that the accuracy and the precision has out-performed the machine learning methods by addressing the context awareness and resolution window problems with the help of the studies which increase the corpus for models.

2.3.1 Rule-Based Methods

NegExpander (Aronow ve ark., 1999), NegEx, DeepNeg, NegFinder and NegBio are some of the best known pioneering works in rule-based systems (Mahany ve ark., 2021). Another common point of these studies is that they were conducted in the healthcare field. Some of these models developed as an update of NegEx system which are shown in Figure 2.1.

These systems scan text with grammatical or word/suffix-based flags and triggers, usually defined with Regex features (Slater ve ark., 2021). In NegExpander, an algorithm is reported in this study to solve the problem of finding negated UMLS, coding standard for EHRs, by constructing conjunctive phrases. (Chapman ve ark., 2001) applied a simple regular expression algorithm called NegEx uses trigger terms to detect a list of UMLS semantic vocabulary which refers to negation with grammatical negation phrases. (Huang ve Lowe, 2007) applied a hybrid system enforced by a regular expression mechanism and a grammatical parser. By this approach, they were able to identify the negation and its concept even if they had a distance to the expression on their corpus (Kim ve ark., 2008). Later on ConText (Harkema

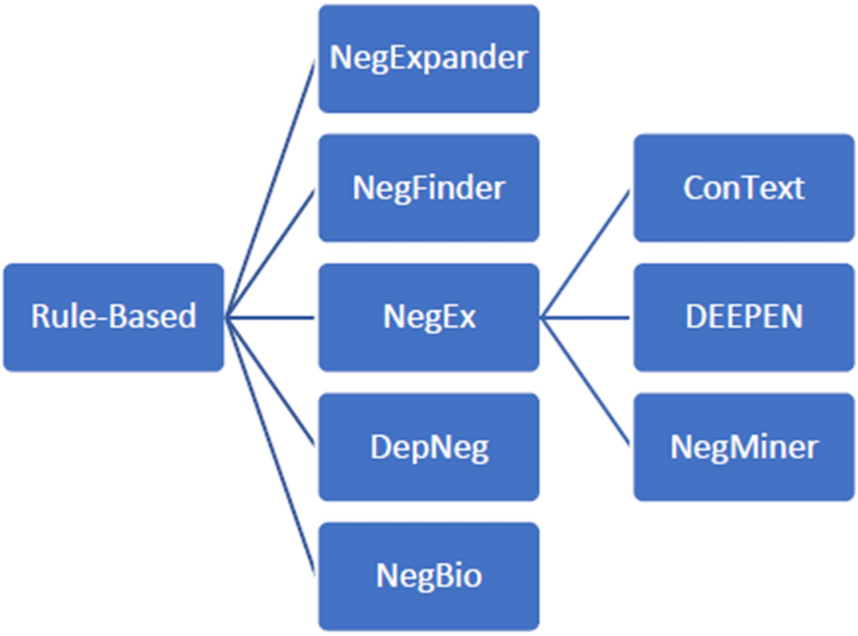


Figure 2.1: Pioneer Rule Based Methods (Mahany ve ark., 2021)

ve ark., 2009), derived from NegEx, covers more objectives as temporal, and the subject concerned by this information in the clinical texts.

Rule-based method densely studied in English and some negation extraction tools have been developed as NegEx and DEEPEN. Rule based methods also applied in ML and some DL application to train models or pre-process data to produced negation annotated dataset. Also rule based methods can be a good helper to boost system accuracy by making heuristic models and define some rules on deep learning and ML models.

2.3.2 ML Methods

Negation study in NLP generally depends on classifying and labeling the sequence of input. This approach is also indicated in this study. (Singh ve Paul, 2021) summarized the most used algorithms as Support Vector Machine (SVM), Hidden Markov Model (HMM) and Conditional Random Field (CRF). SVM is a machine learning model based on linear methods, has demonstrated strong performance in various classification tasks, making it the default choice for identifying negation cues. Respect to this study, when it comes to detecting the scope of negation, two algorithms, Hidden Markov Model (HMM) and Conditional Random Field (CRF), have consistently produced the best results in sequence labeling tasks. They also underline that both HMM and CRF, which are based on Markov chains, often struggle with handling long sequential dependencies. (Morante ve Daelemans, 2009) use SVM and CRF to detect scope of negation and (Althari ve Alsulmi, 2022) point out that these methods are good at cues but take an average performance for scope.

Addition to this brief mention, (Blanco ve Moldovan, 2011) shows a successful implementation for IE task of negation by using Decision Trees to identify focus. They use well defined rules with the help of role and PoS tagging used in ProBank. This study is also important to show how extra information can help to identify the features of negation and NLP tasks. It is valuable to note that early and mid neural network and deep learning techniques show close and high precision results in the medical corpora which are well prepared and has extra knowledge. (Sykes ve ark., 2020) is can be given as a good example for this phenomena which supported with results and comparison in the same study on radiology reports. They are also discussed, this can be mainly related with the corpora difference, less complexity of the sentences and hard working effort can improve results in biomedical corpora (see Chapter 7).

In Turkish corpora, this part is mostly studied with supervised learning methods. Dataset is enriched for decision trees and is generally used for sentiment analysis to boost accuracy in Turkish corpora. (Eroğul, 2009) explores the use of surface linguistic features such as part-of-speech tags, word unigrams and bigrams and negation markers. Corpus fed into the polarity spanning process. In his study, it is also underlined that the bag-of-word does not contain a word order and this causes loss of significant data. This work relies on a morphological analyses for Turkish, called "Zemberek". Later on (Dehkharghani ve ark., 2015) are studied sentiment analysis and use negation handling with bag-of-word embedding before sentiment analysis. They occurred one of the highest records in literature comparisons. Also in another parallel study, which produced a lexicon polarity resource, SentiTurkNet, (Dehkharghani ve ark., 2015) used negation detection to decide polarity scores.

Table 2.2: AI Models in Negation (Adapted from (Mahany ve ark., 2021) .

Yıl	Language	Dataset	Input Parameters	Method	Reference
2009	English	BioScope	PoS, Lemma, Syntactic	SVM, CRF, TiMBL	(Morante ve Daelemans, 2009)
2010	English	BioScope	PoS	CRF	(Councill ve ark., 2010)
2013	English	BioScope	Syntactic	SVM Tree Kernel	(Councill ve ark., 2010)
2016	English	BioScope	Syntactic	CNN	(Qian ve ark., 2016)
2016	English	SFU Corpus	WE	BiLSTM	(Lazib ve ark., 2019)
2016	English	Conan Doyle	WE, PoS	BiLSTM	(Fancellu ve ark., 2016)
2017	English	Connan Doyle	Lexical, Syntactic	SVM+CRF	(Enger ve ark., 2017)
2017	Chinese	Chinese Clinical Texts	WE, BoW, BoC	CRF	(Kang ve ark., 2017)
2017	English	Wikipedia	Lexical Lookup	ML Tech.	(Mukherjee ve ark., 2017)
2018	English, Çince	SEM2012 CDS-CO	Latent Structural, Cue	Semi-CRF	(Li ve Lu, 2018)
2018	Spanish	SFU Review	WE, PoS	LSTM	(Fabregat ve ark., 2018)
2018	English, Çince	CNeSp	PoS	BiLSTM	(Fancellu ve ark., 2018)
2019	English	Bioscope, SFU, Connon Doyle	Cue	NegBERT	(Khandelwal ve Sawant, 2020)
2019	English, Chinese	Bioscope, NegPar	BERT, PoS	BiLSTM	(Sergeeva ve ark., 2019)
2019	French	ESSAI, CAS	WE, PoS	BiLSTM-CRF	(Dalloux ve ark., 2019)
2020	English, Chinese	Bioscope, CNeSp	Syntactic	RNN-CRF	(Fei ve ark., 2020)
2020	English	Bioscope	Syntactic	BiLSTM & CNN	(Lazib ve ark., 2020)
2020	English, French	Connan Doyle	Syntactic	STRNN	(McKenna ve Steedman, 2020)
2020	Brazilian, Portuguese	ESSAI, CAS	WE, PoS	BiLSTM-CRF	(Dalloux ve ark., 2021)
2020	English	MiPACQ, i2b2 Medical	WE, Domain Adaptation	BERT	(Lin ve ark., 2020)
2021	Arabic	KSUCCA, Wikipedia	WE	BiLSTM	(Mahany ve ark., 2021)
2021	Mexican, Spanish	T-MexNeg Corpus	PoS	CRF	(Bel-Enguix ve ark., 2021)

Table 2.2 devamıdır.

Year	Language	Dataset	Input Parameters	Method	Reference
2021	English	MIMIC-III Clinical Notes	WE, UM	ASGD Weight-Dropped LSTM	(Grinde ve Johansen, 2021)
2021	English	MIMIC-III, HCM Dataset	Syntactic	Dependency Graphs	(Slater ve ark., 2021)
2021	English, Japanese, Chinese	WMT2021	WE	NMT models	(Trinh, 2021)
2021	English, Spanish	IULA, NUBes, ESSAI, CAS	WE	BERT	(Hartmann ve Søgaard, 2021)
2023	German	Erasmus Medical Center Dutch	Rule-based, WE	ContextD, BiLSSTM, RoBERTa	(van Es ve ark., 2023)

Main cons of ML techniques can be addressed the high dependency of afforded engineering features and bias or relativity factor of annotation process. Also domain expert dependency ratio is high.

2.3.3 DL Methods

Negation has been studied with nearly all historical components of deep learning models. Negation with scope feature has been tried to be identified by CNN (Qian ve ark., 2016) and (Argüello-González ve ark., 2023), RNN (Ren ve ark., 2018), Bi-directional LSTM , Bi-directional LSTM + CRF (Dalloux ve ark., 2021), Feed-forward and Bi-LSTM models (Fancellu ve ark., 2016). Later (Fancellu ve ark., 2017)used their model with different language corpus and discussed the result. (Chen, 2019) fed embedding vectors of sentences into the BiLSTM layer and apply attention at the end of the chain to detect the hiding important relation of target concepts. The improvement of self attention mechanism and later on BERT has shown significant improvements for general NLP tasks with also negation tasks. (Khandelwal ve Sawant, 2020) in their study, NEGBERT, implements BERT mechanism on 3 known datasets which are Bioscope, the Conan Doyle Stories and SFU Review Corpus. They achieve F1 score on scope resolution of 92.36% on the Sherlock dataset, 95.68% on the BioScope Abstracts subcorpus and show that BERT outperform previous state-of-art models by taking 90.95% score. (Pabón ve ark., 2022) studied Bi-LSTM & CRF and BERT on NUBES and IULA datasets to identify negation and its aspects, two public corpora for the Spanish language (F-score of 92% for negation).

The performance of the model is highly dependent on corpus properties and task identification, it is supported by these studies that latest improved models are out-performing or as good as the early ones. Beyond this observations, (Sykes ve ark., 2020) pointed out that well prepared, tagged and enriched datasets shows good results with early methods, parallel to latest deep learning model. This studies proves that the performance of the model strongly correlated with new model achievements and dataset properties. (Mahany ve ark., 2021) and (Devlin ve ark., 2019) underlined the effect of BERT models to address the lack of resource problem. In these studies, it has also been reported that datasets without detailed data engineering features still perform good measurement values with transfer learning mechanism. Transfer Learning has also been shown in the literature to be useful for systems that are aware of language components such as negation and ambiguity. Researchers in (Pabón ve ark., 2022) performed well in identifying negation signs and their scope. They trained models using the NUBES dataset. Then, they applied transfer learning to a Cancer dataset to improve performance. Table 2.2 provides a comprehensive summary list of the datasets done so far depending on their labeling techniques.

(Görgülü, 2017) studied negative polarity and its interaction with other language elements using deep learning models in Turkish language research. This neural network-based approach achieved high precision in detecting complex patterns in Turkish negation. Despite progress in negation detection, challenges and gaps in the literature on Turkish negation remain. Moreover, issues such as domain-specific markup, incomplete or unclean datasets, spelling errors, and grammatical mistakes complicate the learning of language elements like negation in Turkish.

3 FUNDAMENTALS

3.1 Negation

Negation is a universal linguistic phenomenon that transforms expression into the opposition. By referring to (Horn, 1989) definition, this transform functions could be denial, contradiction, misrepresentation, and irony. Negation is in the first place a phenomenon of semantic opposition. This relation may be realized syntactically and pragmatically in various ways. Furthermore, (Siegel, 2007) in their study refers that certain studies indicate that apes and, surprisingly, even non-primates, can be taught like rejection, refusal, and non-existence. Also this study underlines the fact that these concepts align with stages observed in the development of negation understanding in children, but they do not encompass the concepts of denial or truth-based negation.

Negation occurs frequently in literature and it is a foremost challenge in biomedical (Nawaz ve ark., 2013). The change of meaning has a great importance to annotate the corpus for decision making, categorization and feature extraction models. In their study, (Vincze ve ark., 2008) reported that 13% of the corpus sentences are found to be related with negation in biomedical texts. (Vincze ve ark., 2011), pointed out that in translation, the scope of negation will be very important to translate phrases with correct semantic interpretation. This study is prepared to seek an improvable, adaptable model that can be used as an automatic annotation tool for biomedical texts which will reach the corpus with knowledge. It is also valuable to note here that although the correlation between negation and negativity may seem obvious; it is not always the case that negation expresses a negative opinion (Emilie, 2018). Negated statements may carry an implicit positive meaning (Morante ve Sporleder, 2012).

It can be distinguished between two primary forms of negation in language: complement (or local) negation and clausal (or sentence) negation. This generally can be referred to the scope of the negation (Klima, 1964).

❊ Clausal negation negates an entire proposition or statement:

[Eski tekrarlayan lezyonlara ait skar dokusu ise *izle*n**me**di].

❊ In contrast, complement negation is associated with a specific constituent or clause within a sentence:

[Pnömokok aşısının yapıl**ma**dığı] öğrenildi.

This categorization is structurally correct yet it is obvious that the same sentence can be categorized in different groups for Turkish and English. This is caused by the linguistic properties of languages. To be given an example:

He has **no** [*money*].
Here the negation is constituent but when we translate it into Turkish:

[Onun hiç *parası* **yok**].
Here negation became clausal as it effect verb.

(Diaz, 2014) summarize the other classification of negation by studies as follow:

❊ Totti defines 5 types of negation in clausal negation as Denials, Rejection, Imperatives, Questions and Support and Repetitions (Van Ostade ve ark., 1999).

❊ Payne defines different types of clausal and constituent negation in any language. This comprehensive division of types is also accepted by our dataset (Diaz, 2014).

In the Turkish dataset, the researchers followed the category format defined in (Pabón ve ark., 2022), which originates from Payne's work.

- ❋ Lexical Negation : The concept of negation is semantically part of the negation layer, with the incorporation of a specific verb or additional verb.
- ❋ Morphological negation: It is a form of negation made with affixes in phrases and structures such as verbs, verb phrases, etc.
- ❋ Syntactical Negation: Here the negation marker acts directly as an adverb or an element. In Turkish, only *"ne ... ne ..."* phrase has been categorized in this category.

Negation cues are the trigger components where it defines a negation. Cues can be expressed by a single syllable, word or compounds. These signs can consist of a prefix, a suffix, a single word or more than one word. (Emeksiz Erk, 2010) studied two forms of negation in Turkish, namely the suffix *"-mA"*, which is the basic negation marker, and the free morpheme *"-değil"*, which is used to negate noun or adjective-predicate copulative clauses. (Nadir ve Kabadayı, 2015) noted that negation shows structural differences in verbs and nouns. In sentences, verbs consistently acquire negative meaning through the verb-forming suffix *"-mA"*, while the characterization of negation in nouns is a matter of debate. In this study, the definitions of negativity are analyzed in two main sections.

According to (Morante ve Blanco, 2021), the scope of negation refers to the part of the meaning that is negated. Identifying the scope of negation presents different challenges depending on the language. In most cases, the scope is the largest element associated with a keyword, determiner, or phrase. Negation markers themselves are not included in the scope. The literature suggests that scope ambiguity is difficult to define, even for experts, as it is latent and subjective (Pröllochs ve ark., 2020). Therefore, negation labeling often requires clear annotation guidelines.

The concept of negation events is primarily applied in the biomedical field (Mahany ve ark., 2022). These domains include many medical or non-medical events and in most cases they are verbs, nouns or adjectives (Kim ve ark., 2008). An event can defines a process, an action or a state. The most popular event tagging work for negation is the GENIA (Ohta ve ark., 2007) event annotation guidelines. Although event studies are very common in biomedical corpora (see Table 2.1), in ConnanDoyle-neg the researchers tagged events for Conan Doyle Stories in their corpus (Morante ve Daelemans, 2012). Used dataset tagged events in the same way as the ConnanDoyle-neg study, regardless of whether they are medical or not. Scope and event tagging rules are summarized from the ongoing study Negation Tagging in Turkish Health Texts and shared in Section 8.

- **Semantic Negation**

  This negation type is semantic operators (affix, word, phrase, mimic etc.) that invert or oppose the meaning or situation. The negation cues of this class can be words such as nouns, adjectives, verbs, conjunctions and prepositions that have acquired different negative meanings over time. For example, the word *"felaket"* is a structurally negative polarity word and has a negative semantic connotation. In addition, the same word is sometimes used in similes, metaphors, etc. to create negative connotations with neutral or positive polarity words. Semantic negation is not marked in the dataset.

- **Structural Negation**

  Structural Negation refers to grammatical structures obtained through negation affixes and words. Commonly used negativity suffixes include *-mA*, *"-mAz"*, *"-sIz/-sUz"*, *"-mI/-mU"*, *"-ImsA/-UmsA"* etc. In addition to words that express negativity such as *"hayır"* (*"no"*), *"değil"* (*"not"*), *"hiç"* (*"never"*), there are also nouns and verbs that have inherently negative meanings. In addition, there are signs that can be selected specifically for the research area or that semantically fulfill the definition of negation. In the used dataset *"durdur-"*, *"sonlan-"*, *"çıkar-"* are labeled as negation cues according to their meaning in the sentence. For example, the sentence *"Akan kan durduruldu."* means that a situation has been eliminated. Usually *"çıkar-"*, *"kaldır-"*, they denote negation by absence. This class is divided into 3 groups in Turkish grammar, and the term double negation was added as the fourth in the dataset study.

▲ Morphological Negation Cues

In Turkish *"-mA"*, *"-mAz"*, *"-sIz"* suffix and *"a-"*, *"an-"*, *"de-"*, *"dez-"* prefixes can be examples for this category.

[Hastanın PA akciğer grafisinde patolojik bulgu *sap*tan]**ma**[dı].

Fotoduyarlılık normal hatta sınırlı ultraviyole ışığına maruziyet sonrasında gelişen **a**[*normal* bir tepki] olarak tanımlanır.

▲ Lexical Negation Cues

Here negation is achieved by a word or a sequence of words. The most familiar lexical markers are "yok" and "değil". These words are referred to in grammar as structural negation (Emeksiz Erk, 2010). The word *"yok"* is a logical negation marker expressing the state of non-existence, while *"değil"* is a logical negation marker expressing the opposite of the state or action.

[Tüm vücutta yaygın makülopapüler döküntüler, dinlemekle bilateral solunum seslerinde kabalaşmadan başka patolojik fizik muayene *bulgu*su **yok**tu].

[Hastanın geldiği dış merkezde başlanan *tedavi*si 7. gün **kes**ildi].

- ▲ Syntactic Negation Markers

  Although it does not appear in the medical dataset, only the pattern *"ne ... ne ..."* is used as a syntax in Turkish.

- ▲ Double Negation

  When there are two or more negation markers in the same sentence, markers with the same scope may cancel each other out. In this case negation cues will not be marked.

  Osteogenesis imperfektalı bazı hastalar dentisyonda Clinical veya radyografik olarak hiçbir **a**normallik göster**me**zken, bazıları önemli dişsel gelişim bozuklukları gösterirler.

  In this example, both "a-" and "-mez" are negation marks. The suffix "-mez" is attached to the verb of a verb phrase. Since "anomaly" is the subject of the phrase and "göstermezken" is the verb, both have scope over the whole phrase. Also, "not showing abnormality" is the same as saying "normal" in this context. This expression is not labeled as a negation.

3.2 NLP and DL

NLP history can be summarized by referencing (Roberts, 2023). In the 1940s, this field was started by researchers to translate languages to each other. Around the same period in history, from 1957-1970, symbolic and stochastic NLP approaches have been theorized. Symbolic or rule-based, researchers focused on formal languages and generating syntax; this group consisted of many linguists and computer scientists who considered this branch the beginning of artificial intelligence research. Stochastic researchers were more interested in statistical and probabilistic methods of NLP, working on problems of optical character recognition and pattern recognition between texts.

As (Chollet, 2021) pointed out, since the emergence of kernel methods in the 1990s, field researchers have been dedicated to feature engineering, manually transforming input data into a format suitable for shallow learning methods to generate valuable predictions. However, this approach was highly time-consuming, task-specific, and not easily accessible to non-experts. The book also pinned the arrival of deep learning in approximately 2012 as a significant revolution in NLP. Neural networks' ability to automatically craft relevant features in their layers has made these methods more applicable to new tasks and challenges. Subsequently, human effort has shifted towards designing the suitable neural network architecture for specific tasks and fine-tuning various hyper parameters during training. All these studies and later research divide NLP into several tasks like Speech Recognition, Part of Speech Tagging, Word Sense Disambiguation, Named Entity Recognition (NER), Sentiment Analysis, Text Generation etc. (IBM, 2023).

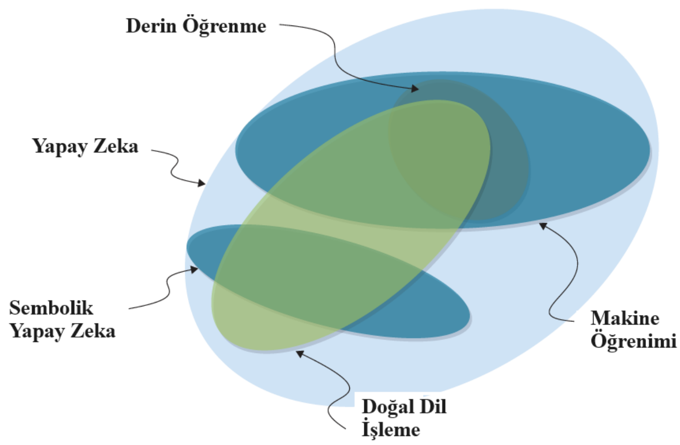


Figure 3.1: AI ve NLP Study Area Venn Representation Chollet (2021)

To give a short history in this paragraph, we can refer to the (Chollet, 2021) 's book. In the early 2010s, the rise of neural networks and deep learning in the field dramatically transformed NLP. Such techniques were shown to achieve state-of-art results for the most difficult NLP tasks, such as machine translation and text classification. The mid-2010s witnessed the development of the word2vec models and its variants sent2vec, doc2vec. The meaning of a word was assumed to be tied to its context, that is, the words surrounding it. In 2014, Seq to Seq models raise and extend the capability of keeping long term relation in a text further than bag-of-words which only has a short window size capability. RNN and LSTMs and later Bi-LSTMs show great improvement to understand long text input yet still they can lose information as the length increases and are computationally hard to process. In 2017, (Vaswani ve ark., 2017), researchers defined a new algorithm of Attention which dramatically boosted all these improvements, decreased the computational complexity and allowed the model to focus the input parts most relevant with output. Later, transformer models further improved the training time and results. Yet still there are lack of features that these models have hardened to cover like spelling errors, words that are not in dictionary or emojis which are highly used in social media etc. (Zhang ve ark., 2015) showed this in the context of character-level CNNs for text classification and demonstrated a remarkable robustness to misspellings.

3.2.1 Name Entity Recognition

Named Entity Recognition (NER) is a fundamental task in natural language processing (NLP) that involves identifying and classifying specific entities, such as names of people, locations, organizations, dates, and more, within a text. The goal of NER is to extract and classify these labels accurately to understand and process text more effectively. (Durango ve ark., 2023) defines NER as one of the most extensively studied tasks in information extraction. Study gives examples with references to show that current clinical NER systems have a wide range of capability for clinically relevant entities, including medical issues, diagnostic tests, treatment methods , medication names, and adverse events, as well as protected health information. Also, the latest advantage of the NER mechanism is giving the flexibility to annotate information not only by tagging, but also, by giving relational information about these entities (Chen, 2019).

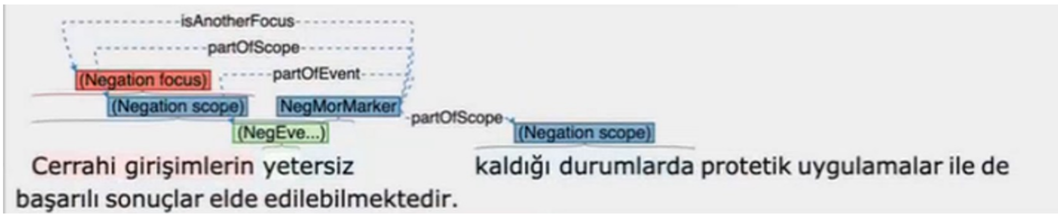


Figure 3.2: NER Labels

This method can be also applied into the tagging systems by identifying the relations over the multi word entities as in our case. In our dataset, tagging system designed to identify features and all related with the negation cue over a relational layer.

3.2.2 Tokenization

Tokenization is the process of converting text into a list of meaningful words or components. Words, suffixes or punctuation marks that appear for the first time are taken as unique components. Beyond basic character and word base tokenization, sub-word tokenization has became popular for identifying components of the texts. Additionally, in some research topics, sentence tokenization can also be applied when processing documents. Character-based and sentence-based approaches are beyond the scope of this paper. The tokenization techniques can be exemplified by the word "gözlenmedi".

① Sentence : ["gözlenmedi."]

② Word : ["gözlenmedi", "."]

③ Root : ["gözlen", "#me", "#di", "."]

④ Character : ["g", "#ö", "#z", "#l", "#e", "#n", "#m", "#e", "#d", "#i", "."]

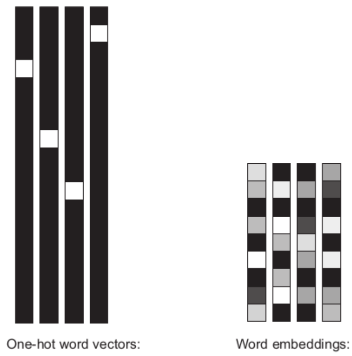
Here '#' claim continuity of the phrase. Tokenization process and method are explained in Chapter 5. Tokenization is important as all vectors will represent the pieces of the words which needs well spelled, representative tokens. These tokens are then saved as a dictionary, which is a set. The purpose of calling the dictionary a set is to indicate that the elements are unique. The dictionary contains the token and its assigned number (word-ids). The vectors representing these tokens can then be retrained according to the new dataset. This is why in most cases the third option is more useful than the second one. There are many tokenization tools developed for different types of techniques, such as Spacy (Honnibal ve Montani, 2017), Fastai(Howard ve ark., 2018), Bert-tokenizer (Devlin ve ark., 2019), etc. In this study BERT tokenizers have been used.

3.2.3 Word Embedding

Word Embedding (WE) is a unit vector representation of a word, placed in an index of a given length, containing mathematical weights. With these vectors, lexical words can be represented as a vectorial quantity in space (of the same size as the length value of the representation). The first examples of this representation are the One-hot encoding. One-hot encoding refers to the binary representation of categorical variables. In this method, all vectors are orthogonal to be unique. For this purpose, a vector of the same length will assigned 1 in a unique place and all other vector elements will be 0. Because of it vector length should have same with the number of unique elements. This reduces the efficiency of the system by increasing the computational cost and memory usage. Recently, this structure has been used for special cases, for representing classes from classification problem models and for small dictionary databases.

WE vectors consist of vector lengths of lower dimensions respect to one-hot encoding. In addition, all words can be uniquely expressed without being orthogonal. The computational complexity of WE has been reduced by decreasing the dimension. In addition, it has been observed that the representation vectors can also geometrically express semantic relations such as similarity, contrast, etc. between tokens (Figure 3.3). The fact that word representations can be trained using other datasets has resulted in more realistic representations, where the same word can have different meanings depending on the subject of the corpus it is connected to (Chollet, 2021). The word groups where this is most effective are homophones and polysemous words.

Figure 3.3: Word Embeddings (Chollet, 2021)



One-hot Vector:

Dictionary Size: 10000

Vector Length = 50:

token-A:	[0, 0, ... , 0, 0, 0, 0, 1]
token-B:	[0, 0, ... , 0, 0, 0, 1, 0]
token-C:	[0, 0, ... , 0, 0, 1, 0, 0]

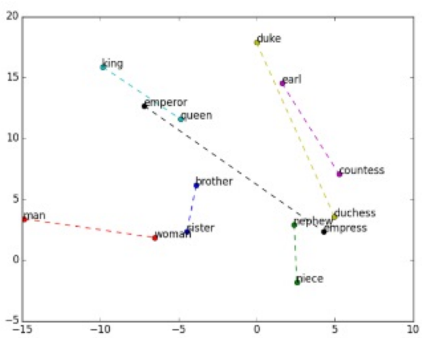
Word Embeddings:

Dictionary Size: 10000

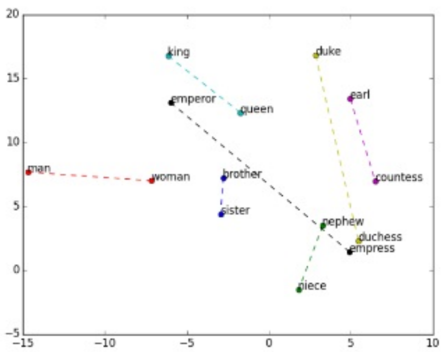
Vector Length = 50:

token-A:	[0.08, 0, ... , 0.32, 0.12, 0.06, 0, 0.13]
token-B:	[0, 0.7, ... , 0.8, 0.3, 0.12, 0.21, 0.01]

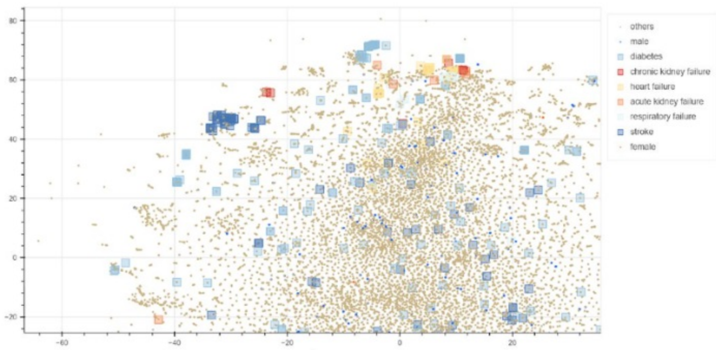
token-C: [0.42, 0.2, ... , 0.01, 0.02, 0, 0, 0.25]



(a) Stanford (2017)



(b) Stanford (2017)



(c) Wu ve ark. (2022)

Figure 3.4: Geometric Relation Between Word Embedding Vectors

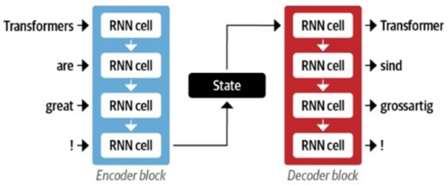
WE methods has been evolved in decent years. Bag of Words (BoW), TF-IDF, Word2Vec, GloVe and BERT can be given as examples for this (Turing, 2022). First generation methods (BOW and TF-IDF) represent input by using statistical and algebraic calculation with packages of words (Bag of Word) and with the frequency of the terms in sentences or documents (TF-IDF). The second generation of techniques uses probabilistic calculations based on words within a certain window on a word or a co-occurrence matrix (Almeida ve Xexéo, 2019) (Word2Vec, Glove, FastText etc).

The key feature of the embedding is the ability to re-train vectors on different topics using various datasets. This allowing them to store more information or specialize on a more specific topic. This means they can be used with similar datasets that take the token vectors generated by a model as input, without needing to retrain them from scratch. These token representations can also be downloaded and retrained on target corpus (see Chapter 3.2.5).

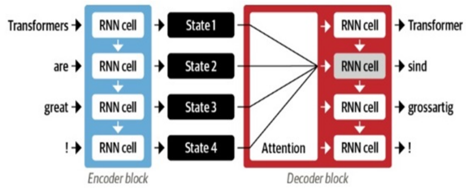
3.2.4 Transformers

Transformers is a neural network architecture introduced by (Vaswani ve ark., 2017) that has revolutionized the field of Deep Learning. It has made remarkable improvements in overall context understanding and computational efficiency where state-of-the-art works such as RNN, LSTM, BiLSTM, etc. have struggled with. In RNN, each input processed with an instance of time and in each round related weights and feedback calculated. The outputs are generated once per time unit and the errors reflected in the backward direction. Therefore, these models are time consuming and need computational resources. In addition, in sequential models, the context and semantic impact of a distant word weakens as the distance increases (Chollet, 2021). To reduce the impact of this effect, the successors of sequential models added memory units and inside feedback mechanisms to the cell. In addition, sequential models use the Encoder-Decoder architecture for sequence-to-sequence tasks to better understand the content in translation tasks. (Tunstall ve ark., 2022) argues that the results produced in the hidden layers of the encoder-decoder structure, output of the encoder, can be aggregated and transmitted. This can cause a bottleneck for the system in the decoder input and some important points of the data can be missed. The authors note that if these outputs are given to the decoder individually without being aggregated, this would create a large amount of data that would be difficult to process. Applying attention mechanism to the encoder output to select which features and information go to the decoder will reduce this effects. (Vaswani ve ark., 2017) summarized the successes achieved in sequential models, but emphasized that processing time and other fundamental problems of serialized processing remain. Figure 3.5 summarizes the RNN model and its successors developed to overcome the limitations.

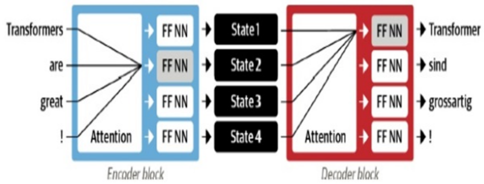
Figure 3.5: RNN Models



(a) RNN Enc-Dec Structure



(b) RNN Enc-Dec + Attention



(c) RNN Transformer

Figure 3.6 shows the complexities for the related models. n is the length of input, d is dimension k is the kernel number. As it can be seen from the table, the attention mechanism reduces complexity.

Layer Type	Complexity per Layer	Sequential Operations	Maximum Path Length
Self-Attention	$O(n^2.d)$	$O(1)$	$O(1)$
Recurrent	$O(n.d^2)$	$O(n)$	$O(n)$
Convolutional	$O(k.n.d^2)$	$O(1)$	$O(\log_k(n))$
Recurrent	$O(r.n.d)$	$O(1)$	$O(n/r)$

Figure 3.6: Complexities of Some DL Models

In the first step of the high-level architecture of Transformers structures, WEs are generated by aggregating positions and representations. Transfer learning is used to generate these first step output vectors. The models used in this study:

- Methods where Masked Language Model use unlabeled large Turkish corpora, next token/sentence prediction predict the next item by unsupervised techniques and produce word embedding. On this step also positions and representations aggregated,
- Second step is to configure prediction stage which use word embedding vectors in a neural network and maps classification label of input (Chapter 5.2).

The result vectors contain the semantic data of the tokens in the corpus. The pre-trained vectors trained with new word sequences. These vectors are then enriched with position information. This data is passed to a multi-head attention system that incorporates self-attention mechanisms to capture different types of relationships between tokens. Here the matrices containing the query, key and value vectors are dot product multiplied and the values undergo some corrections. These mathematical adjustments include layer normalization and residual connections. These parts speed up the system by normalizing the magnitudes of the values and combining the outputs for computational simplicity. Finally, this hidden output pass through a FFNN where last layer uses soft-max activation function. These structures are described by (Vaswani ve ark., 2017) in Figure 3.7. Here, encoders evaluate the input corpus for various features and generate corpus-aware WEs. The encoder output is enriched with additional semantic information found by the attention mechanism. The decoder system combines these different input representations and passes them to the final layer model to generate the result. It is worth to note that Encoder-Decoder models are generally well suited for sequence-to-sequence tasks such as translation and Q&A tasks. Another feature of this architecture is the WE extensibility mentioned in Section 3.2.3, where only encoders can generate WEs to be used in other models (see 3.8).These WEs

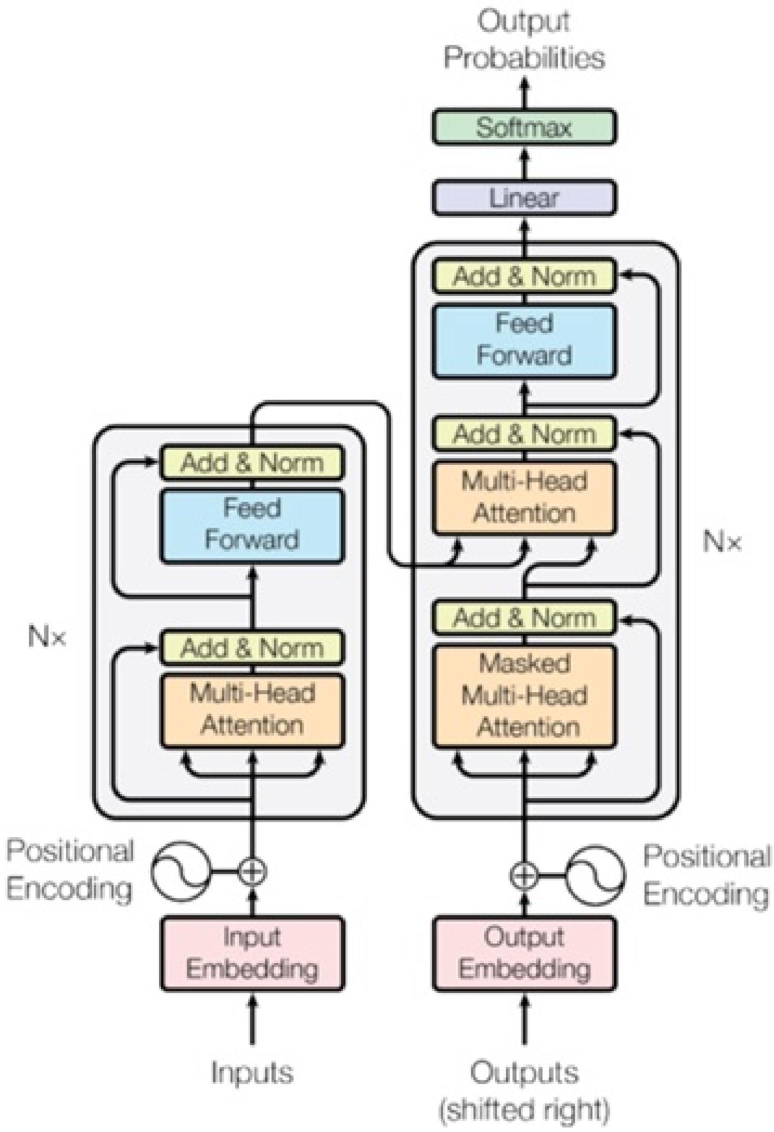


Figure 3.7: Transformer Architecture (Vaswani ve ark., 2017)

can be retrained on a narrower corpus if desired, or they can be used directly without the need for training. This study is used WEs in the HuggingFace library without retraining.

3.2.5 Transfer Learning

Transfer Learning can be understood as applying knowledge gained from solving one problem to help solve a new or related problem. Instead of starting the learning process with randomly generated vectors, Transfer Learning capable to carry the knowledge gained from previous training and use pre-trained vectors. This process can be applied sequentially to different dataset domains and by using more special task corpus, WEs also can be specialized.

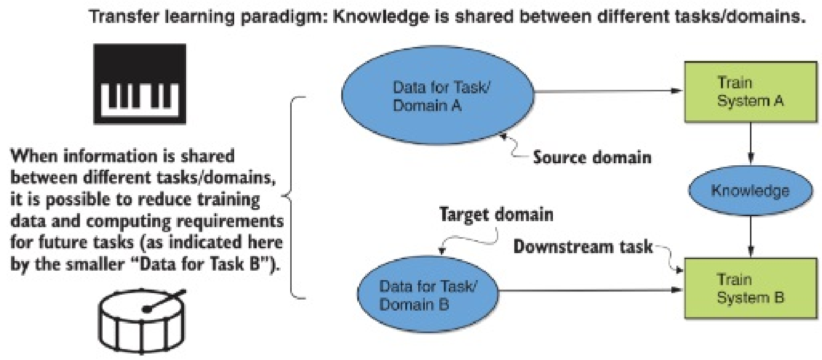


Figure 3.8: Knowledge transfer on Transfer Learning (Azunre, 2021)

Domain Adaptation (DA) is the process of training a machine learning model on a source corpus domain and then adapting it to perform well on a different (but related) problem dataset with same task. In the context of NLP, the source corpus can consist of many different datasets and data types such as news articles, social media posts, job interviews or scientific papers. Figure 3.2.5 shows the simple algorithm behind this mechanism. Unsupervised Domain Adaptation (UDA) updates weights on unlabeled big data. Structural Correspondence Learning (SCL) applies multi-layer perceptions to infer different features and increase their overall impact on the results.

Google's NLP model, called BERT (Devlin ve ark., 2019), was trained on a large amount of English text from Wikipedia and BookCorpus. ULMFiT (Howard ve Ruder, 2018) is a method that fine-tunes any neural network language model for specific tasks which usually used for text classification. It tunes different parts of the model at different rates. OpenAI's GPT (OpenAI, 2023) has adapted the transformer model into a fine-tunable language model for NLP by retaining certain components and modifying others. BERT does

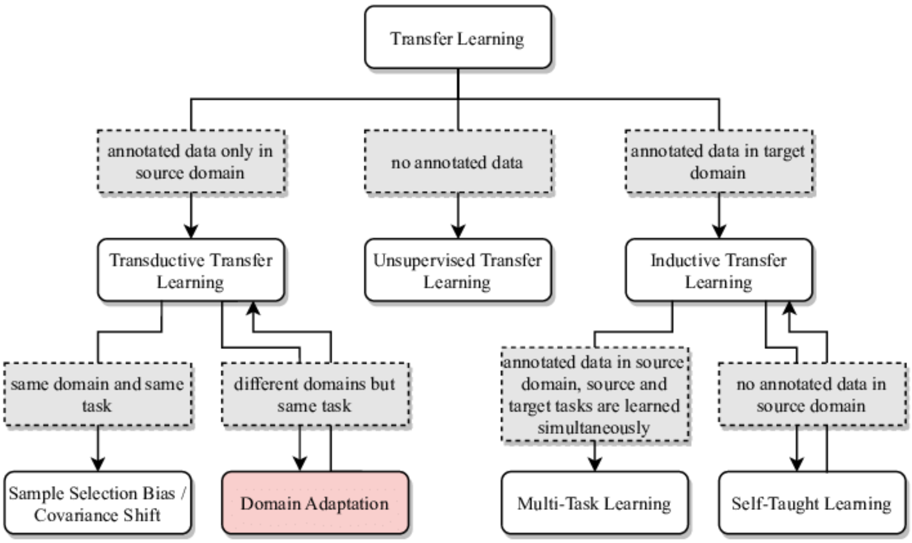


Figure 3.9: Transfer Learning Types (Triess ve ark., 2021)

the opposite, it keeps different parts and focuses on masking and correctly guessing next words during training. These models have shown that word representations can be tuned for specific NLP tasks with small amounts of labeled data. (Chollet, 2021).

4 DATASET

This study uses an abstract corpus, the first test dataset of the Negation Annotation in Turkish Health Texts project. The dataset of this study is an ongoing project. It contains case reports and articles produced by medical personnel in hospitals. The sub-groups of reports can be classified under 12 main departments:

- Department of Emergency Medicine
- Department of Pediatrics
- Department of Skin and Venereal Diseases
- Department of Infectious Diseases and Clinical Microbiology
- Department of Physical Medicine and Rehabilitation
- Department of Chest Diseases
- Department of Public Health
- Department of Internal Medicine
- Department of Cardiology
- Department of Neurology
- Department of Mental Health and Diseases
- Department of Medical Genetics

Dataset includes negation cues, scope and event (see Table 2.1). Researchers from Computer Engineering and Turkish Language and Literature departments of Ege University pooled pdfs published in those 12 departments and filtered out duplicate publications. Labeling started after linguistic rules have been determined. Finally, the results that passed the compliance and control checks were pre-processed with the generated xmi format (see Figure 4.1 for the output format of Inception tool). To be able to parse the output of Inception, (Klie ve de Castilho, 2021) library has been used. The analysis of the dataset which are contains only the case reports are given in Figure 4.3.

It can be seen that in Turkish dataset, the negated sentence ration and negation cues used in a single sentence are high regarding to other datasets in different languages. Max number of cues used in a sentence is 4 in some documents.


```
<type5:Token xmi:id="30687" sofa="1" begin="16191" end="16205" order="0"/>
<type5:Token xmi:id="30700" sofa="1" begin="16205" end="16206" order="0"/>
<type5:Token xmi:id="30713" sofa="1" begin="16207" end="16210" order="0"/>
<type5:Token xmi:id="30726" sofa="1" begin="16211" end="16213" order="0"/>
<type5:Token xmi:id="30739" sofa="1" begin="16214" end="16222" order="0"/>
<type5:Token xmi:id="30752" sofa="1" begin="16223" end="16232" order="0"/>
<type5:Token xmi:id="30765" sofa="1" begin="16233" end="16238" order="0"/>
<type5:Token xmi:id="30778" sofa="1" begin="16239" end="16243" order="0"/>
<type5:Token xmi:id="30791" sofa="1" begin="16244" end="16252" order="0"/>
<type5:Token xmi:id="30804" sofa="1" begin="16253" end="16259" order="0"/>
<type5:Token xmi:id="30817" sofa="1" begin="16260" end="16269" order="0"/>
<type5:Token xmi:id="30830" sofa="1" begin="16270" end="16273" order="0"/>
<type5:Token xmi:id="30843" sofa="1" begin="16274" end="16286" order="0"/>
<type5:Token xmi:id="30856" sofa="1" begin="16287" end="16292" order="0"/>
<type5:Token xmi:id="30869" sofa="1" begin="16293" end="16304" order="0"/>
<type5:Token xmi:id="30882" sofa="1" begin="16304" end="16305" order="0"/>
<custom2:NegationMarker xmi:id="30919" sofa="1" begin="329" end="332" markerType="NegMorMarker" scope="31333 31336" focus="31348" event="31359" coordinatior
<custom2:NegationMarker xmi:id="30930" sofa="1" begin="394" end="397" markerType="NegMorMarker" scope="31369" focus="31380" event="31435" coordinationpartic
<custom2:NegationMarker xmi:id="30895" sofa="1" begin="546" end="549" markerType="NegMorMarker" scope="31968 31971" focus="31983" event="" coordinationparti
<custom2:NegationMarker xmi:id="30941" sofa="1" begin="573" end="576" markerType="NegMorMarker" scope="31449 31452" focus="" event="32447" coordinationparti
<custom2:NegationMarker xmi:id="30954" sofa="1" begin="648" end="651" markerType="NegMorMarker" scope="31989 31992" focus="" event="32436" coordinationparti
<custom2:NegationMarker xmi:id="30906" sofa="1" begin="800" end="803" markerType="NegMorMarker" scope="32479 32482" focus="" event="32004" coordinationparti
<custom2:NegationMarker xmi:id="30967" sofa="1" begin="846" end="849" markerType="NegMorMarker" scope="32493" focus="" event="32504" coordinationparticle="'
<custom2:NegationMarker xmi:id="30980" sofa="1" begin="1017" end="1020" markerType="NegMorMarker" scope="31504 31507" focus="32469" event="32458" coordinati
<custom2:NegationMarker xmi:id="30991" sofa="1" begin="2181" end="2184" markerType="NegMorMarker" scope="31522 31525" focus="32526" event="32515" coordinati
<custom2:NegationMarker xmi:id="31002" sofa="1" begin="3012" end="3015" markerType="NegMorMarker" scope="32540 32543" focus="32566" event="32555" coordinati
<custom2:NegationMarker xmi:id="31283" sofa="1" begin="3299" end="3301" markerType="NegMorMarker" scope="32813 32816" focus="32577" event="32015" coordinati
<custom2:NegationMarker xmi:id="31013" sofa="1" begin="3412" end="3414" markerType="NegMorMarker" scope="31540 31543" focus="32828" event="32026" coordinati
```


Figure 4.1: XMI format example of tagged sentence

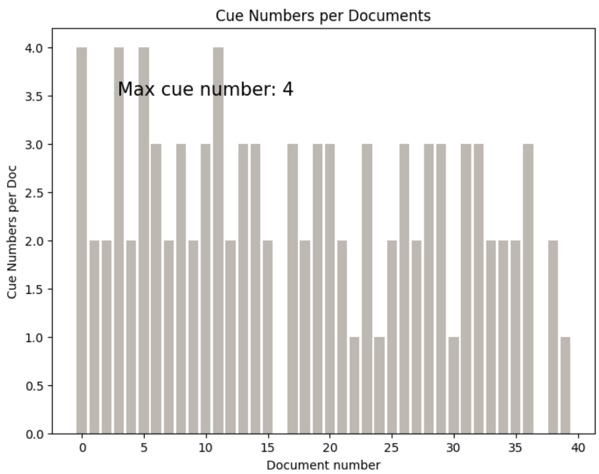


Figure 4.2: Maximum number of cues in sentences according to the documents

```
Cue Label Numbers on Train Validation and Test Dataset:

labels
O                  30554
B-NegMorMarker       437
B-NegLexMarker       168
I-NegLexMarker         2
Name: count, dtype: int64
labels
O                  7844
B-NegMorMarker       95
B-NegLexMarker       34
I-NegLexMarker        2
Name: count, dtype: int64
labels
O                  9275
B-NegMorMarker      136
B-NegLexMarker       58
I-NegLexMarker        1
Name: count, dtype: int64
     Dataset      Total Sent. #  Negated Sent. #   Sent. Negation Rate     Token Negation Rate

Train :                 1736            492                % 28.30                       % 1.90

Validation :             434            109                % 25.10                       % 1.60

Test :                   543            160                % 29.50                       % 2.10
```

Figure 4.3: Dataset Statistics

5 Methodology and System Design

There are two main methods to identify labels in negation studies.

- Span detection: The output layer of the model is designed to predict the index of start and end points for labeling. The sequence number of the characters in sentence or document used to predict cues, scopes or events. (Althari ve Alsulmi, 2022) reduced this method to a binary classification method as if the start and end index is same with input or not.
- Token classification: In this method, tokens tagged with BIO format (which is widely used in medical NLP tasks) and for each token the method worked as a classification model. Related tags prepared as a sequence and aligned with the sentence token. Then they fed into model as labels.

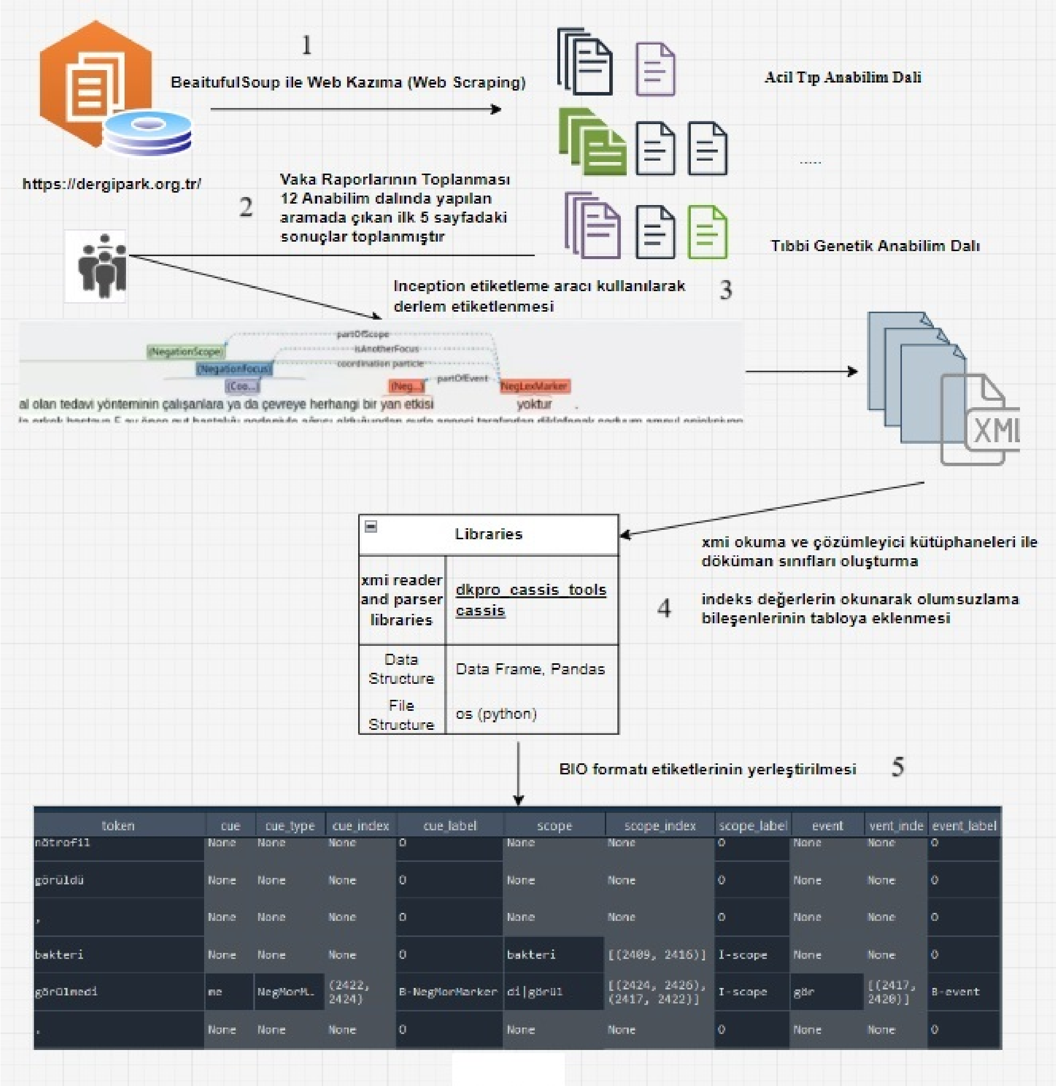


Figure 5.1: Dataset Preparation

Figure 5.1 steps can be summarize as:

① : Collecting articles and case reports with Web Scraping

② : First 5 pages results taken in all main department listed and same articles filtered

③ : Annotation Stage by using Inception Tool (Project group)

④ : Parsing xmi output to produced csv datasets

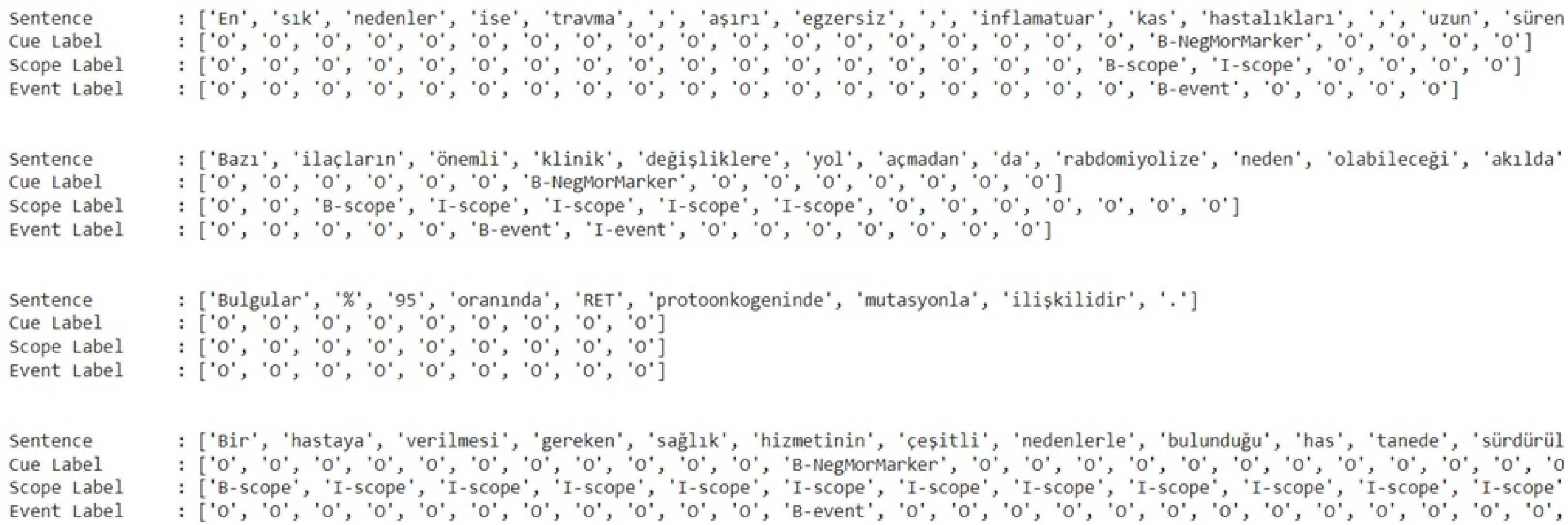
⑤ : BIO Tagging

Turkish medical corpus have many challenges. Some of negation features are discontinuous which lead double or triple index value for start and end indexes of span detection method. Token classification method, also, is a challenge for Turkish language as available models are not optimized to separate words into its compounds (prefixes, roots, affixes). There are a lot of sentences which include multiple negation (see Figure 4.2). To overcome this affect, sentences have been repeated regarding their cue number and fed to model as all replicas include one cue. This may lead a confusion for a single turn model, as model saw some negations are not annotated although there are negation. For this, process divided into two stage. First; cues are found by not repeating the sentences and secondly, for the features, the sentences replicate regarding to their cue number. In features model, cue information is provided with the embedding vectors for each words (see Figure 5.4).

5.1 Pre-processing of Dataset

Inception Annotation Tool (Klie ve ark., 2018) supports xmi and tsv format for special labeling. This study used xmi format as shown in Figure 4.1). "token", "cue_type", "cue_index", "scope_index", "event_index" parameters has been extracted and placed in Data Frame objects. BIO tag labels are also assigned to tokens for label classification tasks (see Figure 5.1). These labels are applied to words because Turkish tokenizers do not efficiently separate tokens according to grammar and prefixes (see Chapter 7).

Two datasets prepared with Data Frame objects. Each have document number, sentence_id, ... Cue dataset include unique sentences with multiple negation cue available option. Other dataset is a word indexed dataset where unique_sentence_id column also included to be able to make sentence dataset in second step. This last type include sentences repeated with the number of cues that they have. Second dataset is used for feature (scope, focus) detection.



```
Sentence     : ['En', 'sık', 'nedenler', 'ise', 'travma', ',', 'aşırı', 'egzersiz', ',', 'inflamatuar', 'kas', 'hastalıkları', ',', 'uzun', 'süren
Cue Label    : ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-NegMorMarker', 'O', 'O', 'O', 'O']
Scope Label  : ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-scope', 'I-scope', 'O', 'O', 'O', 'O']
Event Label  : ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-event', 'O', 'O', 'O', 'O']

Sentence     : ['Bazı', 'ilaçların', 'önemli', 'klinik', 'değişliklere', 'yol', 'açmadan', 'da', 'rabdomiyolize', 'neden', 'olabileceği', 'akılda'
Cue Label    : ['O', 'O', 'O', 'O', 'O', 'O', 'B-NegMorMarker', 'O', 'O', 'O', 'O', 'O', 'O', 'O']
Scope Label  : ['O', 'O', 'B-scope', 'I-scope', 'I-scope', 'I-scope', 'I-scope', 'O', 'O', 'O', 'O', 'O', 'O', 'O']
Event Label  : ['O', 'O', 'O', 'O', 'O', 'B-event', 'I-event', 'O', 'O', 'O', 'O', 'O', 'O', 'O']

Sentence     : ['Bulgular', '%', '95', 'oranında', 'RET', 'protoonkogeninde', 'mutasyonla', 'ilişkilidir', '.']
Cue Label    : ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O']
Scope Label  : ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O']
Event Label  : ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O']

Sentence     : ['Bir', 'hastaya', 'verilmesi', 'gereken', 'sağlık', 'hizmetinin', 'çeşitli', 'nedenlerle', 'bulunduğu', 'has', 'tanede', 'sürdürül
Cue Label    : ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-NegMorMarker', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O
Scope Label  : ['B-scope', 'I-scope', 'I-scope', 'I-scope', 'I-scope', 'I-scope', 'I-scope', 'I-scope', 'I-scope', 'I-scope', 'I-scope', 'I-scope'
Event Label  : ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-event', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O',
```

Figure 5.2: Grouping Dataset into Sentences where Each Word has One Label

After tokenization, the sentence and negation components in the dataset should also be expanded in a way that preserves the integrity of the words allocated to tokens. Figure 5.2 shows examples of sentences and tags. When synchronizing tags to tokens, 3 basic methods can be used.

① First method is to assign only the first token of the word with word label. Other tokens belonging to the same word are assigned as "-100". In this scenario only the first token results taking into account and others not. Although it leads to high results, this method may make it difficult to understand the location of the negation marker for affix-intensive languages. Also, model will not apply enough attention to insider tokens which include the cue, scope or event where prefixes used in negation. For this reason, when components such as scope and event are included, it is inferred that it may be difficult to find the boundaries of the components. (Huggingface, 2021) and NUBes model (Pabón ve Montenegro, 2021) can be given good examples for this method.

② In second method all token can be assigned with the word label. This method is lack to show if the feature is in the word if the label is beginning ("B-label"). Niels (2020) used this approach.

③ In last method, tokens are treated as separate terms and the first token is taken as the initializer marker. In this method, if the word tag starts with “B-label”, the following tokens are marked as “I-label”. In cases where the word label is “I-label” and “O”, all tokens are marked with the same label. Since each token has a separate representation vector, this method is also included in the model training. Approach 3 was tested on the models in this study.

The results has been compared in Chapter 6, regarding to their labeling technique.

5.2 Model Selection

5.2.1 Models

Bidirectional Encoder Transformer (BERT) (Devlin ve ark., 2019) is a layer of models developed by Google that significantly improves NLP. Unlike context-independent models such as Word2vec, BERT is context-based and provides superior performance on a variety of NLP tasks such as question answering and text generation. The BERT-base consists of 12 connected encoder architectures, each with 12 attention layers, and units that hold the output of each hidden layer. Embeddings in BERT are generated by token, segment and position embeddings. BERT functions as an auto-encoding language model that reads sentences bidirectionally to predict masked words in a masked language modeling task, improving contextual language understanding. Thanks to this initial pre-trained masked model, it can also apply initialized vector representations to be used in transfer learning when processing training. The tested models in this study can be listed as:

- ajtamayoh/Negation_Scope_Detection_NubEs _Training_Testing_mBERT_fine_tuned (Tamayo, 2023)
- akdeniz27/bert-base-turkish-cased-ner (Taner Akdeniz, 2023)
- alpcansoydas-turkish-ner-bert-base (Soydaş, 2023)
- bert-base-multilingual-cased (Devlin ve ark., 2019)
- bvanakenclinical-assertion-negation-bert (van Aken ve ark., 2021)
- mys/electra-base-turkish-cased-ner (Mys, 2023)
- liatoutou/distilbert-base-uncased-finetuned-negation-scope-classification (Liatoutou, 2023)
- loodos/bert-base-turkish-uncased (Loodos, 2023)
- savasy/bert-base-turkish-ner-cased (Yildirim, 2024)
- shoubhikc/negation_bert (Shoubhikc, 2023)
- UMCU/MedRoBERTa.nl_NegationDetection (van Es ve ark., 2023)

5.2.2 Model Selection Method

The listed Hugging-face models (see Chapter 5.2.1) have been download and used in Google Colab platform. Hyper-parameters are given in Table 5.1. Cue dataset used to record model performance in a loop with different combinations. 240 test made over these different hyper-parameters. Best 25 results take into account to select next models for feature detection task. Some of the tested model results and best results model measurements shared in Chapter 6).

In this study, each rounds for cue and feature models used pre-trained downloaded models. No trained cue model used in feature detection task. This event added as a reference in future works (see 7.2). In the second part, the number of layers and neurons of the feed-forward model, classifier part, were changed and tested. The appropriate number of layers was found as 2 and 3. Thus, a multi-layer classification model used to avoid the risk of bottle-neck. In addition, for the hyper-parameter, combinations tested with the reduced combination list given in Table 5.2.

Drop out	Learning Rate	Weight Decay
0.15	3e-4	0
0.30	1e-4	0.005
0.45	5e-5	0.01
	1e-5	0.015
	5e-6	

Table 5.1: Hyper-parameters of Cue Models

2 experiment set up results regarding to parameters and methods are given in Chapter 6 and results discussed in Chapter 7.

Drop out	Learning Rate	Weight Decay
0.2	1e-4	0
0.4	5e-5	0.01
		0.015

Table 5.2: Hyper-parameters of Scope and Event Models

5.3 Experiments

The cue detection and feature detection models are trained over downloaded BERT models to extract embedding vectors (see Figure 5.4). Produced embedding fed into a FFNN which calculates the probabilities of labels. In feature detection, the cue labels of tokens inserted as an additional one-hot encoding vector with token embedding (similar to the approaches (Cruz Díaz ve ark., 2012) and (Althari ve Alsulmi, 2022)).

Tokenized input tokens will not be align with labels. Three tagging method has been used in literature to align the input tokens with their labels (see Figure 5.4). To apply method which is shown in to inputs the layers should be aligned with the tokens.

① Tagging first token with the word label. Other tokens belonging to same word will labeled as "-100" and will not calculate while error analysis.

② Tagging all tokens with the same label as their word has.

③ Continuous tagging where program tagged upcoming tokens of word as "B-label" if it is first token and word label is "B-label" and then other tokens will take "I-label". If word label is "I-label" or "O" then others will be "I-label" respectively.

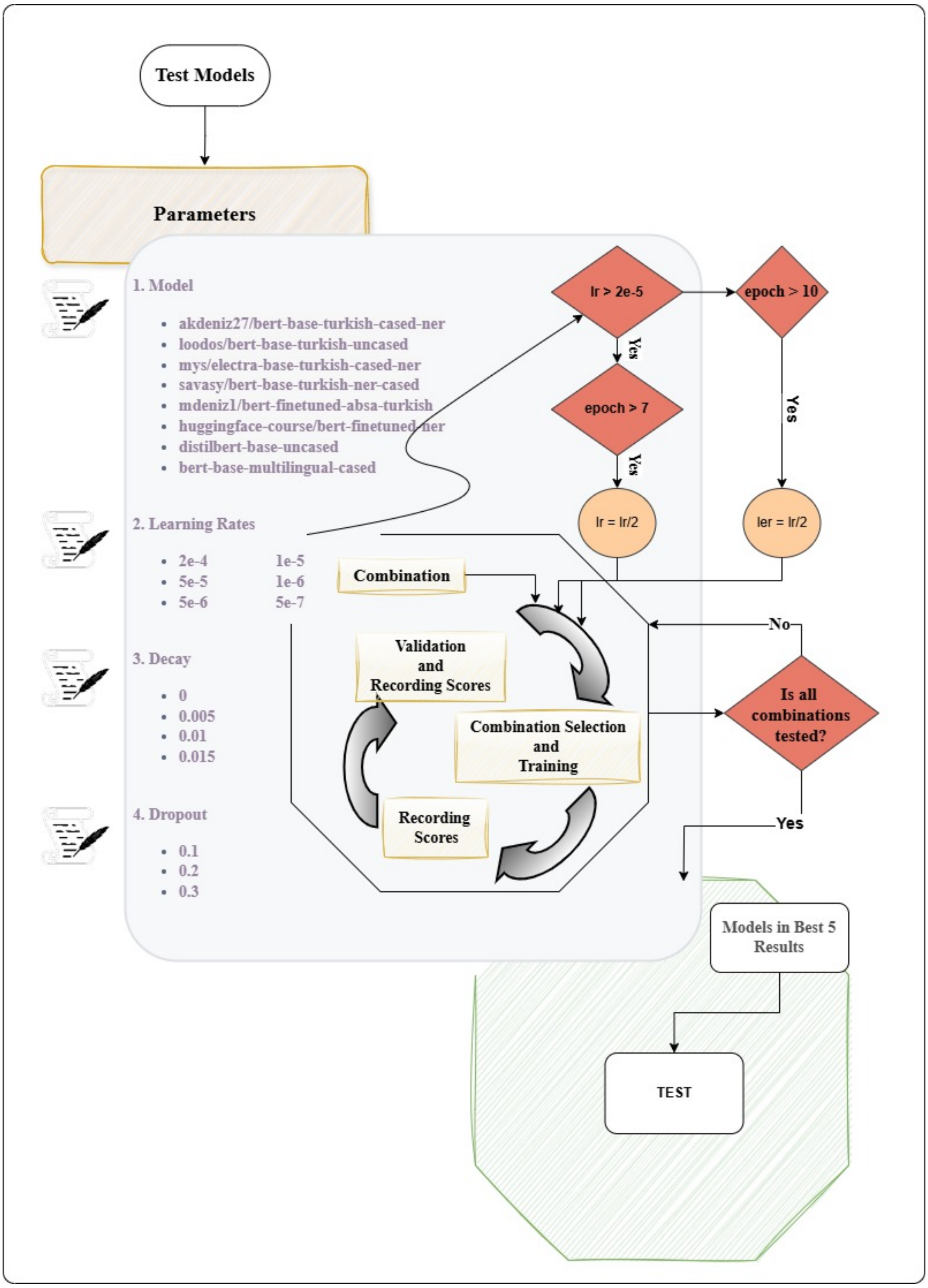


Figure 5.3: System Design

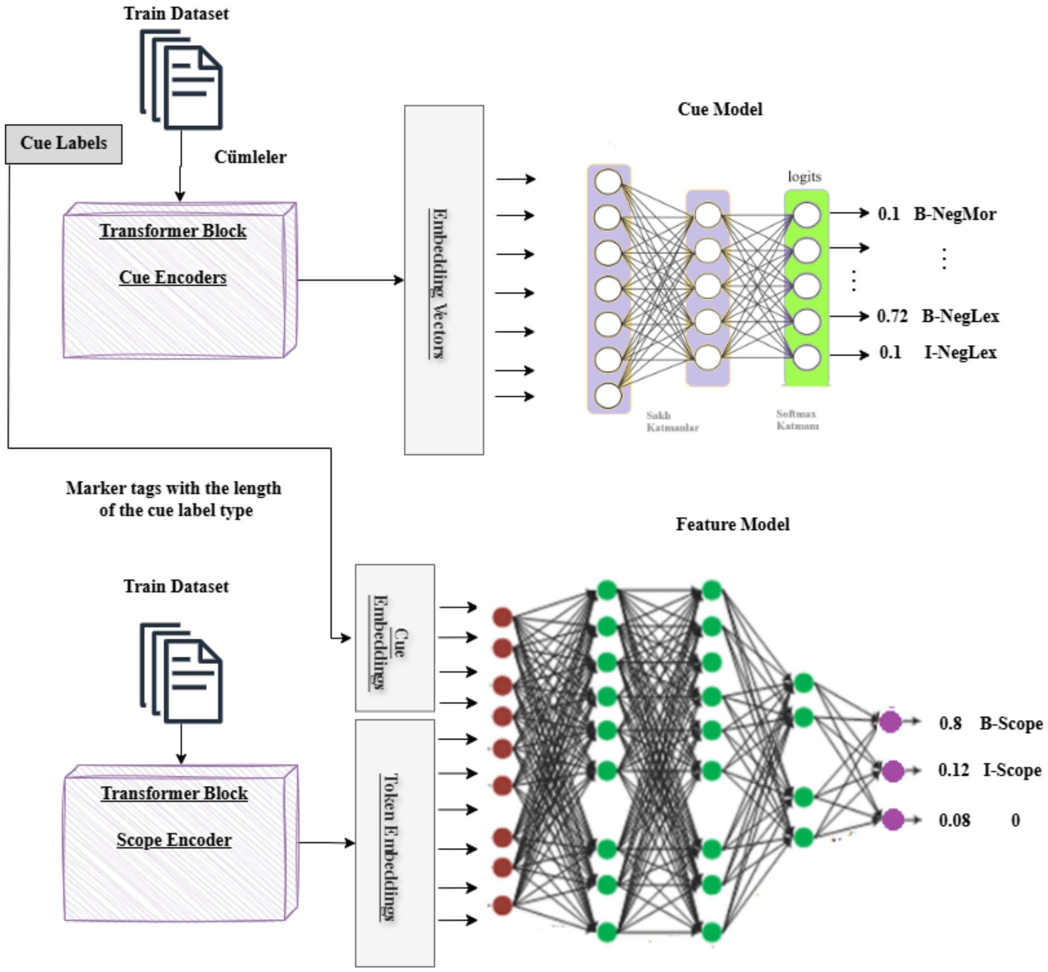
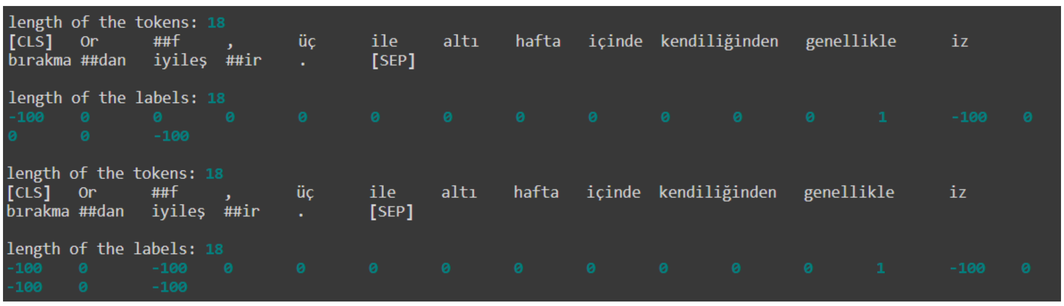


Figure 5.4: Model Training

Chapter 6 also discussed performance of these different labeling methods. To compare methods, results are shown for 3 different labeling methods. Figure 5.5, Figure 5.6 and Figure 5.7 shows the labeling methods.



```
length of the tokens: 18
[CLS]   Or      ##f     ,       üç      ile     altı    hafta   içinde  kendiliğinden   genellikle      iz
bırakma ##dan   iyileş  ##ir    .       [SEP]

length of the labels: 18
-100    0       0       0       0       0       0       0       0       0       0       0       1       -100    0
0       0       -100

length of the tokens: 18
[CLS]   Or      ##f     ,       üç      ile     altı    hafta   içinde  kendiliğinden   genellikle      iz
bırakma ##dan   iyileş  ##ir    .       [SEP]

length of the labels: 18
-100    0       -100    0       0       0       0       0       0       0       0       0       1       -100    0
-100    0       -100
```

Figure 5.5: First Token Labeling

```
Belirteçler:
[CLS]           Bir             hastaya         verilmesi       gereken         sağlık          hizmetinin      çeşitli         nedenlerle      bulunduğu
has             tane            ##de            sürdürül        ##emeyeceği     ##ne            ,               muayene         eden            doktorun
karar           vermesi         ##yle           hastanın        başka           bir             hastaneye       nakil           süreci          başlar
.               [SEP]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]
[PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]

İşaretleyiciler:
0               0               0               0               0               0               0               0               0               0
0               0               0               B-NegMorMarker  B-NegMorMarker  B-NegMorMarker  0               0               0               0
0               0               0               0               0               0               0               0               0               0
0               0               0               0               0               0               0               0               0               0
0               0               0               0               0               0               0               0               0               0

Kapsam:
0               B-scope         I-scope         I-scope         I-scope         I-scope         I-scope         I-scope         I-scope         I-scope
I-scope         I-scope         I-scope         I-scope         I-scope         I-scope         0               0               0               0
0               0               0               0               0               0               0               0               0               0
0               0               0               0               0               0               0               0               0               0
0               0               0               0               0               0               0               0               0               0
```

Figure 5.6: Same Label Tagging

```
Belirteçler:
[CLS]           Bir             hastaya         verilmesi       gereken         sağlık          hizmetinin      çeşitli         nedenlerle      bulunduğu
has             tane            ##de            sürdürül        ##emeyeceği     ##ne            ,               muayene         eden            doktorun
karar           vermesi         ##yle           hastanın        başka           bir             hastaneye       nakil           süreci          başlar
.               [SEP]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]
[PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]           [PAD]

İşaretleyiciler:
0               0               0               0               0               0               0               0               0               0
0               0               0               B-NegMorMarker  I-NegMorMarker  I-NegMorMarker  0               0               0               0
0               0               0               0               0               0               0               0               0               0
0               0               0               0               0               0               0               0               0               0
0               0               0               0               0               0               0               0               0               0

Kapsam:
0               B-scope         I-scope         I-scope         I-scope         I-scope         I-scope         I-scope         I-scope         I-scope
I-scope         I-scope         I-scope         I-scope         I-scope         I-scope         0               0               0               0
0               0               0               0               0               0               0               0               0               0
0               0               0               0               0               0               0               0               0               0
0               0               0               0               0               0               0               0               0               0
```

Figure 5.7: Continuous Label Tagging

5.4 Evaluation

Evaluation is done by "F1 Score", *"Precision" (P)* and *"Recall" (R)* metrics. Also the loss analysis has been made while training and validation. F1 score calculated by using Precision *"P"* and Recall *"R"* weighted average. Figure 5.8 shows these binary classification metrics which are used to calculate P and R.

P and R are F1 score metrics used in classification, pattern recognition, image processing and information extraction tasks. Figure 5.8 shows the binary metrics for calculation of these metrics. It is important to note that these calculations are separately applied to each binary classification of the labels so they are calculated over each labels. Actual values are represented as True and False while predicted values were positive and negative.

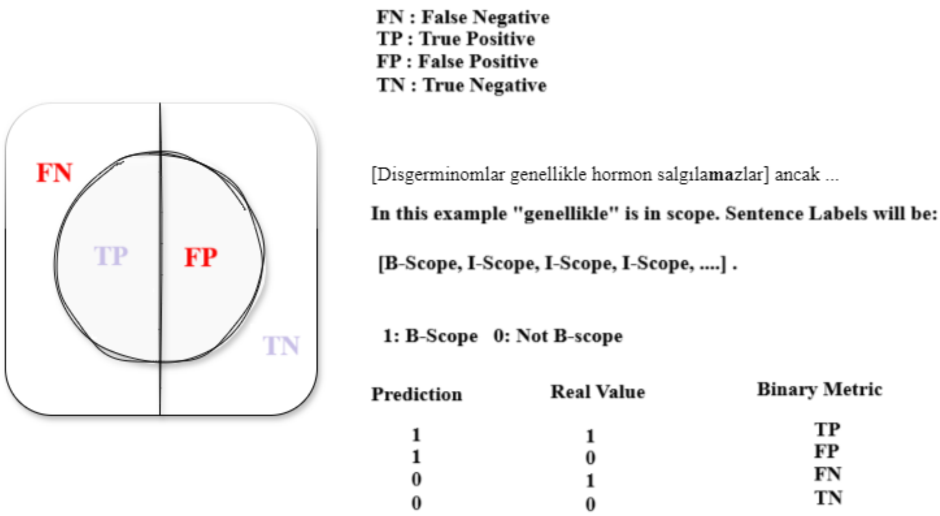


Figure 5.8: Binary Metrics

$$\mathrm{P} = \frac{TP}{TP + FP} \qquad \text{(Positive Prediction Value)} \tag{5.1}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \qquad \text{(Relevant Result Measurement)} \tag{5.2}$$

Precision and accuracy are important metrics. However, one may take priority over the other where false positive or false negative results are important. For example, in medical diagnostics, a false positive test may result in unnecessary treatment and costs, while a false negative result for some diseases may result in the disease being missed. The decision of which parameter is more important depends on the properties of datasets and tasks. For fraud detection systems, FN may become more important than the FP. Not raising an alarm in a fraudulent situation can cause much more financial damage than raising an alarm in a non-fraudulent situation. The F1 Score is a scalar measurement term that is the harmonic mean of precision and accuracy measurements.

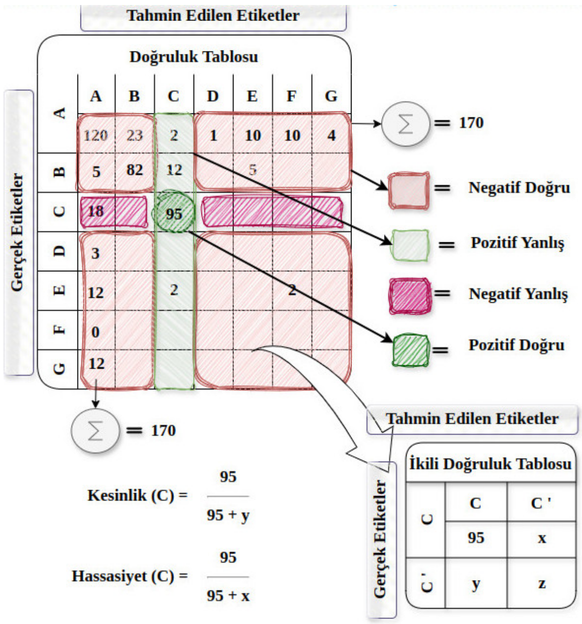


Figure 5.9: Multi-Class Complexity Matrix and Binary Projection Matrix on Selected Class

$$\mathrm{F}_1 = \frac{2 \cdot K \cdot H}{K + H} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{5.3}$$

For classification tasks with more than two classes, the process involves calculating, summing and normalizing these binary values for each label separately. For each label there is a binary confusion matrix (see Figure 5.9). Performance metrics P, R ve F1 calculated for each and take into account for calculating over all F1.

6 RESULTS

The test results of methods (see Chapter 5.2) for the cue detection has been given in the Figure 6.1 and Figure 6.2. For cue recognition task, Figure 6.3 and Figure 6.4 shows F1 and loss graphs of the best performance model.

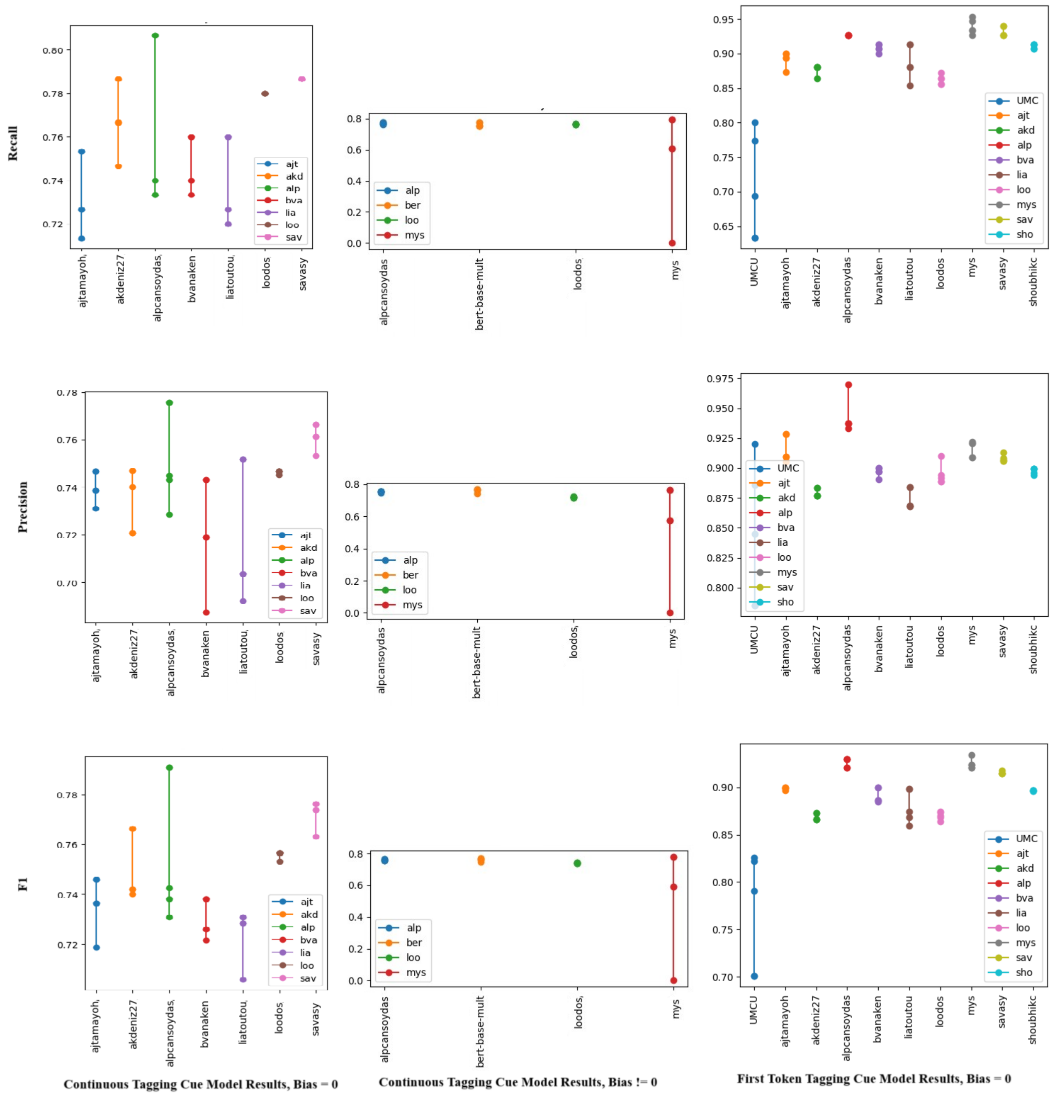


Figure 6.1: Negation Cue Model Results with the Highest Performance Metrics on Different Labeling Techniques

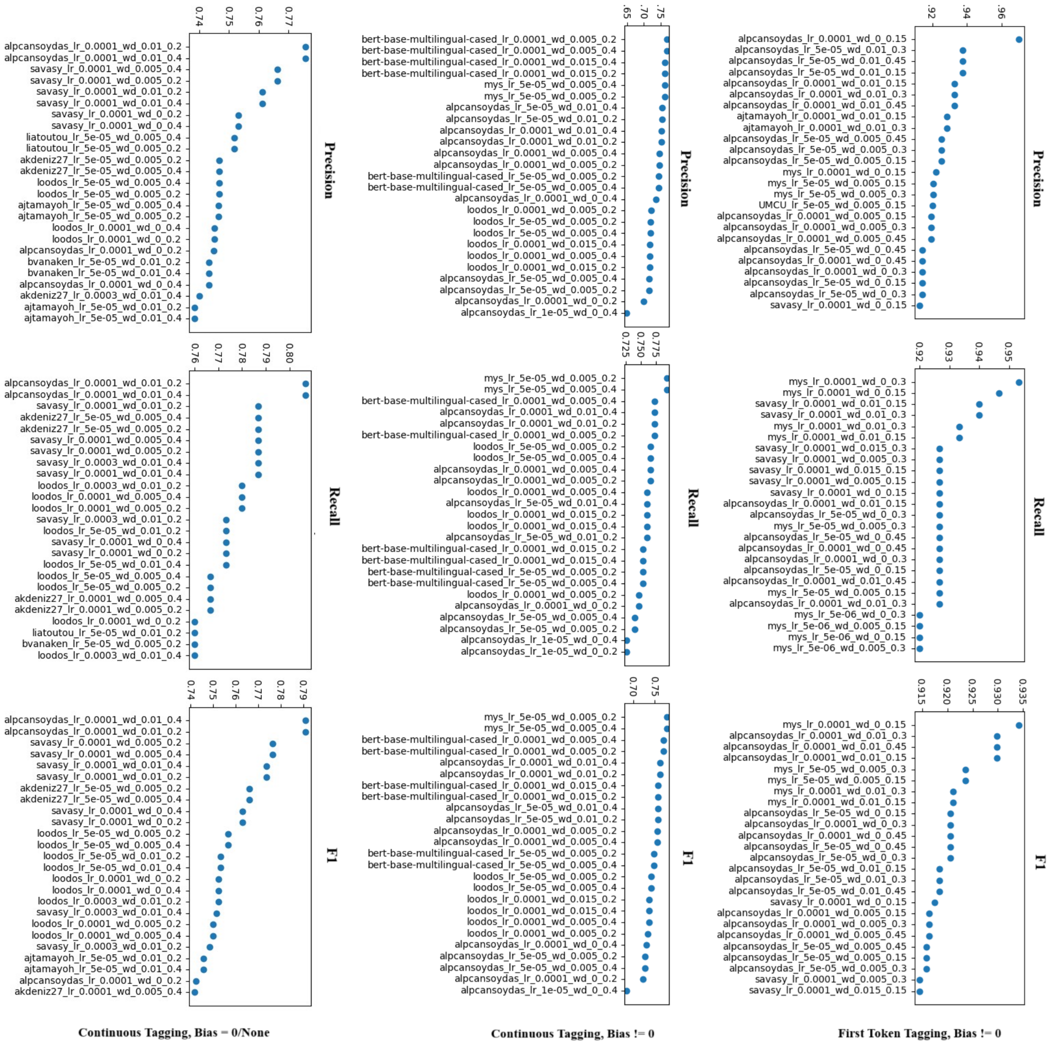


Figure 6.2: General Cue Model Comparison

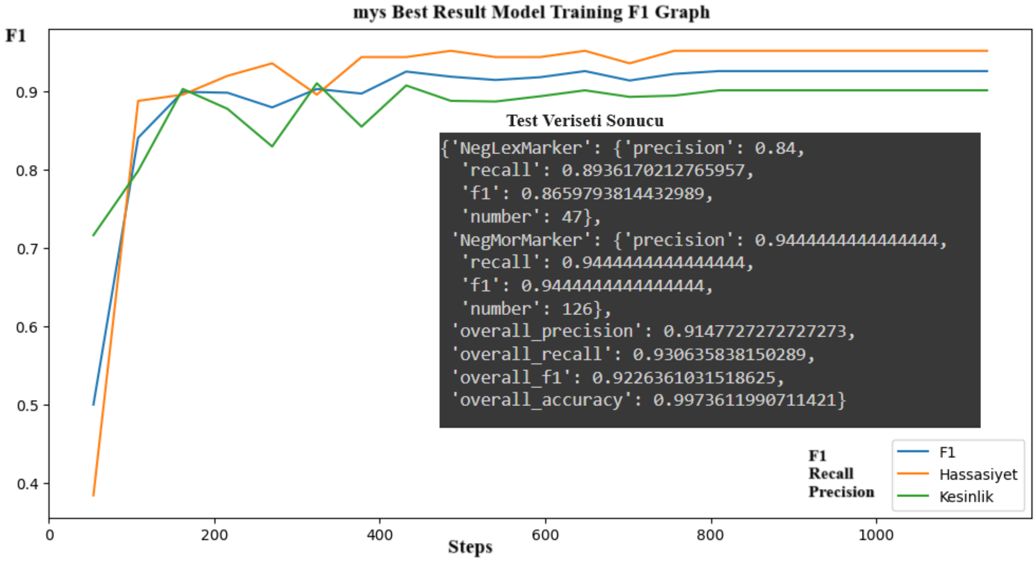


Figure 6.3: Best Cue Model F1 Graph

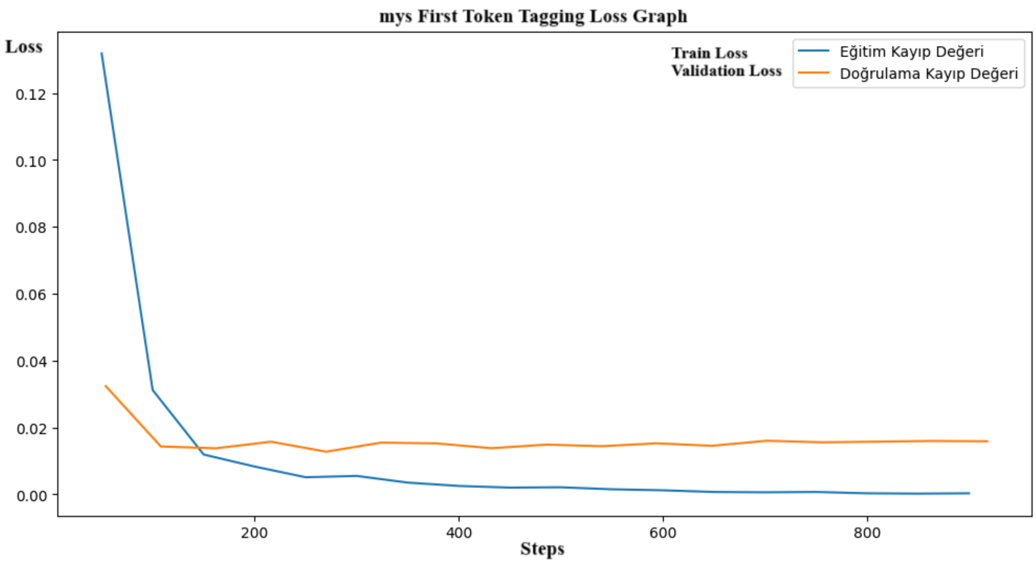


Figure 6.4: Best Cue Model Loss Graph

It is observed that model has high prediction rate and learning precision and recall turn wise manner in sequence (Figure 6.3 and Figure 6.4). These two component measures initially moved in inverse proportion, but increased with the common F1 score after learning.

The effect of the method used for label and tag length equalization on the model result can be clearly seen when looking at the models with the highest measured values. The highest results, which is first token tagging method in cue detection, can be explained with the effect of not taking other tokens into calculations. In addition, although the most appropriate tagging style for the correct representation of the negation during tagging is considered to be tagging with the same label, it is also seen in the literature that first tagging is used in different languages (see Section 5.4). Beyond labeling method, the best method for prefix rich languages is considered to be span detection. This case is noted on future work section. For feature detection task (ie. scope and event recognition), the additional cue label vector of the token is assign regarding to their labels. For this method PyTorch model have been utilized from (Rogge, 2020).

SCOPE TEST	Model	First Token Tagging	Same Label Tagging	Continuous Tagging
	savasy/bert-base-turkish-ner-cased			
Test Loss:		0.306	0.2788	0.2588
Test Accuracy:		0.928	0.9305	0.9213
Test F1:		0.7	0.7116	0.706
Test Recall:		0.7125	0.7353	0.7391
Test Precision:		0.69	0.694	0.679
	mys/electra-base-turkish-cased-ner			
Test Loss:		0.2732	0.2945	0.2723
Test Accuracy:		0.9263	0.9219	0.9066
Test F1:		0.667	0.7108	0.517
Test Recall:		0.6409	0.7508	0.5491
Test Precision:		0.6995	0.6796	0.4948
	alpcansoydas/turkish-ner-bert-base			
Test Loss:		0.2926	0.3556	0.2302
Test Accuracy:		0.9305	0.922	0.9248
Test F1:		0.709	0.69	0.5484
Test Recall:		0.6884	0.6835	0.5363
Test Precision:		0.7352	0.7052	0.8703
	liatoutou/distilbert-base-uncased-finetuned-negation-scope-classification			
Test Loss:		0.3047	0.3573	0.3334
Test Accuracy:		0.9141	0.9033	0.8944
Test F1:		0.5744	0.5326	0.6175
Test Recall:		0.5394	0.5022	0.592
Test Precision:		0.7451	0.6686	0.6576

Figure 6.5: Scope Test Results

Models seems to memorize the pattern and after nearly 5 epochs the model start over fitting. At these breakpoints, it is observed that precision

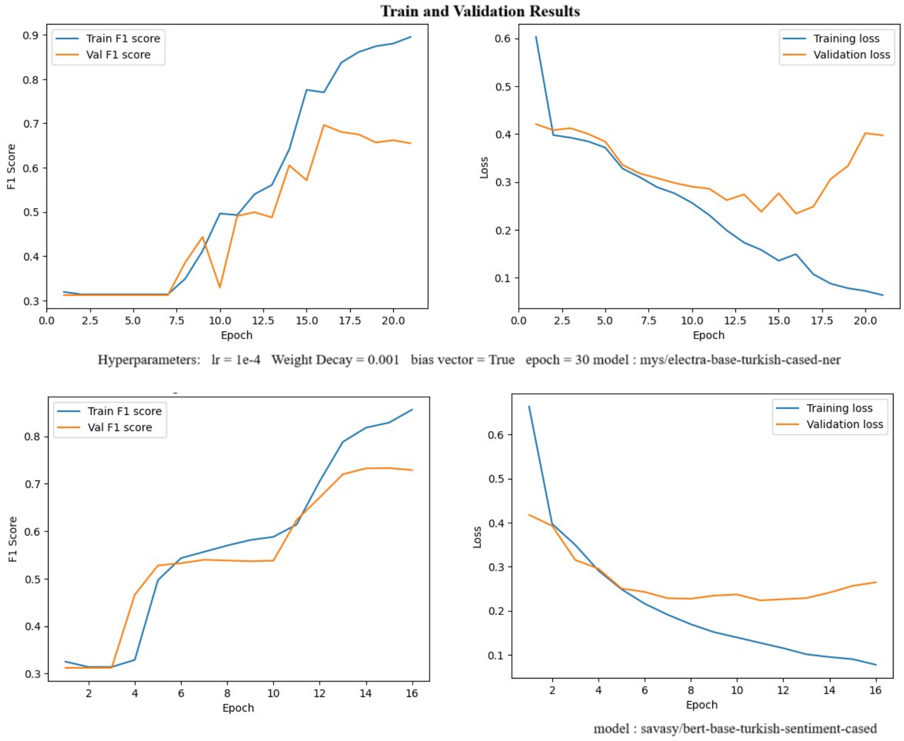


Figure 6.6: Scope Detection Best Results

and accuracy are tried to be learned alternately. As expected, the model measurement values are lower than the cue model result despite increasing the number of layers. The results are evaluated in Section 7. It is estimated that using larger models against model complexity does not provide much benefit to the system without increasing the capacity of the dataset. Also it is observed a big difference between cue and feature recognition.

Event prediction results are just quite low respect to scope prediction on (Mys, 2023) model. The F1 scores and other metrics are nearly same. As Figure 6.5 and Figure 6.7 shows in both tasks the first token tagging method has lower scores respect to other methods opposite to cue prediction task. Main result thought to be the dataset size as scope and event tasks have more complexity which is addressed in scope ambiguity (see Chapter 7.3.1).

EVENT	Model	First Token Tagging	Same Label Tagging	Continious Tagging
	savasy/bert-base-turkish-ner-cased			
Test Loss:		0.0502	0.0742	0.046
Test Accuracy:		0.9817	0.9839	0.9855
Test F1:		0.499	0.5247	0.657
Test Recall:		0.4725	0.5576	0.8087
Test Precision:		0.538	0.501	0.5877
	mys/electra-base-turkish-cased-ner			
Test Loss:		0.0659	0.0728	0.0649
Test Accuracy:		0.9835	0.981	0.9797
Test F1:		0.6067	0.5067	0.6877
Test Recall:		0.6188	0.4988	0.7114
Test Precision:		0.6047	0.5155	0.6707
	alpcansoydas/turkish-ner-bert-base			
Test Loss:		0.0472	0.0658	0.0982
Test Accuracy:		0.9843	0.9837	0.9786
Test F1:		0.5112	0.5041	0.6392
Test Recall:		0.5051	0.5771	0.6606
Test Precision:		0.5176	0.4661	0.6392
	liatoutou/distilbert-base-uncased-finetuned-negation-scope-classification			
Test Loss:		0.099	0.1096	0.0957
Test Accuracy:		0.9811	0.9734	0.9761
Test F1:		0.4876	0.5272	0.657
Test Recall:		0.5444	0.5604	0.6956
Test Precision:		0.4682	0.5073	0.6264

Figure 6.7: Event Detection Test Results

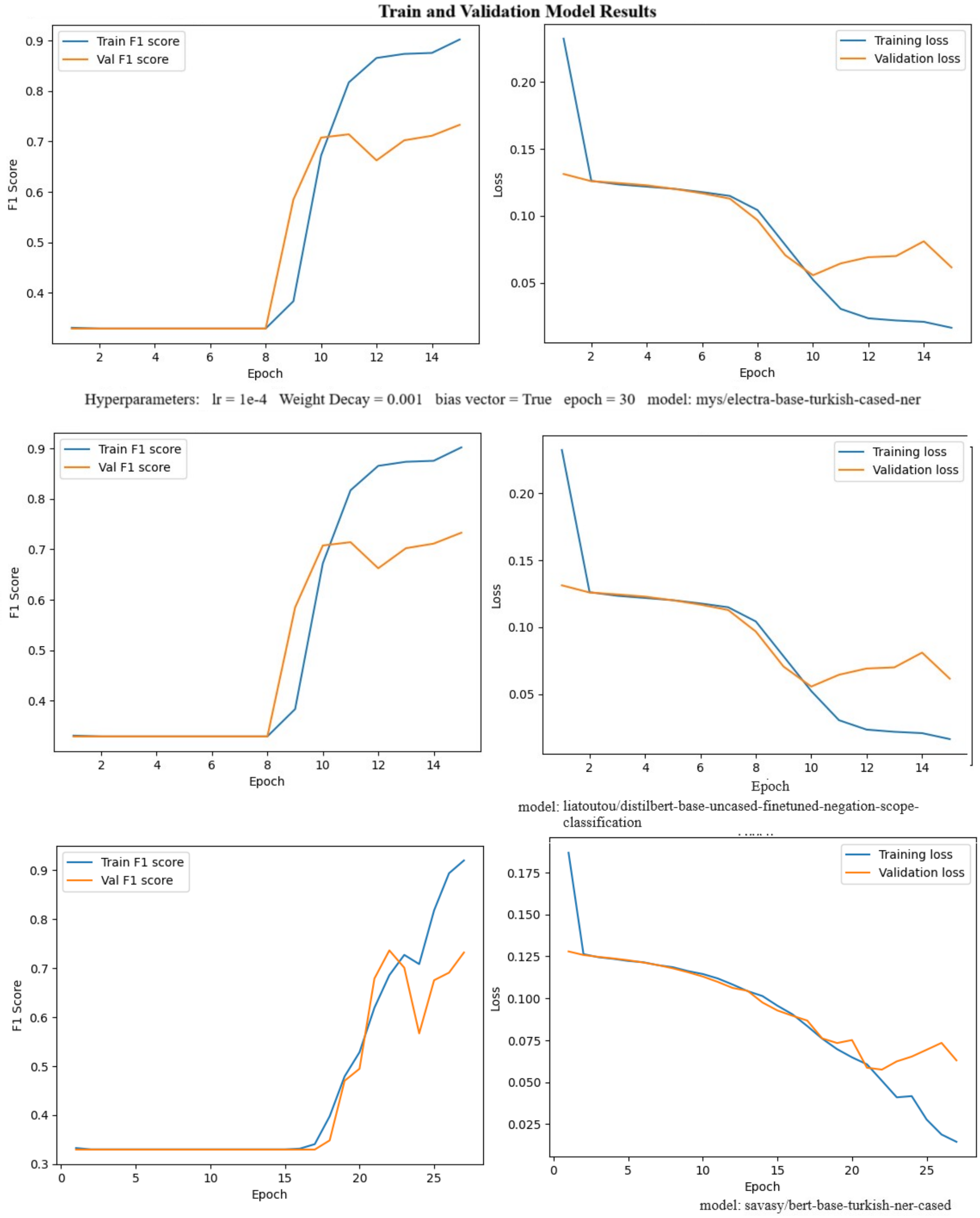


Figure 6.8: Event Detection Best Results

Figure 6.9: Confusion Matrix of Scope and Event

7 DISCUSSION

7.1 Result Evaluation

As a result of the literature survey, it was observed that the existing models and negation studies use first token tagging (NUBes BERT (Pabón ve Montenegro, 2021)) and same label tagging methods. For this reason, it can be seen that the cue models give good results despite the mentioned limitations. In the same-label tagging method and the continuous labeling method, the system becomes more complex. In these models, it can be said that RoBERTa and DistilBERT models show high success in cues. The problem of scope and event detection is more complex than that marker. For this reason, the number of layers of the classifier head was increased to 2 and 3 respectively. However, even though the batch size was increased to 128, it was observed that the models had difficulty in finding the scope complexity. Since increasing the number of layers would impose computational constraints, the learning coefficient and weighted reduction ratios were tested. In the scope and event detection problems, it was observed that the event definition was a more complex problem than the scope boundary. Also to prevent fast learning, memorizing and over-fitting; drop out parameter is set to 0,4. Although this effect positive to model metrics, still not boost the correct prediction. It was thought that the dataset size may not enough to see such configuration results clearly. Also span prediction models may have better insights for prefix reach languages.

7.2 Future Works

For the future work, the most important thing to be done is to expand the dataset as an academic study group in order to reduce the limitations mentioned in the final evaluation. In addition, in the light of the test findings, it is suggested that the number of parameters of the model should be expanded, the number of neurons in the layer should be increased, and more comprehensive research should be carried out on Turkish models. It is also important to emphasize that only pre-trained transfer learning was used in this study. The encoder model is not trained over the corpus again. No domain adaptation applied. For this reason, it is planned to apply unsupervised training method with unlabeled health articles in addition to the increased dataset by restructuring the transformers mechanism. Thus, it is thought that the use of domain adaptation method in addition to transfer learning can improve the model results. For this reason, testing the domain adaptation method is also included in future studies. Finally, a pipeline is envisioned in which these two separate models are sequentially added one after the other.

7.3 Restrictions

7.3.1 Scope Ambiguity

Scope ambiguity refers to the ambiguity in the grammatical rules found in all languages, or the ambiguity in the grammatical rules for representation of an expression as a whole in a sentence. Scope ambiguity is specially discussing topic in prefix rich languages where semantic and structure bonds are done with prefixes such as Turkish, Hungarian, Japanese. In their study, (Han ve ark., 2004) report that the scope ambiguity is a problematic issue on negation. In this paper, also the referenced thesis study summarize scope ambiguity issue by pointing unclear resulting image. In Turkish, adjectives etc. can be inserted between phrases connected by verbs or fused with phrasal suffixes. This situation, as well as the attachment of clauses or phrases to the sentence with affixes, leads to uncertainty as to whether the semantic relation continues structurally. Which elements of the sentence are included in the scope is also discussed in the literature. In addition, in a study on Hungarian, it is stated that it is unclear whether the scope of negation is semantic or structural, and that the scope boundaries can be accepted in both classes according to linguists (Wedgwood, 2006).

7.3.2 Auxiliary Verb Ambiguity

Compound verbs are verb types that are formed by the combination of a noun and an auxiliary verb, two separate verb forms or one or more words with a noun family and a main verb and correspond to a single concept. The phrases *"et-"*, *"ettir-"*, *"edil-"*, *"edin-"*, *"eyle-"*, *"ol-"*, *"olun-"*, *"bul-"* etc. can be used as a main verb with a phrase. For this reason, it is sometimes unclear when these words are used as auxiliary verbs.

7.3.3 Multiple Negation

As shown in Figure 4.2, a sentence can contain up to 4 negation in corpora. This is also shows how complex sentences can be. The method discussed in Chapter 5 applied and for feature models the sentences copied regarding to their cue number. By this, study aim to use single cue task for each round and reduce complexity.

[Son vakamızda [*kitle* tam olarak **çıkar**tılmış] ve [*metastaz* da saptan**ma**dığı için] kemoterapi *ver*il**me**miştir.]

Turkish dataset observed to have high complexity with respect to other medical reports. Also the Turkish dataset is prepared over a general medical articles. Not prepared in a constructure structure but free typing articles. While the BioScope dataset contains at most one marker per sentence, NUBes states that the average number of affected sentences per marker is 1.25. In the used dataset, 613 signs were marked in a total of 1950 sentences, and the ratio of signs per negated sentence was calculated as 1.143. The ration would be much higher when it calculated for the sentences that has at least one negation.

7.3.4 Sentence Structures and Misspellings

Turkish is an agglutinative language. Although this feature enriches the flexibility of the language and the dense meaningful sentences, it can bring complexity problems. In addition, suffixes can be combined with sentences to form semantically dense, long and compound sentences. The examples in the dataset show errors such as misspellings, incorrect use of affixes, and incorrect Turkish adaptations of foreign words. These can be human mistake or lack of time to fill or write reports in the field.

7.3.5 Model Restrictions

When calculating loss, it is more difficult for the model to find a pattern in datasets where negation are expressed affixes or prefixes respect to syntactical negation. One reason for this is that the same affixes may be used in different meanings or tasks. In the WE method, in order to correctly assign tags to affixes, tokenization must produce tokens that can be completely (or nearly completely) separated into roots and affixes. However, unfortunately, Hugginface models (for Turkish) do not have tokenizer models that can differentiate between roots and words at an acceptable level. Although there are valuable studies on it, tokenization of Turkish phrases has an urgent need to improve. Deep learning models are proficient at inferring meaning on many features, albeit with incomplete separation of components.However, they still need further development in order to be able to handle additional variation and, in the next step, to be able to clearly identify the scope and the relationship between the event and the cue. One way to overcome this is to use more data. At this point, labeling methods were investigated and the 3 methods tested on two phase models. Results discussed in Chapter 6.

[Pnömokok aşısının yapıl**ma**dığı] öğrenildi.

It was learned that [the pneumococcal vaccine was **not** *administered*].

8 APPENDIX

8.1 Scope Annotation

(Morante ve Blanco, 2021) states that the scope of negation is the part of the meaning that is negated. Identifying negation scopes have different challenges in each language. (Pröllochs ve ark., 2020) point to the ambiguity of scope and the fact that negation scopes are difficult to define because they are latent, unobservable and highly subjective - even among experts. Because of this, the dataset (produced for such cases) should include a guideline to define a general consensus for annotators. The Bioscope (Vincze ve ark., 2008), (Morante ve ark., 2011), The SFU Review of Negation (Jiménez-Zafra ve ark., 2017), (Emilie, 2018) and (Lima ve ark., 2020) are all include a short or expanded annotation rules.

The Turkish dataset which is still under process have an early release guideline. In order to create a comprehensive guideline, linguistic studies have been analyzed and taken together with Turkish grammar and rules. In this study, the scope can be extended to as many affected syntactic units as possible. Thus, the scope is defined as the maximum possible unit that contains the possible keywords of the negation concept. Turkish is a Subject-Object-Verb structured language. With this basic structure, the meaning can be extend with prefixes to include more dense semantic with combining phrases or complex logical propositions. Although this is a rich language feature, the negation structure and features come with ambiguity, as in similar languages such as Japanese and Hungarian (Wedgwood, 2006). Similar to Turkish, studies in Japanese, which is an agglutinative language, have debated whether the suffixes that come after the sign should be included or not. (Han ve ark., 2004) reported that linguists in their studies stated that both methods can be based on various points. Scope ambiguity in Turkish is similar along this axis. Negation is also marked in the phrases used in parentheses in the sentence.Some of the scope annotation rule can be listed as:

➢ If the negation marker is in the main predicate of the sentence, the scope is the whole sentence.

[Hastanın olgularında herhangi bir bulgu saptan]**ma**[dı].

➢ Cues in double negation are not tagged and are not assigned as a negation if their scopes are same.

[**A***normal* tutulum] ile osteopoikiloz tanısı dışlanmaz.

In this example, *"dışla-"* is a word (lexical) negation and a verb and *"-maz"* is a morhological negation prefix. The scope of both two negation cues are the same except for themselves. They are in the event-determining element of the sentence.

➢ For the negations in the nominal phrase of a sentence or clause, the scope is found in the nominal phrase.

[Cerrahi girişimlerin *yet*er]**siz** [kaldığı durumlarda] protetik uygulamalar ile de başarılı sonuçlar elde edilebilmektedir.

➢ If the negation suffix is attached to an adverb that qualifies the verb, the negation covers the whole sentence. If the adverb is a verb phrase, it covers only the phrase. This rule is same with the study of (Vincze ve ark., 2011).

[Akciğer grafisi normal olan hastanın *PPD*' si] **negatif** [saptandı].

Bu durum, [test değerleri **negatif** olan hastanın grafisinde] gözlemlendi.

➢ “Nominal phrases” refer to groups of words, usually nouns, that are connected to each other. In these phrases, a noun or noun phrase is usually taken as the main element. Other nouns, adjectives or determiners may complement the phrase. If adjective verbs have a negation suffix, the scope can extend to qualifying words. "The scope of qualifying adjectives usually extends to the following noun phrase." (Vincze ve ark., 2008). The directive is parallel to this rule. For detailed rules defined on phrases, reference can be made to the labeling guideline of the dataset study.

➢ Derivational suffixes following negation clues are not included. Since words with a construction suffix are new words, negation only affects the meaning of the old word.

Bruselloz endemik bölgelerde ciddi morbiditeye neden olabileceğinden erken tanı ve tedavi komplikasyonların önlenmesi açısından önemli olmakla birlikte hastalığın çoğu zaman ateş yüksekliği, eklem ağrısı, kas ağrısı, sırt ağrısı, [*hal*]**siz**lik, [*iştah*]**sız**lık ve yorgunluk gibi özgün olmayan belirtiler ile seyretmesi tanı konulmasını geciktirmektedir.

➢ If conjunctions occur within a scope, it takes over the scope and joins broken scopes. These conjunctions are represented on the dataset by the label coordination snippets.

➢ The scope does not include negated sentences containing conditional or prepositional conjunctions (if, whereas, so that, etc.). In other words, if there is a negation in one of the sentences connected to each other by conditional and prepositional conjunctions, these expressions are not included.

Hastanın genel durumunun aniden kötüleşmesi ve solunum sıkıntısı başlaması nedeniyle [nozokomiyal pnömoni *düşün*ül]**me**[mesine] rağmen meropenem tedavisine devam edildi.

➢ There may be unique parts used as a sequence and shared phrases within a phrase or sentence. An example of this would be adjectives within a noun phrase. In such a case, the "disjunctive negation scope" can be applied. In the study's dataset, the frequency of these cases in Turkish is also quite high.

Ultraviyole radyasyon maruziyeti sonrası, ışığa maruz kalan deride, ataklar şeklinde gelişen [skar *yap***ma**yan kaşıntılı papüller, veziküller ve plaklar] ile tanımlanır.

8.2 Event Annotation

Event in negation is generally studied in Biomedical areas (Mahany ve ark., 2022).These domains include many medical or non-medical phenomena and in most cases they are verbs, nouns or adjectives (Kim ve ark., 2008). An event can refer to a process, an action or a situation. One of the most popular negation event annotation guideline is the GENIA (Ohta ve ark., 2007) guideline (see Table 2.1). In (Morante ve Daelemans, 2012), event annotated in Conan Doyle Stories. This study define event annotation in more general consensus respect to GENIA. The dataset study tags events regardless of whether they are medical or not in the same way as the ConnanDoyle-neg study.

- ➢ The event is the part directly affected by the negation marker. The smallest possible syntactic unit is tagged.
- ➢ If the negation marker is inside the parenthesis, the negation marker, scope and event are labeled in parentheses. If the negation sign is outside the parenthesis, only the scope is tagged in parentheses.
- ➢ The event is not labeled if the whole sentence contains future tense, necessity, possibility, imperative, command, prohibition, prediction and conditional expressions. There can be no certainty in sentences using these expressions.

- ➢ In negatives formed with morphological markers, the root of the word is tagged when tagging the event. This rule is not a factor in this study since word-based tag transformation is performed.

  [Sol overden biyopsi *al*ın]**ma**[dı].

- ➢ In negatives formed with lexical and syntactic markers, the root of the word is taken when labeling the event.Inflectional suffixes in the word are not tagged.

[PA AC grafisinde görülemeyen *patoloji*leri] **ekarte** [etmek] ...

- Linking prepositions connect words, phrases and sentences to create a unity of meaning.Prepositions such as “comma, and, with, or, or or” form a linking group.
- “If the name of the test in the negation marker “Negative” corresponds to something that exists, the event is not labeled" (Lima ve ark., 2020).

[Olgumuzun bakılan Widal testi] negatif [olarak sonuçlanmıştı].

References

Almeida, F. ve **Xexéo**, 2019, Word Embeddings: A Survey, *CoRR*, abs/1901.09069.

Althari, G. ve **Alsulmi, M.**, 2022, Exploring transformer-based learning for negation detection in biomedical texts, *IEEE Access*, 10, 83813–83825.

Aras, G., **Makaroğlu, D.**, **Demir, S.**, ve **Cakir, A.**, 2021, An evaluation of recent neural sequence tagging models in Turkish named entity recognition, *Expert Systems with Applications*, 182, 115049, ISSN 0957-4174, doi:https://doi.org/10.1016/j.eswa.2021.115049.

Argüello-González, G., **Aquino-Esperanza, J.**, **Salvador, D.**, **Bretón-Romero, R.**, **Del Río-Bermudez, C.**, **Tello, J.**, ve **Menke, S.**, 2023, Negation recognition in clinical natural language processing using a combination of the NegEx algorithm and a convolutional neural network, *BMC Medical Informatics and Decision Making*, 23(1), doi:10.1186/s12911-023-02301-5.

Aronow, D.B., **Fangfang, F.**, ve **Croft, W.B.**, 1999, Ad Hoc Classification of Radiology Reports, *Journal of the American Medical Informatics Association*, 6(5), 393–411, doi:10.1136/jamia.1999.0060393.

Azunre, P., 2021, *Transfer Learning for Natural Language Processing*, Simon and Schuster.

Banjade, R. ve **Rus, V.**, 2016, DT-Neg: Tutorial Dialogues Annotated for Negation Scope and Focus in Context, in N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, ve S. Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3768–3771, European Language Resources Association (ELRA), Portorož, Slovenia.

Batet, M. ve **Grau, L.**, 1995, The acquisition of negation in English, *Atlantis: Revista de la Asociación Española de Estudios Anglo-Norteamericanos*, 17(1-2), 27–44, ISSN 0210-6124.

Bel-Enguix, G., **Gómez-Adorno, H.**, **Pimentel, A.**, **Ojeda-Trueba, S.L.**, ve **Aguilar-Vizuet, B.**, 2021, Negation Detection on Mexican Spanish Tweets: The T-MexNeg Corpus, *Applied Sciences*, 11(9), ISSN 2076-3417, doi:10.3390/app11093880.

Beukeboom, C.J., **Burgers, C.**, **Szabó, Z.P.**, **Cvejic, S.**, **Lönnqvist, J.E.M.**, ve **Welbers, K.**, 2019, The Negation Bias in Stereotype Maintenance: A Replication in Five Languages, *Journal of Language and Social Psychology*, 39(2), 219–236, doi:10.1177/0261927x19869759.

Blanco, E. ve **Moldovan, D.**, 2011, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 581–589, Association for Computational Linguistics, Portland, Oregon.

Bos, J., **Basile, V.**, **Evang, K.**, **Venhuizen, N.**, ve **Bjerva, J.**, 2017, The Groningen Meaning Bank, in N. Ide ve J. Pustejovsky, editors, *Handbook of Linguistic Annotation*, volume 2, pages 463–496, Springer.

Buroker, J.V., 1993, The Port-Royal semantics of terms, *Synthese*, 96(3), 455–475, doi:10.1007/bf01064012.

Chapman, W.W., **Bridewell, W.**, **Hanbury, P.**, **Cooper, G.F.**, ve **Buchanan, B.G.**, 2001, A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries, *Journal of Biomedical Informatics*, 34(5), 301–310, doi:10.1006/jbin.2001.1029.

Chen, L., 2019, Attention-Based Deep Learning System for Negation and Assertion Detection in Clinical Notes, *International Journal of Artificial Intelligence & Applications*, 10(01), 1–9, doi:10.5121/ijaia.2019.10101.

Chollet, F., 2021, *Deep Learning with Python, Second Edition*, Simon and Schuster.

Chomsky, N., 1957, *Syntactic Structures*, Mouton.

Çöltekin, Ç., **Doğruöz, A.S.**, ve **Özlem Çetinoğlu**, 2022, Resources for Turkish natural language processing: A critical survey, *Language Resources and Evaluation*, 57(1), 449–488, doi:10.1007/s10579-022-09605-4.

Corrigan, Kevin ve Harrington, L.M., 2023, Pseudo-Dionysius the Areopagite, in E.N.Z. ve Uri Nodelman, editor, *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, Summer 2023 edition, [Accessed 12.03.2023].

Councill, I., **McDonald, R.**, ve **Velikovich, L.**, 2010, What's great and what's not: learning to classify the scope of negation for improved sentiment analysis, in R. Morante ve C. Sporleder, editors, *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 51–59, University of Antwerp, Uppsala, Sweden.

Cruz Díaz, N.P., **Maña López, M.J.**, **Vázquez, J.M.**, ve **Álvarez, V.P.**, 2012, A machine-learning approach to negation and speculation detection in clinical texts, *Journal of the American Society for Information Science and Technology*, 63(7), 1398–1410, doi:10.1002/asi.22679.

Dalloux, C., **Claveau, V.**, ve **Grabar, N.**, 2019, Speculation and Negation detection in French biomedical corpora, in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 223–232, INCOMA Ltd., Varna, Bulgaria, doi:10.26615/978-954-452-056-4_026.

Dalloux, C., **Claveau, V.**, **Grabar, N.**, **Oliveira, L.E.S.**, **Moro, C.M.C.**, **Gumiel, Y.B.**, ve **Carvalho, D.R.**, 2021, Supervised learning for the detection of negation and of its scope in French and Brazilian Portuguese biomedical corpora, *Natural Language Engineering*, 27(2), 181–201, doi:10.1017/S1351324920000352.

Dehkharghani, R., **Saygin, Y.**, **Yanikoglu, B.**, ve **Oflazer, K.**, 2015, SentiTurkNet: a Turkish polarity lexicon for sentiment analysis, *Language Resources and Evaluation*, 50(3), 667–685, doi:10.1007/s10579-015-9307-6.

Devlin, J., **Chang, M.W.**, **Lee, K.**, ve **Toutanova, K.**, 2019, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Diaz, N.C., 2014, Negation and Speculation Detection in Clinical and Review Texts. Ph.D. dissertation awarded the SEPLN (Spanish Association for Natural Language Processing) 2014 award for the Best Research in Natural Language Processing, doi:10.13140/2.1.2061.4400.

Durango, M.C., **Torres-Silva, E.A.**, ve **Orozco-Duque, A.**, 2023, Named Entity Recognition in Electronic Health Records: A Methodological Review, *Healthcare Informatics Research*, 29(4), 286–300, doi:10.4258/hir.2023.29.4.286.

Emeksiz Erk, Z., 2010, Negation in Turkish, *Dilbilim Arastırmaları*.

Emilie, 2018, SFU-Discourse-Lab/SOCC: SFU Opinion and Comments Corpus.

Enger, M., **Velldal, E.**, ve **Øvrelid, L.**, 2017, An open-source tool for negation detection: a maximum-margin approach, in E. Blanco, R. Morante, ve R. Saurí, editors, *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*, pages 64–69, Association for Computational Linguistics, Valencia, Spain, doi:10.18653/v1/W17-1810.

Erguvanlı, E.E., 1984, *The Function of Word Order in Turkish Grammar*, volume 106, University of California Press.

Eroğul, U., 2009, *Sentiment Analysis in Turkish*, Master's thesis, Middle East Technical University.

Fabregat, H., **Martinez-Romo, J.**, ve **Araujo, L.**, 2018, Deep learning approach for negation cues detection in Spanish, in *CEUR Workshop Proceedings*, volume 2174, pages 43–48, CEUR-WS, ISSN 1613-0073.

Fancellu, F., **Lopez, A.**, ve **Webber, B.**, 2016, Neural Networks For Negation Scope Detection, in K. Erk ve N.A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 495–504, Association for Computational Linguistics, Berlin, Germany, doi:10.18653/v1/P16-1047.

Fancellu, F., **Lopez, A.**, ve **Webber, B.**, 2018, Neural Networks for Cross-lingual Negation Scope Detection.

Fancellu, F., **Lopez, A.**, **Webber, B.**, ve **He, H.**, 2017, Detecting negation scope is easy, except when it isn't, in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 58–63, Association for Computational Linguistics, Valencia, Spain.

Farkas, R., **Vincze, V.**, **Móra, G.**, **Csirik, J.**, ve **Szarvas, G.**, 2010, The CoNLL-2010 Shared Task: Learning to Detect Hedges and Their Scope in Natural Language Text, in *Proceedings of the Fourteenth Conference on Computational Natural Language Learning — Shared Task*, CoNLL '10: Shared Task, page 1–12, Association for Computational Linguistics, USA, ISBN 9781932432848.

Fei, H., **Ren, Y.**, ve **Ji, D.**, 2020, Negation and speculation scope detection using recursive neural conditional random fields, *Neurocomputing*, 374, 22–29, ISSN 0925-2312, doi:https://doi.org/10.1016/j.neucom.2019.09.058.

Flickinger, D., **Zhang, Y.**, ve **Kordoni, V.**, 2012, Deepbank: A Dynamically Annotated Treebank of the Wall Street Journal, in *Proceedings of the Eleventh International Workshop on Treebanks and Linguistic Theories*, pages 85–96, Edições Colibri.

Giora, R., **Fein, O.**, **Ganzi, J.T.**, **Levi, N.A.**, ve **Sabah, H.**, 2005, On Negation as Mitigation: The Case of Negative Irony, *Discourse Processes*, 39, 100 – 81.

Grinde, A.B. ve **Johansen, B.M.**, 2021, *Using Natural Language Processing with Deep Learning to Explore Clinical Notes*, Master's thesis, The University of Bergen.

Görgülü, E., 2017, Negative Polarity in Turkish: from negation to nonveridicality, *Macrolinguistics*, 5(7), 51–69, doi:10.26478/ja2017.5.7.3.

Han, C.h., **Storoshenko, D.R.**, ve **Sakurai, Y.**, 2004, Scope of Negation and Clause Structure in Japanese, *Annual Meeting of the Berkeley Linguistics Society*, 30(1), 118, doi:10.3765/bls.v30i1.955.

Hao, T., **Huang, Z.**, **Liang, L.**, **Weng, H.**, ve **Tang, B.**, 2021, Health Natural Language Processing: Methodology Development and Applications, *JMIR Medical Informatics*, 9(10), e23898, doi:10.2196/23898.

Harkema, H., **Dowling, J.N.**, **Thornblade, T.**, ve **Chapman, W.W.**, 2009, ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports, *Journal of Biomedical Informatics*, 42(5), 839–851, ISSN 1532-0464, doi:https://doi.org/10.1016/j.jbi.2009.05.002, biomedical Natural Language Processing.

Hartmann, M. ve **Søgaard, A.**, 2021, Multilingual Negation Scope Resolution for Clinical Text, in E. Holderness, A. Jimeno Yepes, A. Lavelli, A.L. Minard, J. Pustejovsky, ve F. Rinaldi, editors, *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*, pages 7–18, Association for Computational Linguistics, online.

Honnibal, M. ve **Montani, I.**, 2017, spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Horn, L.R., 1989, *A Natural History of Negation*, University of Chicago Press.

Howard, J. ve **Ruder, S.**, 2018, Fine-tuned Language Models for Text Classification, *CoRR*, abs/1801.06146.

Howard, J. ve ark., 2018, fastai, `https://github.com/fastai/fastai`.

Huang, C.C. ve **Lu, Z.**, 2015, Community challenges in biomedical text mining over 10 years: success, failure and the future, *Briefings in Bioinformatics*, 17(1), 132–144, doi:10.1093/bib/bbv024.

Huang, Y. ve **Lowe, H.J.**, 2007, A novel hybrid approach to automated negation detection in clinical radiology reports, *Journal of the American medical informatics association*, 14(3), 304–311.

Huggingface, 2021, examples/token_classification, `https://huggingface.co/docs/transformers/tasks/token_classification` [Accessed: (01.010.2023)].

IBM, 2023, Natural Language Processing, [Erişim Tarihi: 21.03.2023].

Ince, A., 2011, Negation and Implicature in Turkish Child Language, in *Proceedings of the 35th Annual Boston University Conference on Language Development*, pages 284–296.

Jiménez-Zafra, S.M., **Taulé, M.**, **Martín-Valdivia, M.T.**, **Ureña-López, L.A.**, ve **Martí, M.A.**, 2017, SFU ReviewSP-NEG: a Spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns, *Language Resources and Evaluation*, 52(2), 533–569, doi: 10.1007/s10579-017-9391-x.

Kang, T., **Zhang, S.**, **Xu, N.**, **Wen, D.**, **Zhang, X.**, ve **Lei, J.**, 2017, Detecting negation and scope in Chinese clinical notes using character and word embedding, *Computer Methods and Programs in Biomedicine*, 140, 53–59, doi:10.1016/j.cmpb.2016.11.009.

Kelepir, M., 2001, *Topics in Turkish syntax: clausal structure and scope*, Ph.D. thesis, Massachusetts Institute of Technology, uRI: `http://dspace.mit.edu/handle/1721.1/8196`.

Khandelwal, A. ve **Sawant, S.**, 2020, NegBERT: A Transfer Learning Approach for Negation Detection and Scope Resolution, in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5739–5748, European Language Resources Association, Marseille, France, ISBN 979-10-95546-34-4.

Kim, J.D., **Ohta, T.**, ve **Tsujii, J.**, 2008, Corpus annotation for mining biomedical events from literature, *BMC Bioinformatics*, 9(1), doi:10.1186/1471-2105-9-10.

Klie, J.C., **Bugert, M.**, **Boullosa, B.**, **de Castilho, R.E.**, ve **Gurevych, I.**, 2018, The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation, in *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Association for Computational Linguistics, event Title: The 27th International Conference on Computational Linguistics (COLING 2018).

Klie, J.C. ve **de Castilho, R.E.**, 2021, *dkpro-cassis*, doi:10.5281/zenodo.3994108.

Klima, E., 1964, Negation in English, in J. Fodor ve J. Katz, editors, *The Structure of Language: Readings in the Philosophy of Language*, Prentice-Hall.

Kolhatkar, V., **Wu, H.**, **Cavasso, L.**, **Francis, E.**, **Shukla, K.**, ve **Taboada, M.**, 2019, The SFU Opinion and Comments Corpus: A Corpus for the Analysis of Online News Comments, *Corpus Pragmatics*, 4(2), 155–190, doi:10.1007/s41701-019-00065-w.

Lazib, L., **Qin, B.**, **Zhao, Y.**, **Zhang, W.**, ve **Liu, T.**, 2020, A syntactic path-based hybrid neural network for negation scope detection, *Frontiers of computer science*, 14, 84–94.

Lazib, L., **Zhao, Y.**, **Qin, B.**, ve **Liu, T.**, 2019, Negation scope detection with recurrent neural networks models in review texts, *International Journal* *of High Performance Computing and Networking*, 13(2), 211, doi:10.1504/ijhpcn.2019.097501.

Li, H. ve **Lu, W.**, 2018, Learning with Structured Representations for Negation Scope Extraction, in I. Gurevych ve Y. Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 533–539, Association for Computational Linguistics, Melbourne, Australia, doi:10.18653/v1/P18-2085.

Liatoutou, 2023, distilbert-base-uncased-finetuned-negation-scope-classification.

Lima, S., **Pérez, N.**, **Cuadros, M.**, ve **Rigau, G.**, 2020, NUBes: A Corpus of Negation and Uncertainty in Spanish Clinical Texts, in *International Conference on Language Resources and Evaluation*.

Lin, C., **Bethard, S.**, **Dligach, D.**, **Sadeque, F.**, **Savova, G.**, ve **Miller, T.A.**, 2020, Does BERT need domain adaptation for clinical negation detection?, *Journal of the American Medical Informatics Association*, 27(4), 584–591, ISSN 1527-974X, doi:10.1093/jamia/ocaa001.

Loodos, 2023, bert-base-turkish-uncased.

Mahany, A., **Fouad, M.M.**, **Aloraini, A.**, **Khaled, H.**, **Nawaz, R.**, **Aljohani, N.R.**, ve **Ghoniemy, S.**, 2021, Supervised Learning for Negation Scope Detection n Arabic Texts, in *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pages 177–182, doi:10.1109/ICICIS52592.2021.9694175.

Mahany, A., **Khaled, H.**, **Elmitwally, N.S.**, **Aljohani, N.**, ve **Ghoniemy, S.**, 2022, Negation and Speculation in NLP: A Survey, Corpora, Methods, and Applications, *Applied Sciences*, 12(10), ISSN 2076-3417, doi:10.3390/app12105209.

McKenna, N. ve **Steedman, M.**, 2020, Learning Negation Scope from Syntactic Structure, in I. Gurevych, M. Apidianaki, ve M. Faruqui, editors, *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics*, pages 137–142, Association for Computational Linguistics, Barcelona, Spain (Online).

Messikommer, N., **Gehrig, D.**, **Gehrig**, ve **Scaramuzza, D.**, 2021, Bridging the Gap between Events and Frames through Unsupervised Domain Adaptation, *CoRR*, abs/2109.02618.

Morante, R. ve **Blanco, E.**, 2021, Recent advances in processing negation, *Natural Language Engineering*, 27(2), 121–130, doi:10.1017/S1351324920000534.

Morante, R. ve **Daelemans, W.**, 2009, A Metalearning Approach to Processing the Scope of Negation, in S. Stevenson ve X. Carreras, editors, *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 21–29, Association for Computational Linguistics, Boulder, Colorado.

Morante, R. ve **Daelemans, W.**, 2012, ConanDoyle-neg: Annotation of negation cues and their scope in Conan Doyle stories, in N. Calzolari, K. Choukri, T. Declerck, M.U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, ve S. Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1563–1568, European Language Resources Association (ELRA), Istanbul, Turkey.

Morante, R., **Schrauwen, S.**, ve **Daelemans, W.**, 2011, Annotation of negation cues and their scope: Guidelines v1.0., [Accessed 21.01.2023].

Morante, R. ve **Sporleder, C.**, 2012, Modality and Negation: An Introduction to the Special Issue, *Computational Linguistics*, 38(2), 223–260, doi:10.1162/coli_a_00095.

Mukherjee, P., **Leroy, G.**, **Kauchak, D.**, **Rajanarayanan, S.**, **Diaz, D.Y.R.**, **Yuan, N.P.**, **Pritchard, T.G.**, ve **Colina, S.**, 2017, NegAIT: A new parser for medical text simplification using morphological, sentential and double negation, *Journal of Biomedical Informatics*, 69, 55–62, doi: 10.1016/j.jbi.2017.03.014.

Mutalik, P.G., **Deshpande, A.**, ve **Nadkarni, P.M.**, 2001, Use of General-purpose Negation Detection to Augment Concept Indexing of Medical Documents: A Quantitative Study Using the UMLS, *Journal of the American Medical Informatics Association*, 8(6), 598–609, doi:10.1136/jamia.2001.0080598.

Mys, 2023, electra-base-turkish-cased-ner.

Nadir, I. ve **Kabadayı, C.**, 2015, *Türk dilinde olumsuzluk*, Kesit Yayınları.

Nawaz, R., **Thompson, P.**, ve **Ananiadou, S.**, 2013, Negated bio-events: analysis and identification, *BMC Bioinformatics*, 14(1), doi:10.1186/1471-2105-14-14.

Niels, R., 2020, Transformers Tutorials.

Ohta, T., **Kim, J.D.**, ve **Tsujii, J.**, 2007, Guidelines for event annotation, *Department of Information Science, Graduate School of Science, University of Tokyo*.

OpenAI, 2023, GPT-4 Technical Report, *OpenAI Technical Report*, accessed: 2024-05-20.

Ozcelik, O. ve **Toraman, C.**, 2022, Named entity recognition in Turkish: A comparative study with detailed error analysis, *Information Processing & Management*, 59(6), 103065, ISSN 0306-4573, doi:https://doi.org/10.1016/j.ipm.2022.103065.

Pabón, O.S., **Montenegro, O.**, **Torrente, M.**, **González, A.R.**, **Provencio, M.**, ve **Ruiz, E.M.**, 2022, Negation and uncertainty detection in clinical texts written in Spanish: a deep learning-based approach, *PeerJ Computer Science*, 8.

Pabón, O.S. ve **Montenegro, O.**, 2021, Transformers Tutorials.

Pagel, M., 2017, Q&A: What is human language, when did it evolve and why should we care?, *BMC Biology*, 15(1), doi:10.1186/s12915-017-0405-3.

Perreault, C. ve **Mathew, S.**, 2012, Dating the Origin of Language Using Phonemic Diversity, *PLoS ONE*, 7(4), e35289, doi:10.1371/journal.pone.0035289.

Pröllochs, N., **Feuerriegel, S.**, **Lutz, B.**, ve **Neumann, D.**, 2020, Negation scope detection for sentiment analysis: A reinforcement learning framework for replicating human interpretations, *Information Sciences*, 536, 205–221, ISSN 0020-0255, doi:https://doi.org/10.1016/j.ins.2020.05.022.

Pyysalo, S., **Ginter, F.**, **Heimonen, J.**, **Björne, J.**, **Boberg, J.**, **Järvinen, J.**, ve **Salakoski, T.**, 2007, BioInfer: a corpus for information extraction in the biomedical domain, doi:10.1186/1471-2105-8-50.

Qian, Z., **Li, P.**, **Zhu, Q.**, **Zhou, G.**, **Luo, Z.**, ve **Luo, W.**, 2016, Speculation and Negation Scope Detection via Convolutional Neural Networks, in J. Su, K. Duh, ve X. Carreras, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 815–825, Association for Computational Linguistics, Austin, Texas, doi:10.18653/v1/D16-1078.

Ren, Y., **Fei, H.**, ve **Peng, Q.**, 2018, Detecting the Scope of Negation and Speculation in Biomedical Texts by Using Recursive Neural Network, in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 739–742, doi:10.1109/BIBM.2018.8621261.

Roberts, E., 2023, An Overview of Natural Language Processing, `https://cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/nlp/overview_history.html`, [Accessed 21.03.2023].

Rogge, N., 2020, Transformers Tutorials.

Sergeeva, E., **Zhu, H.**, **Tahmasebi, A.**, ve **Szolovits, P.**, 2019, Neural Token Representations and Negation and Speculation Scope Detection in

Biomedical and General Domain Text, in E. Holderness, A. Jimeno Yepes, A. Lavelli, A.L. Minard, J. Pustejovsky, ve F. Rinaldi, editors, *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, pages 178–187, Association for Computational Linguistics, Hong Kong, doi:10.18653/v1/D19-6221.

Seydi, M., 2020, Existential Negation in Turkish: A Functional Approach, *Journal of Language and Linguistic Studies*, 16(2), 647–660.

Shi, J., **Gao, X.**, **Kinsman, W.C.**, **Ha, C.**, **Gao, G.G.**, ve **Chen, Y.**, 2022, Di++: A deep learning system for patient condition identification in clinical notes, *Artificial Intelligence in Medicine*, 123, 102224, doi:10.1016/j.artmed.2021.102224.

Shickel, B., **Tighe, P.J.**, **Bihorac, A.**, ve **Rashidi, P.**, 2018, Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis, *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589–1604, doi:10.1109/jbhi.2017.2767063.

Shoubhikc, 2023, negation_bert.

Siegel, J., 2007, Language Contact and Grammatical Change by Bernd Heine and Tania Kuteva, *Journal of Sociolinguistics*, 11(4), 549–552, doi:10.1111/j.1467-9841.2007.00337_9.x.

Singh, P.K. ve **Paul, S.**, 2021, Deep Learning Approach for negation handling in sentiment analysis, *IEEE Access*, 9, 102579–102592, doi:10.1109/access.2021.3095412.

Slater, L.T., **Bradlow, W.**, **Motti, D.F.**, **Hoehndorf, R.**, **Ball, S.**, ve **Gkoutos, G.V.**, 2021, A fast, accurate, and generalisable heuristic-based negation detection algorithm for clinical text, *Computers in Biology and Medicine*, 130, 104216, ISSN 0010-4825, doi:https://doi.org/10.1016/j.compbiomed.2021.104216.

Soydaş, A., 2023, turkish-ner-bert-base.

Speranza, J. ve **Horn, L.R.**, 2010, A brief history of negation, *Journal of Applied Logic*, 8(3), 277–301, doi:10.1016/j.jal.2010.04.001.

Stanford, S.U., 2017, *Exploring the Effects of External Semantic Data on Word Embeddings*.

Sykes, D., **Grivas, A.**, **Grover, C.**, **Tobin, R.**, **Sudlow, C.**, **Whiteley, W.**, **Mcintosh, A.**, **Whalley, H.**, ve **Alex, B.**, 2020, Comparison of rule-based and neural network models for negation detection in Radiology reports, *Natural Language Engineering*, 27(2), 203–224, doi:10.1017/s1351324920000509.

Tagliani, M., **Vender, M.**, ve **Melloni, C.**, 2022, The Acquisition of Negation in Italian, *Languages*, 7(2), ISSN 2226-471X, doi:10.3390/languages7020116.

Tamayo, A.J., 2023, Negation_Scope_Detection_NubEs_Training_Testing_mBERT_fine

Taner Akdeniz, 2023, bert-base-turkish-cased-ner (Revision b247a7f), doi:10.57967/hf/0949.

Thelwall, M., **Buckley, K.**, **Paltoglou, G.**, **Cai, D.**, ve **Kappas, A.**, 2010, Sentiment strength detection in short informal text, *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558, doi:10.1002/asi.21416.

Triess, L.T., **Dreissig, M.**, **Rist, C.B.**, ve **Zöllner, J.M.**, 2021, A Survey on Deep Domain Adaptation for LiDAR Perception, *arXiv (Cornell University)*, doi:10.48550/arxiv.2106.02377.

Trinh, D., 2021, Negation Learning in Transformer-based Neural Machine Translation Models.

Tunstall, L., **Von Werra, L.**, ve **Wolf, T.**, 2022, *Natural Language Processing with Transformers, Revised Edition*, "O'Reilly Media, Inc.".

Tura, S., 1981, A study on negation in Turkish, not published.

Turing, 2022, A Guide on Word Embeddings in NLP.

van Aken, B., **Trajanovska, I.**, **Siu, A.**, **Mayrdorfer, M.**, **Budde, K.**, ve **Loeser, A.**, 2021, Assertion Detection in Clinical Notes: Medical Language Models to the Rescue?, in *Proceedings of the Second Workshop on Natural Language Processing for Medical Conversations*, Association for Computational Linguistics, Online, doi:10.18653/v1/2021.nlpmc-1.5.

van Es, B., **Reteig, L.C.**, **Tan, S.C.**, **Schraagen, M.**, **Hemker, M.M.**, **Arends, S.R.S.**, **Rios, M.A.R.**, ve **Haitjema, S.**, 2023, Negation detection in Dutch clinical texts: an evaluation of rule-based and machine learning methods, *BMC Bioinformatics*, 24(1), doi:10.1186/s12859-022-05130-x.

Van Ostade, I.T.B., **Tottie, G.**, ve **Van Der Wurff, W.**, 1999, *Negation in the History of English*, Walter de Gruyter.

Vaswani, A., **Shazeer, N.**, **Parmar, N.**, **Uszkoreit, J.**, **Jones, L.**, **Gomez, A.N.**, **Kaiser, L.**, ve **Polosukhin, I.**, 2017, Attention Is All You Need.

Vincze, V., **Szarvas, G.**, **Farkas, R.**, **Móra, G.**, ve **Csirik, J.**, 2008, The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes, *BMC Bioinformatics*, 9(S11), doi:10.1186/1471-2105-9-s11-s9.

Vincze, V., **Szarvas, G.**, **Móra, G.**, **Ohta, T.**, ve **Farkas, R.**, 2011, Linguistic scope-based and biological event-based speculation and negation annotations in the BioScope and Genia Event corpora, *Journal of Biomedical Semantics*, 2(Suppl 5), S8, doi:10.1186/2041-1480-2-s5-s8.

Vural, A.G., 2013, *Sentiment-focused web crawling*, Ph.D. thesis, Middle East Technical University.

Wedgwood, D., 2006, Predication, focus and the positions of negation in Hungarian, *Lingua*, 116(3), 351–376, ISSN 0024-3841, doi:https://doi.org/10.1016/j.lingua.2004.08.009.

Wu, X., **Zhao, Y.**, **Yang, Y.**, **Liu, Z.**, ve **Clifton, D.A.**, 2022, A Comparison of Representation Learning Methods for Medical Concepts in MIMIC-IV, *Cold Spring Harbor Laboratory*, doi:10.1101/2022.08.21.22278835.

Yildirim, S., 2024, Fine-tuning Transformer-based Encoder for Turkish Language Understanding Tasks.

Zhang, X., **Zhao, J.**, ve **LeCun, Y.**, 2015, Character-level Convolutional Networks for Text Classification, in C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, ve R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28, Curran Associates, Inc.

İnci Kavak, V., 2019, The acquisition and use of negation in the early child language, *Journal of Language and Linguistic Studies*, pages 587 – 604, ISSN 1305-578X, doi:10.17263/jlls.586745.

THANKS

The names of people to thank do not fit in this section. I wanted to start this section with the wish that most of the negations are used in sentences that make you feel good, as expressed here. I would like to thank my family, relatives, friends, housemates and loved ones whose support I felt in every difficulty. I would like to thank my advisor Prof. Dr. Oğuz Dikenelli, who guided me during the thesis and helped me in my studies, Academic Lecturer Oylum Alatlı, who read my thesis and gave her support and suggestions and to our study partners Prof. Dr. Caner Kerimoğlu and my colleague Esra Mollaoğlu for all of their efforts. Learning a language is not only knowing it but also understanding it. I would like to take this opportunity to thank my primary, high school and undergraduate Turkish teachers for all their efforts in helping me understand Turkish as a language. I would like to thank my colleagues Tolga Kışlal, Mustafa Başak, Patrick Trane, ICCORE Technologies Smart Card Systems Inc. and CardCentric Ltd. with whom I have been working for almost 7 years, for their supports. I would also like to thank Umut Erdal for his support and Neophoria Medical Limited for supporting this study with healthy sportive life suggestions, and Alp Öğüt for sharing ideas about the use of the models and the methodology.

21/03/2024

İmzası

Zana Söğüt

CV

Graduated from Ankara Science High School between 2003-2007. Completed his bachelor's degree in Electrical and Electronics Engineering at Middle East Technical University and master's degree in Computer Engineering at Ege University. Worked as a software developer at Ankira Ltd. in Ankara, working on surgery room control panels. Since 2019, has been working on smart cards and applications at ICCORE Smart Card Technologies and CardCentric in Izmir.