HEAD FINALIZATION AND MORPHOLOGICAL ANALYSIS IN FACTORED PHRASE-BASED STATISTICAL MACHINE TRANSLATION FROM ENGLISH TO TURKISH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

HAYDAR İMREN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

MAY 2015

Approval of the thesis:

HEAD FINALIZATION AND MORPHOLOGICAL ANALYSIS IN FACTORED PHRASE-BASED STATISTICAL MACHINE TRANSLATION FROM ENGLISH TO TURKISH

submitted by **HAYDAR İMREN** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** ______________

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** ______________

Dr. Ruket Çakıcı
Supervisor, **Computer Engineering Department, METU** ______________

Examining Committee Members:

Prof. Dr. Adnan Yazıcı
Computer Engineering Department, METU ______________

Dr. Ruket Çakıcı
Computer Engineering Department, METU ______________

Prof. Dr. İlyas Çiçekli
Computer Engineering Department, Hacettepe University ______________

Assoc. Prof. Dr. Tolga Can
Computer Engineering Department, METU ______________

Dr. Ayşenur Birtürk
Computer Engineering Department, METU ______________

Date: ______________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: HAYDAR İMREN

Signature :

ABSTRACT

HEAD FINALIZATION AND MORPHOLOGICAL ANALYSIS IN FACTORED PHRASE-BASED STATISTICAL MACHINE TRANSLATION FROM ENGLISH TO TURKISH

İmren, Haydar
M.S., Department of Computer Engineering
Supervisor : Dr. Ruket Çakıcı

May 2015, 61 pages

Machine Translation is a field of study which deals with translating text from one natural language to another automatically. Statistical Machine Translation generates the translations using statistical methods and bilingual text corpora. In this study, an approach for translating from English to Turkish is introduced. Turkish is an agglutinative language with a free constituent order, whereas English is not agglutinative and the constituent order is strict. Besides these differences, there is a lack of parallel corpora for this language pair which makes SMT a challenging problem. Up to now, most of the work and research done for this language pair suggest representing the languages at the morpheme-level. The difference of this study is not only representing English and Turkish at morpheme-level but also applying a different reordering technique which was successfully used for other languages, which are grammatically similar to Turkish. The technique is called Head Finalization. To report the results of this study, BLEU metric is used. With improvements in reordering and morpheme-level representation, we have increased our BLEU score from a baseline score of 19.62 to 30.93, which corresponds to an increase of 57%. The experiments can be successfully applied to other languages which are close to Turkish in terms of word order, morphological structure and suffixation.

Keywords: Natural Language Processing, Statistical Machine Translation, Morphology, Reordering, Head Finalization

ÖZ

İNGİLİZCEDEN TÜRKÇEYE FAKTÖRLÜ SÖZCÜK ÖBEĞİ TABANLI İSTATİSTİKSEL MAKİNE ÇEVİRİSİNDE BAŞ SONLANDIRMA VE MORFOLOJİK ÇÖZÜMLEME

İmren, Haydar
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi : Dr. Ruket Çakıcı

Mayıs 2015 , 61 sayfa

Makine Çevirisi, bir metni bir doğal dilden başka bir doğal dile yazılımlar yardımıyla çevirmekle uğraşan bir çalışma alanıdır. İstatistiksel Makine Çevirisi ise bu işi istatistiksel metotlar ve paralel metinleri kullanarak yapar. Bu çalışmada, İngilizceden Türkçeye çeviri için bir yaklaşım tanıtılmıştır. Türkçe sondan eklemeli ve serbest öğe sıralı bir dildir, aksine İngilizce sondan eklemeli olmayan ve katı bir öğe sıralaması olan bir dildir. Bu farklılıkların yanında, iki dil arasındaki paralel metin eksikliği, bu iki dil arasında istatistiksel makine çevirisini zor bir problem haline getirmektedir. Şimdiye kadar, bu iki dil için yapılan çalışma ve araştırmaların çoğu, iki dili de ek-düzeyinde çalışmak gerektiğini önerir. Bu çalışmanın farkı, sadece İngilizce ve Türkçeyi ek-düzeyinde çalışması değil aynı zamanda dilbilgisel açıdan Türkçeye yakın diller için daha önce başarıyla kullanılmış olan farklı bir yeniden sıralama tekniği uygulamasıdır. Bu teknik Baş Sonlandırma tekniğidir. Bu çalışmada sonuçları raporlamak için BLEU ölçüsü kullanılır. Yeniden sıralamada ve ek-düzeyinde yapılan çalışmalarda elde edilen gelişmelerle BLEU skorumuzu 19.62'den 30.93'ye çıkararak %57'lik bir artış sağladık. Bu sonuçlar Türkçe'ye kelime dizilişi bakımından, biçimbilgisel açıdan ve sondan eklenme açısından benzerlik gösteren diğer dillere de başarıyla uygulanabilir.

Anahtar Kelimeler: Doğal Dil İşleme, İstatistiksel Makine Çevirisi, Morfoloji, Yeniden Sıralama, Baş Sonlandırma

To my family...

ACKNOWLEDGMENTS

I would like to thank my supervisor Dr. Ruken Çakıcı for her constant support, guidance and patience. It was a great honor to work with her for the last three years. I also would like to thank Prof. Dr. Fatoş Yarman Vural for motivating me to start my graduate study and influencing me highly in scientific context.

Throughout these three years, I would like to thank my friends Berker Batur, Mehmet Tatlıcıoğlu, Ian Grimaud and Pawel Wos for always motivating and supporting me. I would also like to thank Prof. Dr. Kemal Oflazer for sharing his invaluable experience and also the parallel corpus he used in his studies.

I would also like to thank to my parents for their moral support.

I am most thankful to my wife Hatice Erat İmren, for not losing her hope on me finishing this study, for her patience, endless support and assistance at hardest times.

Lastly, my deepest thanks to my sister Sura İmren for being always there when needed, without her assistance I would not be able to finish.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

MT	Machine Translation
SMT	Statistical Machine Translation
BLEU	Bilingual Evaluation Understudy
SOV	Subject Object Verb
SVO	Subject Verb Object
IWSLT	International Workshop on Spoken Language Translation
SETimes	Southeast European Times
TUBITAK	Scientific and Technological Research Council of Turkey
HPSG	Head-Driven Phrase Structure Grammar
POS	Part of Speech
IRSTLM	IRST Language Modeling
LM	Language Model
MERT	Minimum Error Rate Training
EU	European Union
UN	United Nations
NATO	North Atlantic Treaty Organization

CHAPTER 1

INTRODUCTION

Machine translation, one of the oldest research topics in computational linguistics, is the process of translating from one natural language to another automatically or by computer software. Although, it is one of the oldest research topics, it still attracts attention not only from computer scientists but also linguists, philosophers and mathematicians like many other natural language processing sub-topics.

Until the 1950s, there was not much research done on the subject. In 1949, Warren Weaver first put forward proposals for solving the problems of ambiguity in his *Memorandum on Translation* [75]. In 1951, Yehosha Bar-Hillel started the first research on machine translation at MIT [8]. Since then, because of political, social, commercial and scientific reasons the topic became really popular.

Many approaches have been taken for building machine translation systems or proposing solutions for the problems in machine translation. For the last twenty years, statistical machine translation approach has become widely popular. Before that, rule-based and example-based approaches were dominating the field [53, 64, 73, 74]. Recently, researchers are trying to bring the strengths of those approaches together with hybrid machine translation approaches [28, 65, 21, 56], however statistical approach is still the most popular one.

1.1 Motivation

The need for machine translation have emerged in a variety of fields. International organizations with member states such as EU, NATO, UN etc., translate their documents into many different languages. Likewise, companies distributing products in many countries have to prepare documents and user manuals for their products in numerous languages. However, these tasks are mostly carried on by human translators as machine translation systems are still not as accurate as human translation. Because of this reason, researchers are motivated to work on better quality and efficient machine translation systems.

For the last two decades, statistical machine translation has been the most popular paradigm in the field of machine translation and there has been an extensive amount of work in the field of statistical machine translation [13, 40, 15]. For English - Turkish language pair, on the other hand, the research and work done is not still at desired levels which is caused by many challenges in building English - Turkish translation systems. Word order difference, morphological differences, lack of parallel data can be given as examples to those challenges. In this thesis, our aim is to propose easy solutions to some of these challenges and develop an efficient statistical machine translation system from English to Turkish. There has been some recent work on languages like Finnish and Hungarian that are agglutinative and morphologically rich like Turkish [79, 72, 51]. These studies together with the other language pairs inspire building systems for English - Turkish statistical machine translation.

Our motivation is to build an English to Turkish translation system that are comparable to the results of other language pairs. Although there are pioneering studies such as [17, 18, 58, 10, 20, 80] in the literature, these studies could not achieve as high scores as the other language pairs like English - Japanese [26] or English - Korean [43] which are considered to have the same chatracterics as Turkish. Commercial products such as Google Translate and Bing Translator are still not efficient for pairs involving Turkish language. This problem motivates us to propose solutions to some challenges encountered in English - Turkish machine translation and provide a basis for future research. The techniques proposed in this study can also be used for other agglutinative and head-final languages. This study can also be used for translating to other Turkic languages where it is even more challenging to find parallel corpora.

1.2 Contributions of the Thesis

One of the most challenging problems of English - Turkish statistical machine translation is the word order problem. English has a rather fixed word order which is Subject-Verb-Object, while Turkish has a very flexible word order while being dominantly Subject-Object-Verb. Another challenge is the rich morphology of Turkish, which simply means a phrase (series of words) in English can be translated into just one word in Turkish or function words in English can simply be suffixes of a word in Turkish. In this study, we aim to solve these problems and build a state-of-the-art statistical machine translation system from English to Turkish.

The list of contributions of this thesis are listed below.

1. *Head Finalization* [33] is suggested to overcome the problems associated with the word order differences. This technique is simpler and more intuitive than the techniques proposed so far [78, 76, 44, 77, 30, 80] and applied is successfully for Japanese which is also a head-final language like Turkish.
    (a) For the English side of the parallel data, *Head Finalization* is applied.
    (b) Hand-written rules are added to the process to prevent reordering of sentence parts starting with conjunctions or conjunctive punctuation.
2. The parallel data are morphologically analyzed and disambiguated for Turkish and part-of-speech tagged for English in order to handle the problem of translating from a language with limited morphology to a rich morphology language.
3. The results of this study show that we surpass the results obtained by Yeniterzi [80] and Tatlıcıoğlu [68] in the benchmark studies.

Experiments are carried out separately for each approach. We present the results for reordering, integrating morphology and both.

1.3 Outline

The outline of the thesis is as follows:

Chapter 2 starts with background information about machine translation and continues with explaining the basics of statistical machine translation. Various approaches to statistical machine translation are explained.

Chapter 3 explains the challenges of English - Turkish machine translation in detail and covers a comparative literature review of the research done and the techniques used until now.

Chapter 4 presents detailed information about the data and tools used, explains preprocessing steps applied to the data and *Head Finalization* which is used for reordering English data in this thesis.

Chapter 5 presents all the experiments carried out throughout this study, their results and error analysis. Six different translation systems are built for the experiments. These systems are:

1. Baseline System
2. Baseline-Reordered System
3. Tagged-Baseline System
4. Tagged-Reordered System
5. Factored-Baseline System
6. Factored-Reordered System

Lastly, Chapter 6 summarizes the work done in this thesis and puts forward how this work can be further improved to guide the future researchers.

CHAPTER 2

BACKGROUND INFORMATION

2.1 Introduction to Machine Translation

Machine translation is the process of translating a *source text* from a natural language into a , another natural language, *the target language*, by computers.

Briefly, machine translation can be understood as a substitution of source language words with target language words. However, that approach does not produce good quality translation systems as languages may differ in word order, and the translated text may not achieve the desired meaning. Because of this, many other approaches have been suggested to build efficient, better quality machine translation systems.

This problem was first introduced by Warren Weaver in 1949 in his *Memorandum on Translation* [75]. Weaver suggested that translation can be seen as a cryptography problem, a text can be converted into another language just like cyphers. He said, "I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is strip off the code in order to retrieve the information contained in the text." [35] page 133, which was an inspiration for the latter research. However, future research proved that *the code*, Weaver mentioned, was more complicated than it was thought to be. Instead of being just a substitution of symbols, the solution to the problem needed more analysis on the grammars and semantics of languages to preserve the meaning of the text to be translated. After Weaver's *Memorandum on Translation*, in the 1950s, IBM researchers were able to successfully translate 60 Russian sentences into English fully automatically using 250 words and 6 rules [32].

From that time until the late 1980s, the researchers focused on building translation systems using hand-written linguistic rules which were created by expert linguists. That approach is called rule-based machine translation.

After late 1980s, a new approach arose for the machine translation problem. Researchers started to develop translation systems using parallel texts of language pairs and use statistics. Since then, research on statistical machine translation has rapidly grown [61].

2.2 Approaches to Machine Translation

2.2.1 Rule-based Approach

Rule-based machine translation systems use linguistic rules and dictionaries to translate a piece of text from one language to another. This approach basically links the semantic, morphological and syntactic structure of the source language with the target language by using human defined linguistic rules. This linking is done using a parser and an analyzer for the source language and a generator for the target language and a transfer dictionary for the translation [31, 6]. A rule-based translation system needs:

1. A dictionary that will map each source language word to a target language word.
2. Rules representing the semantic, morphological and syntactic structure of a regular source language sentence.
3. Rules that define the relationships between the structures of source and target languages.

For rule-based translation systems, use of parallel data is not required. Most of the rules are written in a domain independent way, which makes rule-based systems also domain independent. Any exceptional rule can be added to correct unexpected, erroneous translations and as rules are hand-written, these systems are easy to debug.

However, it can be difficult to deal with introducing new rules into large systems and it can also be very costly to build good quality dictionaries.

There are three types of this approach:

1. Dictionary-based Approach

   In this approach, translation is done word-by-word just like a dictionary does. It is mostly used to translate inventories, databases or phrases, but not full sentences [52]. Morphological analysis or lemmatization can also be used.

2. Interlingual Approach

   In this approach, the source language text is first transformed into an interlingua (an auxiliary language) and then the target language text is generated [60, 16]. The source language text is actually transformed into an abstract language which defines its syntactic, morphological and semantic characteristics. Sometimes two interlinguas are used, one covering the characteristics of the source language and one covering the characteristics of the target language. This principle also provides a basis for pivot machine translation systems. The steps of interlingual approach is shown in 2.1, which is also known as Vauquois' pyramid [71].

   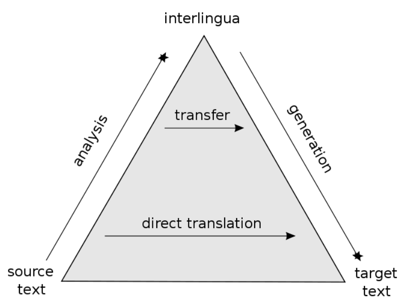

   Figure 2.1: Bernard Vauquois' pyramid

3. Transfer-based Approach

   This approach is similar to the interlingual approach. Generally, the input text is analyzed for morphology and syntax, and an internal representation is created. Using dictionaries and grammatical rules, the translation is generated. The transfer process takes part before generating the translation. Then, the internal representation, which is in the source language, is converted into a representation with the same level of abstraction in the target language. Then, reverse applications of morphological and syntactic analysis are applied and the translation is generated. Some examples of transfer-based approach is given in [27, 42].

2.2.2 Example-based Approach

Example-based translation systems makes use of parallel data. This approach was first introduced by Makoto Nagao in 1984 [53]. When the input sentence is given to the system, similar sentences are selected from the parallel data. The similarity depends on the likeness of sub-sentential components. These similar sentences are used to translate the sub-sentential components and translation is generated. The difference of this approach from rule-based approaches is that it does not need any linguistic rules or analysis, and because of this, the source and target languages can be grammatically different languages.

An example for the example-based approach is shown below.

English: **How much is that** *red umbrella* ?

Turkish: **Şu** *kırmızı şemsiye* **ne kadar** ?

English: **How much is that** *small camera* ?

Turkish: **Şu** *küçük fotoğraf makinesi* **ne kadar** ?

From the example, it is seen that the translation of the phrase "*How much is that ...*" is "*Şu ... ne kadar*", "*red umbrella*" is "*kırmızı şemsiye*" and "*small camera*" is "*küçük fotoğraf makinesi*" in Turkish.

2.2.3 Statistical Approach

Statistical machine translation systems translate input sentences using statistical models whose parameters are derived from the analysis of a parallel corpus which is a body of documents in multiple languages with aligned sentences. Basically the idea is to find the most probable translation of a given sentence. Probabilities of translations are determined automatically by estimating parameters of a statistical translation model from the parallel corpus. This approach can be applied to any language pair that has enough parallel data, and requires the least amount of human effort among all approaches. It was first introduced by Warren Weaver in 1949 [75] and then re-introduced by IBM researchers in 1988 [11, 12, 13].

The mathematics behind statistical machine translation suggests that given a target language model *P(e)*, a sentence in the source language *f* and a translation model *P(f | e)* the goal is to find the translation *e* which maximizes *P(e) x P(f | e)* following Bayes' Theorem.

In order to find the probability *P(e | f)*, we apply the Bayes' Theorem which is shown in Equation 2.1. Then, the translation that gives the highest probability is picked up as the best translation as shown in Equation 2.2.

$$p(e|f) = \frac{p(f|e)p(e)}{p(f)} \tag{2.1}$$

Bayes' Theorem for Statistical Machine Translation

$$\tilde{e} = \arg\max_{e \in e^*} p(e|f) = \arg\max_{e \in e^*} p(f|e)p(e) \tag{2.2}$$

Finding the best translation in Statistical Machine Translation

From the formula, we can see that statistical machine translation problem actually has three parts: (1) building a target language model to estimate *p(e)*; (2) building a translation model to estimate *p(f | e)*; (3) searching for a translation *e* to maximize the product *p(f | e) p(e)*, which is also called decoding [12].

A statistical language model is a probability distribution *P(s)* over all sentences *s* or any other linguistic unit, which was introduced in 1980 [34, 61]. In statistical machine translation, language models are used for assigning a probability to the occurrence of a sequence of *m* linguistic units (mostly words or phrases), by means of the probability distribution of all units. The model, where the probability of a unit depends on the previous *n* units, is called an N-gram language model. In an N-gram language model, the probability of a sentence s with words $w_1 w_2 ... w_m$ is shown in Equation 2.3.

$$P(w_1, ..., w_m) = \prod_{i=1}^{m} P(w_i | w_1, ..., w_{i-1}) \approx \prod_{i=1}^{m} P(w_i | w_{i-(n-1)}, ..., w_{i-1}) \qquad (2.3)$$

The probability to observe a sentence in an n-gram language model

An N-gram language model uses only *n-1* words of prior context to estimate the probability of a word. This comes from the Markov Assumption, which is the presumption that the future state of a dynamical system only depends on its recent history [36]. In particular, a k^{th}*-order Markov Model* suggests that the next state of a system only depends on the k most recent states, therefore an N-gram language model is a *n-1*-order Markov model. That means, the probability of observing a word w_i in a sentence where previous *i-1* words are known, can be approximated to the probability of observing it in the context of previous *n-1* words with an n-gram language model. However, when an unseen word is confronted, this model will fail and assign a probability of 0 to the new word. To eliminate this problem, smoothing methods are usually applied, such as Kneser-Ney smoothing [37] which is also used in this study.

The translation model is created using the parallel corpus. The first step in creating the translation model is the word alignment. After the words are aligned, two notions are derived from alignments; fertility and distortion [12]. Fertility is the number of target language words generated for a source language word. Distortion is the position difference between the target language word and the source language word in the sentence.

Searching is done after finding all possible translations of a given sentence. Using language and translation models, probabilities for partial alignments are computed. The idea is to stack the promising partial alignments, which have higher probabilities and extend the stack until a complete translation is achieved.

Phrase-based approach is the most common statistical method in use today. Phrase-based statistical machine translation was first introduced by Daniel Marcu and William Wong (2002) [46] as a joint probability method for learning words and phrases from the bilingual corpora. The approach was then revised by Koehn et al. (2003) [40]. Axelrod (2006), introduced factored translation models for statistical machine translation [7], which was revised by Koehn and Hoang (2007) [38]. Factored translation approach is an extension of the phrase-based approach and it allows using additional annotation at the word level. In the classic phrase-based approach, a word is represented as a single token, whereas in the factored phrase-based approach, a word is represented as a vector of factors each of which serves as different levels of annotation. Figure 2.2 is taken from [38] to give an illustration of factored representation of words.

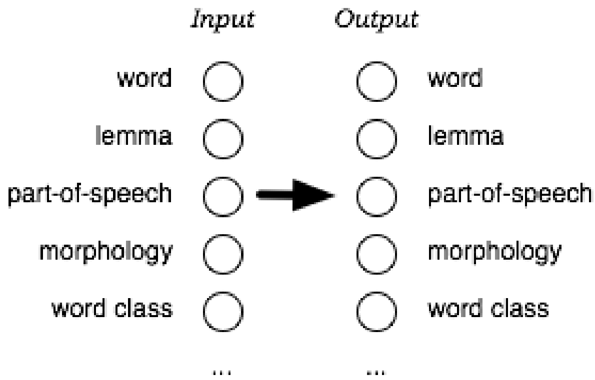

Figure 2.2: Factored representations of input and output words incorporate additional annotation into the statistical translation model

There are other statistical approaches such as word-based [66], syntax-based [78] and hierarchical phrase-based approaches [15]. In this study, for an English - Turkish translation system, we adopt the factored phrase-based approach.

2.2.4 Hybrid Approach

Hybrid approaches use the strengths of both statistical and rule-based approaches. Some hybrid MT systems are given in [65, 21, 56]. Hybrid approaches are mainly divided into two categories in the literature.

1. Translation process is rule-based and statistics are used to post-process the translation. This is also known as statistical smoothing and automatic post-editing. Examples of this approach are introduced by Sanchez-Matinez et al. (2009) [63] and Simard et al. (2007) [67].

2. Rules are used to pre-process the data, statistical approaches are used for trans-

lation and rules are used to post-process the translation. As SMT systems are mostly language independent and need less human effort, this type of hybrid approach is commonly used. Xia and McCord (2004) introduce an example of using pre-processing rules for the SMT system [76], whereas Formiga et al. (2012) show an example of using post-processing rules [23].

CHAPTER 3

STATISTICAL MACHINE TRANSLATION FROM ENGLISH TO TURKISH

English, being currently the dominating language for communicating internationally, it is essential to prioritise the lanaguge pairs involving English as the source or the target language in translation tasks. With this view, research on machine translation on the Turkish-English (or English-Turkish) language pair has increased in recent years. However, statistical machine translation from English to Turkish is still challenging in many aspects. Because of the linguistic differences between these languages, building a robust machine translation system for this language pair is harder than doing it for linguistically closer languages. The performances of baseline SMT systems provided for 11^{th} International Workshop on Spoken Language Translation (IWSLT 2014) [2] MT track show that systems built upon linguistically close language pairs are more successful than the other systems in terms of BLEU scores [1], which can be seen in Table 3.1.

3.1 Challenges

3.1.1 Word Order Problem

In English, the constituent order is more strict than Turkish. English has a Subject-Verb-Object order, while Turkish is much more flexible. Despite being flexible, the most common constituent order of Turkish sentences is Subject-Object-Verb. Changing the order of the words in Turkish is acceptable and even used for information

Table 3.1: IWSLT 2014 MT Track - BLEU scores of baseline smt systems

Language Pair	BLEU Scores (%)
English - Arabic	11.60
English - German	20.42
English - Spanish	31.64
English - Persian	10.63
English - French	29.44
English - Hebrew	18.06
English - Italian	23.17
English - Dutch	25.33
English - Polish	9.33
English - Portuguese	30.29
English - Romanian	20.23
English - Russian	13.10
English - Slovenian	11.01
English - Turkish	**7.59**
English - Chinese	16.93

structure [29]. However, in English, there are a few situations where the order is changed. For example, temporal adverbs can be used at the beginning or at the end of a sentence depending on the emphasis. Word order difference is not only at sentence level but can also be seen in sub-sentential constructions like phrases and clauses.

Examples of different ordered Turkish sentences and sub-sentential constructions with their English translations and reordered English counter-parts are shown in Figure 3.1.

The first four examples in Figure 3.1 indicate the word order difference at sentence level. The fifth and sixth examples show the word order difference at sub-sentential level. From the fifth example, we see that the function words such as `in` or pronouns such as `my` in English are represented as bound suffixes in Turkish. We also see that the English phrase contains eight words, whereas the Turkish phrase contains three words. In the sixth example, we see that conjunctions such as `after` in English become postpositions in Turkish.

(1)	Turkish	Hakan ödevi yaptı.
	English	Hakan did the homework.
	Reordered English	Hakan (the) homework did.
(2)	Turkish	Hakan yaptı ödevi.
	English	Hakan did the homework.
	Reordered English	Hakan did (the) homework.
(3)	Turkish	Ödevi Hakan yaptı.
	English	Hakan did the homework.
	Reordered English	(The) homework Hakan did.
(4)	Turkish	Öğretmen, "Hakan ödevi yaptı." dedi.
	English	The teacher said "Hakan did the homework".
	Reordered English	(The) teacher "Hakan the homework did" said.
(5)	Turkish	Ankara'da okuyan ağabeyim
	English	My elder brother who is studying in Ankara
	Reordered English	Ankara-in (is) studying (who) elder brother-my.
(6)	Turkish	Sunum başladıktan sonra
	English	After the presentation started
	Reordered English	(The) presentation started after

Figure 3.1: An example of word order difference

3.1.2 Turkish Morphology

Turkish, as a member of the Ural-Altaic language family, has a rich inflectional and derivational morphology. This means that a single word in Turkish can consist of a lemma and many morphemes each of which represents a different meaning. Besides, the same morpheme can change form in different words depending on vowel harmony, consonant assimilation or other phonological processes. Thus, a Turkish word can be aligned with a bunch of words in English. An example of this is shown in Figure 3.2.

The example in Figure 3.2 shows that, in order to build a machine translation system, many to one alignment from English words to Turkish word(s) may be required. Morphological analysis and disambiguation is performed on Turkish data in order to aid with this challenge. Figure 3.3 shows a possible alignment between an English

Turkish Word	Turkish Morphological Representation	English Translation
aldıklarımız	al + DHk + lArH + mHz	the things we got
çektiklerimiz	çek + DHk + lArH + mHz	the things we pulled
aldığınız	al + DHk + (s)H + nHz	the thing you got
bitirebilenlerimizinkiler	bitir + (y)Abil + An + lAr + HmHz + Hn + ki + lAr	the things of those of us who can finish
yapabilecekseniz	yap + (y)Abil + AcAk + sA + nHz	if you are going to be able to do

Figure 3.2: An example of Turkish morphology and English translations

phrase and the morphological representation of a Turkish word.

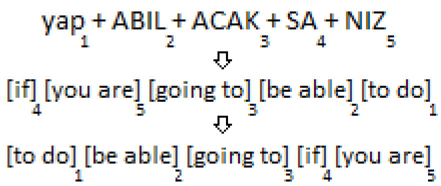

Figure 3.3: Alignment with morphology

3.1.3 Available Parallel Corpora

Compared to other European language pairs, there are not many parallel corpora for English - Turkish. Alperen and Tyers (2010) give a multi-lingual parallel corpus from SE Times news website [70]. SETimes corpus consists of about 200,000 parallel sentences in ten languages including Turkish and English. Tiedemann (2012) also provides a parallel corpus for English and Turkish compiled from OpenSubtitles website [69]. The International Workshop on Spoken Language Translation (IWSLT) provides multi-lingual parallel corpora based on TED Talks [45] every year. The parallel corpus provided for English - Turkish in 2014 contains about 160,000 sentences. These are the largest parallel corpora that are publicly available for this language pair.

In this thesis, a parallel corpus of 54,391 sentences (Oflazer's corpus), obtained from Prof. Kemal Oflazer is used. This relatively small corpus was chosen as the data because it is relatively clean and for the purpose of comparability with the other studies using this corpus. SETimes, OpenSubtitles and TED Talks corpora contain many misalignments and typos and require a significant cleanup process. We also clean up SETimes corpus which reduced to 168,331 parallel sentences and performed an additional experiment with that corpus. Researchers who would like to use this version of the SETimes corpus in their studies can contact us to obtain it.

3.2 Related Work

In recent years, work on Turkish - English statistical machine translation has increased a lot. Oflazer and El-Kahlout (2005) present a preliminary work on aligning Turkish and English parallel texts for statistical machine translation [17]. In 2006, they present the problems in developing a statistical machine translation system from English to Turkish and explore various ways of exploiting morphology and improve their baseline BLEU score from 7.52 to 9.13 [18]. In 2007, again Oflazer and El-Kahlout explore different representational units in English to Turkish statistical machine translation and obtain 25.08 BLEU score [58] where they apply selective morpheme grouping on Turkish data. Bisazza and Federico (2009) perform morphological pre-processing for Turkish to English statistical machine translation for IWSLT 2009 [10], gaining 5 point BLEU improvement. Oflazer and El-Kahlout (2010) investigate the effect of different representations of morphology on both English and Turkish and combine it with local word reordering on English side, which is extensively discussed in [20]. Oflazer and Yeniterzi (2010) introduce a new way to align English syntax with the Turkish morphology by associating function words to their related content words [80] and obtain 23.78 BLEU score. Gorgun and Yildiz (2012) introduce that using different sub-lexical representations instead of word forms. This improves the performance of statistical machine translation from English to Turkish by 21% [25]. The Scientific and Technological Research Council of Turkey (TUBITAK) produced a number of Turkish - English statistical machine translation systems for IWSLT machine translation track [81, 49, 48, 47]. El-Kahlout et al. (2012) present recent improvements in Turkish - English statistical machine translation [19]. Çakmak et al. (2012) compare the effect of morphological segmentation on word alignment for Turkish - English language pair [5]. Mermer et al. (2013) propose a Bayesian approach to word alignment inference and use Gibbs sampling to sample the posterior alignment distributions [50]. Tatlıcıoğlu (2013) increases the BLEU score of the benchmark Turkish to English phrase-based statistical machine translation system from 25.22 to 26.22 by fusing rule-based and stochastic word decomposition methods in his thesis study [68].

In most of these studies, we see that morphological knowledge is exploited in order to build better quality SMT systems and smaller units of representation is better suited for the agglutinative nature of Turkish language. There are also some improvements in BLEU score by reordering English data for better word alignments.

CHAPTER 4

BUILDING FACTORED PHRASE-BASED STATISTICAL MACHINE TRANSLATION SYSTEM

4.1 Data Preparation

In this study, we work on an English - Turkish parallel corpus of 54,391 sentences which is a collection of European Union documents, European Court of Human Rights documents and several treaty texts [80]. The English corpus contains 1,238,169 words, an average of 23 words per sentence and the Turkish corpus consists of 973,442 words, an average of 18 words per sentence. An example from the parallel data is shown in Figure 4.1.

1	26 Mart 2003 tarihinde Komisyon tarafından Üye Ülkelere sunulan gözden geçirilmiş Katılım Ortaklığı belgesi , Üye Ülkelerce bugün Lüksemburg'da kabul edilmiştir .	1	the revised Accession Partnership prepared by the Commission on 26 March 2003 , was adopted by the Member States on 15 April 2003 in Luxembourg .
2	bu belge , Türkiye - AB ilişkileri bakımından , 2002 Aralık Kopenhag AB Devlet ve Hükümet Başkanları Zirvesi kararlarında öngörülen önemli unsurlardan biridir .	2	in terms of Turkey - EU relations , this document is one of the important elements foreseen by the Conclusions of the Copenhagen European Council held in December 2002 .
3	gözden geçirilmiş Katılım Ortaklığı belgesi , 2002 Kopenhag Zirvesi ile tesis edilmiş mekanizmaların öngörüldüğü şekilde işlemekte olduğunun da bir teyididir .	3	the revised Accession Partnership also confirms that the mechanisms instituted with the 2002 Copenhagen European Council are functioning as envisaged .
4	anılan belgede , beklentilerimizin ötesinde unsurlar bulunmamaktadır .	4	the document does not contain elements that exceed our expectations .
5	Türkiye , önümüzdeki haftalarda , yeni Katılım Ortaklığı belgesi çerçevesinde , Ulusal Programının gözden geçirilmiş metnini sonuçlandıracak ve AB Komisyonu'na iletecektir .	5	Turkey will finalize its revised National Program for the Adoption of the Acquis under preparation , in response to the new Accession Partnership and forward it to the EU Commission in the coming weeks .
6	bu yöndeki hazırlık ve çalışmalar AB Genel Sekreterliği'nin eşgüdümünde sürdürülmektedir .	6	preparations to this effect are underway under the coordination of the EU Secretariat General .
7	bununla birlikte , Helsinki Zirve Sonuçları uyarınca AB'nin Türkiye'ye ilişkin tutumunun diğer adaylara yönelik uygulamayla eşit olması gerekmektedir .	7	on the other hand , the EU's approach towards Turkey should be equal in practice to that displayed towards the other candidate countries in accordance with the Helsinki European Council Conclusions .
8	buna karşılık , Kıbrıs konusu ile sınır uyuşmazlıkları , gözden geçirilmiş Katılım Ortaklığı belgesinde de siyasi diyalog ve siyasi kriterler başlığı altında yer almaktadır .	8	however , both the issue of Cyprus and the border disputes , have also been included in the revised Accession Partnership under the political dialogue and political criteria headings .
9	1999 Helsinki Zirve Sonuçları çerçevesinde , her iki konunun da Kopenhag siyasi kriterleriyle bir ilgisi bulunmamaktadır .	9	in the context of the 1999 Helsinki European Council Conclusions neither of these issues are in any way related to the Copenhagen political criteria .

Figure 4.1: An illustrative segment from the parallel corpus

Both Turkish and English sentences are already tokenized and case-normalized. For the experiments in this study, we create 10 datasets from this parallel data. Each dataset consists of 52,391 sentences of training, 1000 sentences of tuning and 1000 sentences of testing data. The sentences are selected randomly from the complete corpus for each of the datasets, provided each one differs from the others.

As mentioned in Chapter 1, six different statistical machine translation systems are created throughout this study. These systems are divided into three categories: *baseline*, *tagged* and *factored* systems. Each category consists of two systems, one built with reordered English data, the other with original English data. Table 4.1 gives a summary of the systems built.

An additional experiment is also carried out with the SETimes courpus [70] to see how systems perform with a larger data set.

Table 4.1: Translation systems built in this study

		Systems	
Categories	**Baseline**	Baseline	Baseline-Reordered
	POS Tagged for English / Morphologically Tagged for Turkish	Tagged-Baseline	Tagged-Reordered
	Factored	Factored-Baseline	Factored-Reordered

Pre-processing of the data is not needed to build the baseline system. However, in order to build the other five systems some pre-processing steps are applied on the data. These steps are as follows:

1. Reordering English data using Head Finalization
   (Used in Baseline-Reordered, Tagged-Reordered and Factored-Reordered systems)
2. Morphologically tagging Turkish data
   (Used in Tagged-Baseline, Tagged-Reordered, Factored-Baseline and Factored-Reordered systems)
3. Part-of-speech tagging English data
   (Used in Tagged-Baseline, Tagged-Reordered, Factored-Baseline and Factored-Reordered systems)
4. Factorizing both sides of the data
   (Used in Factored-Baseline and Factored-Reordered systems)

4.2 Reordering English Data

As discussed in section 3.1.1, different word orders of target and source languages are challenging for English - Turkish machine translation. The constituent order in English is fixed and is *Subject-Verb-Object* (SVO), on the other hand, Turkish constituent order is relatively flexible. However, the canonical sentence structure in Turkish suggests that the *Verb* is usually at the end of a sentence and it is very common that the *Subject* is at the beginning in written text. We can say that Turkish is mostly *Subject-Object-Verb* (SOV). Reordering one language as the other may be helpful to get better word alignments which can result in better translations [78, 76]. In this study, we choose to reorder English side of the parallel data to make it similar to Turkish in terms of word order.

There are many studies that reorder SVO sentences into SOV with a language independent approach. Some are Yamada and Knight (2001) [78], Xia and McCord (2004) [76] and Li et al. (2007) [44]. The main idea of these approaches is to parse the input sentence and learn reordering decisions for each node of the parse tree. Xu et al. (2009) convert SVO sentences into SOV by using a dependency parser [77]. They create a hand-written rule set to move the words in the sentence according to the dependency tags. They show that they increase BLEU scores for all languages including Turkish. Hong et al. (2009) do the same for only English - Korean language pair [30]. They also apply some rules to introduce pseudo-words into the source text and then reorder according to another rule set. For Turkish, a similar approach is taken by Oflazer and Yeniterzi (2010) where they first use syntactic transformations and then apply reordering based on a rule set that covers the most common reordering patterns [80].

4.2.1 Head Finalization

Isozaki et al. (2010) introduce *Head Finalization*, a simple technique to convert SVO sentences into SOV sentences [33]. The technique has one rule, move the syntactic head to the end of corresponding syntactic constituent. A `head` is the word that determines the syntactic type of phrase. For example, head of a noun phrase is a noun and head of a verb phrase is a verb. The other elements of the phrase are dependents or modifiers of the head. There is one exceptional case where the head is not moved to the end: coordination expressions. An expression like `A and B` should not be reordered as `B and A`.

The steps of *Head Finalization* are as follows. First, the sentence which is in SOV order, is parsed by a parser which can output syntactic heads of every sub-tree in the whole parse tree. Then recursively most of the heads are moved to the end of their own sub-trees.

We use Enju Parser [41] to generate parse trees of English sentences. Enju is an HPSG (head-driven phrase structure grammar) parser which also outputs the head node of each sub-tree, POS tag of a node, dependencies of a word, base form of a word, etc. An example of an Enju output is shown in Figure 4.2.

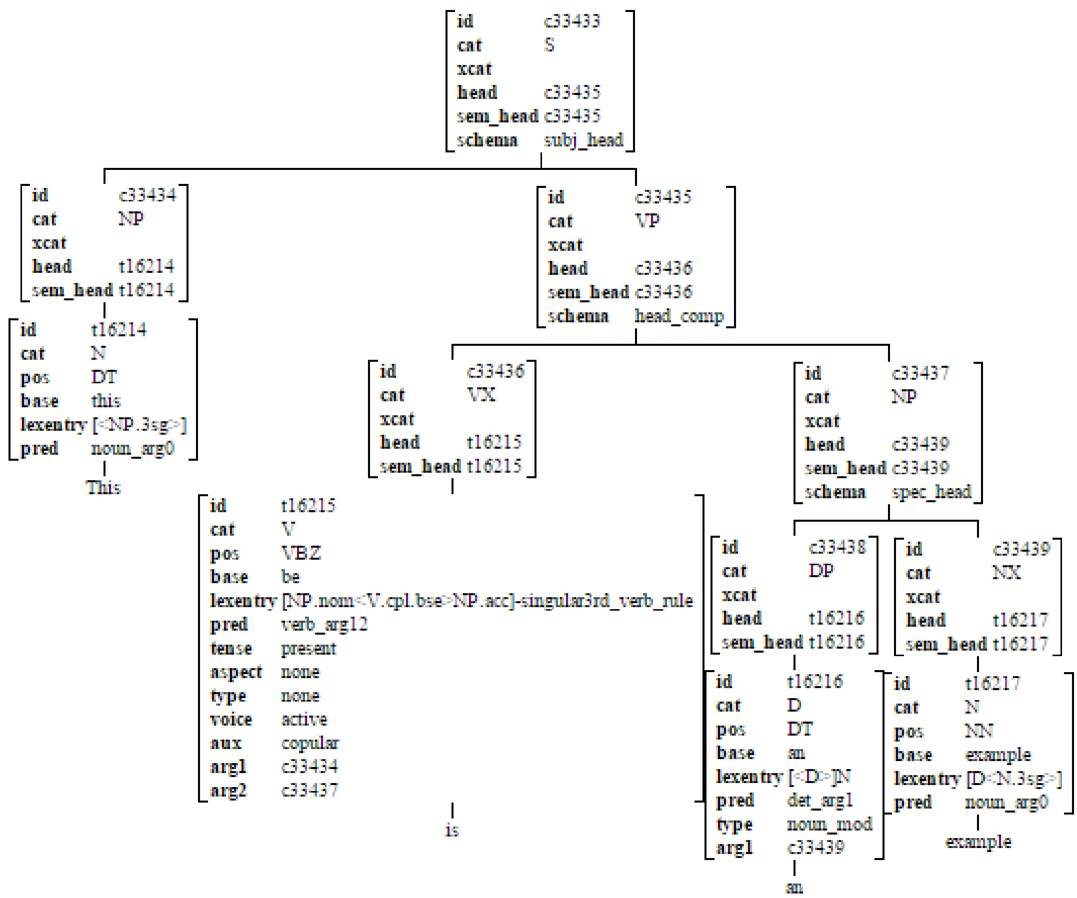

Figure 4.2: An example of Enju Parser output

We use XML style Enju output and convert it into reordered readable text applying head finalization. An example of head finalization can be seen in Figure 4.3. The starred nodes in the parse tree on the left are the head nodes of their own sub-trees. In the parse tree on the right, it is seen that all starred nodes are moved to the end of their own sub-trees. The word order of both parse trees compared to the Turkish translation is shown in Figure 4.4. As seen from the example, the new word order is closer to Turkish word order.

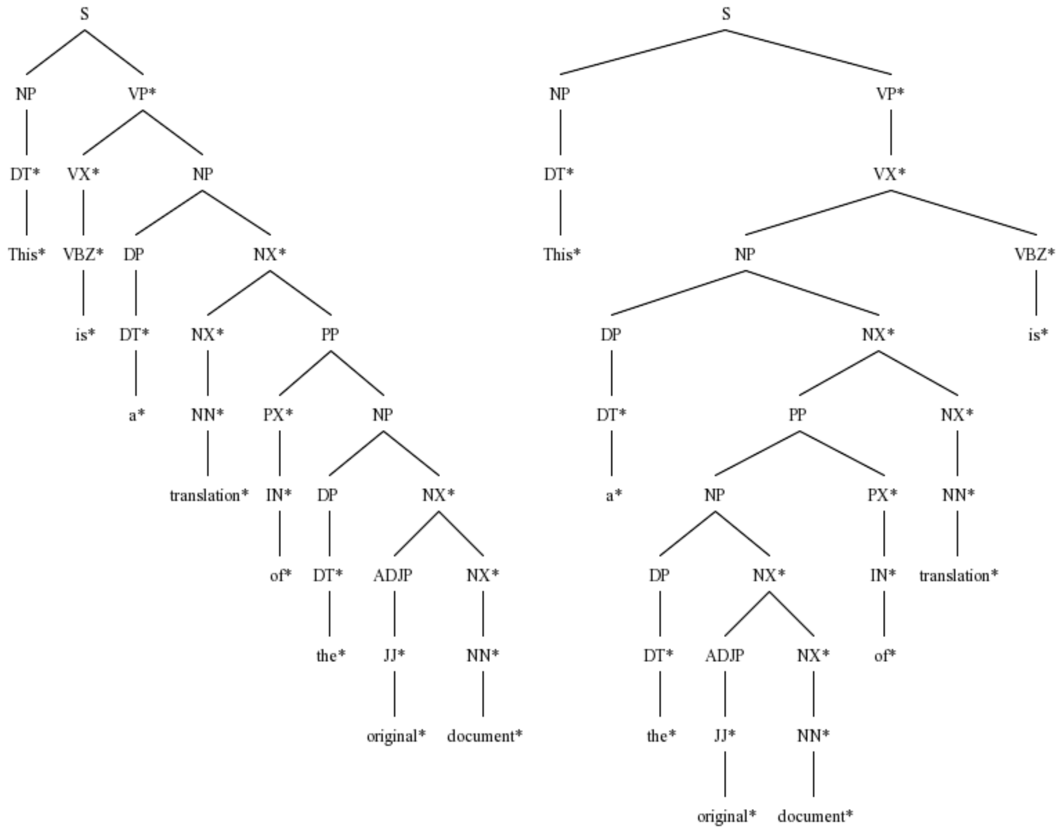

Figure 4.3: An example of head finalization

Determiners pose a problem with this approach. The problem is that the reordered sentence has all the determiners at the beginning. A more clear example can be the noun-phrase " `the determiner of a word in a sentence`". Normally, this sentence should be head-finalized as " `the a a sentence in word of determiner`". In order to solve this issue, while reordering, determiners are moved to the front of the syntactic head of their dependents. So the example in Figure 4.4 will look like Figure 4.5 after handling the problem about determiners.

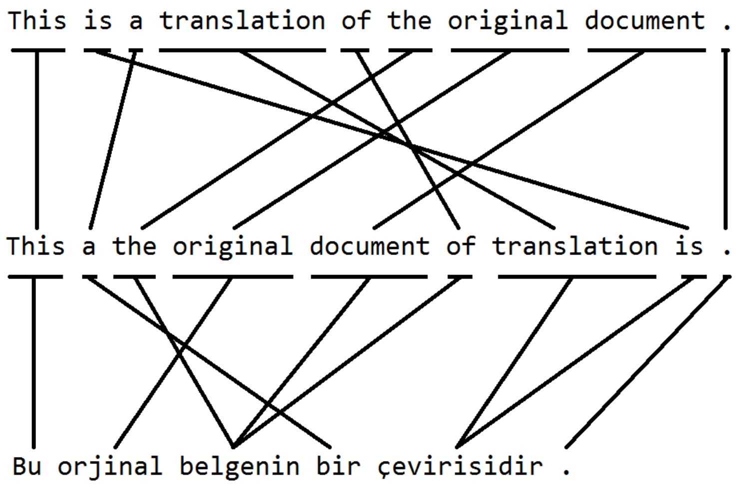

Figure 4.4: Word order change after head finalization

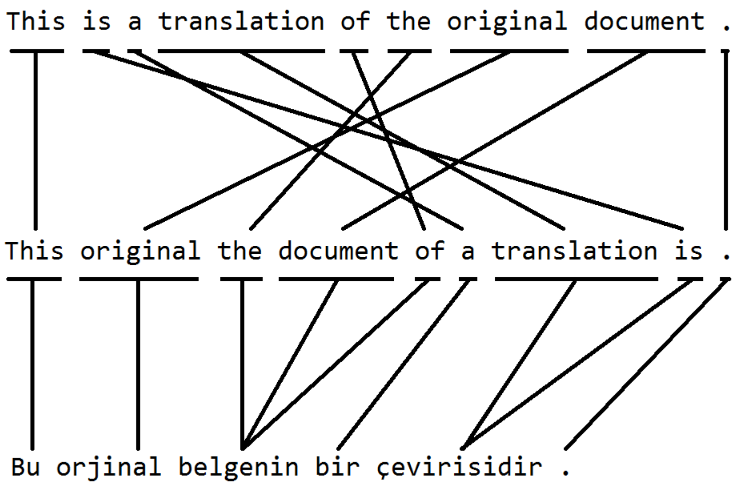

Figure 4.5: Word order after handling determiner problem

4.3 Morphological Tagging of Turkish Data

In Section 3.1.2, we explain how the rich Turkish morphology affects translation to languages with less complex morphology. The examples shown in Figure 3.2 suggest that if we translate directly from English to Turkish without integrating any morphological knowledge, the quality will be lower than expected. Oflazer and Kahlout (2006) show a good example of this problem in their study [18], despite many forms of the root word `faaliyet ('activity')` appear in the parallel corpus, the inflected form `faaliyetlerimizde ('in our activities')` does not exist which means it is impossible to generate a translation for that form. However, with morphological knowledge the system can generate translations for the suffixes `ler`, `imiz` and `de`, and it would be enough for the system to know the root word `faaliyet`. It can generate any inflected form and derivation of the root words with its knowledge of morphemes and their translations which will reduce possible out-of-vocabulary errors. Because of these reasons, following the other studies, we also utilize Turkish morphology. On the other hand, to cover similar grammatical information on the English side, we make use of part-of-speech tags of English words.

We use Oflazer (1993)'s two-level morphological analyzer to find all possible morpheme level representations of words on the Turkish side of parallel corpus [57]. An example output of the analyzer is shown in Figure 4.6. Here, if a word has more than one possible analyses, they are shown in separate lines.

```
ekonomik kriterler ve müktesebat uyumu açısından ise faaliyetlerimiz aynı şekilde sürdürülecektir .

ekonomik    ekonomik    +Adj

kriterler kriter    +Noun+A3pl+Pnon+Nom
kriterler kriter    +Noun+A3sg+Pnon+Nom^DB+Verb+Zero+Pres+A3pl

ve   ve   +Conj

müktesebat   müktesebat+Noun+A3sg+Pnon+Nom

uyumu   uyum    +Noun+A3sg+P3sg+Nom
uyumu   uyum    +Noun+A3sg+Pnon+Acc

açısından    açı+Noun+A3sg+P3sg+Abl

ise  i    +Verb+Pos+Cond+A3sg
ise  is   +Noun+A3sg+Pnon+Dat

faaliyetlerimiz    faaliyet  +Noun+A3pl+P1sg+Nom^DB+Verb+Zero+Pres+A1pl
faaliyetlerimiz    faaliyet  +Noun+A3pl+P1pl+Nom

aynı     aynı+Adj

şekilde  şekil+Noun+A3sg+Pnon+Loc

sürdürülecektir    sür+Verb^DB+Verb+Caus^DB+Verb+Pass+Pos+Fut+A3sg+Cop
```

Figure 4.6: The output of morphological analyzer

After analyzing the data, Sak (2007)'s averaged perceptron-based morphological disambiguator is used to disambiguate the output of morphological analyzer [62]. Disambiguated morphological analysis of the same sentence in Figure 4.6 is shown in

Figure 4.7.

```
ekonomik ekonomik+Adj
kriterler kriter+Noun+A3pl+Pnon+Nom
ve ve+Conj
müktesebat müktesebat+Noun+A3sg+Pnon+Nom
uyumu uyum+Noun+A3sg+P3sg+Nom
açısından açı+Noun+A3sg+P3sg+Abl
ise i+Verb+Pos+Cond+A3sg
faaliyetlerimiz faaliyet+Noun+A3pl+P1pl+Nom
aynı aynı+Adj
şekilde şekil+Noun+A3sg+Pnon+Loc
sürdürülecektir sür+Verb^DB+Verb+Caus^DB+Verb+Pass+Pos+Fut+A3sg+Cop
. .+Punc
```

Figure 4.7: The output of morphological disambiguator

For tagged and tagged-reordered systems (shown in Table 4.1), we use `lemma+MORPH (Surface Form)` representation on Turkish side of the parallel corpus which means a word is represented as its lemma plus the morpheme tags, as opposed to the original form of the words. An example is shown below.

ortak -> *ortak*+Noun+A3sg+Pnon+Nom

piyasa -> *piyasa*+Noun+A3sg+Pnon+Nom

düzenlemesi -> *düzenle*+Verb+PosD̂B+Noun+Inf2+A3pl+P3sg+Nom

4.4 Part-of-speech Tagging of English Data

English does not have a complex morphology, so English side of the parallel corpus is part-of-speech (POS) tagged to match the morphologically tagged Turkish data. We use *Enju Parser* to parse the English data and find the POS tags of English words.

For tagged and tagged-reordered systems (shown in Table 4.1), we use `lemma+POS (Surface Form)` representation on the English side of the parallel corpus, as opposed to the original form of the words. An example of POS-tagged English data is shown below.

this -> *this*+DT

ratio -> *ratio*+NN

is -> *be*+VBZ

represented -> *represent*+VBN

4.5 Preparing Factored Data

We create a factored representation of the data in order to build factored translation systems as explained in Axelrod (2006) [7] and Koehn and Hoang (2007) [38]. In this representation, the data is not only represented by surface forms but also lemmas and POS/Morpheme tags. These three forms are used as factors for the factored translation systems. In the factored data, factors are separated by a `|` symbol and we represent the data as `Surface_Form|Lemma|POS/Morpheme_Tags`. Table 4.2 illustrates examples of all these representations.

Table 4.2: All representations of data

Representation	**English Data/Turkish Data**
Surface Form	country+NNS ülke+Noun+A3pl+Pnon+Nom
Lemma	country ülke
POS/Morpheme Tag	NNS Noun+A3pl+Pnon+Nom
Factored	country+NNS|country|NNS ülke+Noun+A3pl+Pnon+Nom|ülke|Noun+A3pl+Pnon+Nom

4.6 Building Systems

4.6.1 Creating Language Model

As explained in Section 2.2.3, language model is only needed for target language, therefore we create language models for Turkish using IRSTLM toolkit [22].

We use 5-gram language models that perform better than the other ones as explained in Section 5.8. We create the language models for our different systems with Kneser-

Ney smoothing algorithm [14] and binarize the models to shrink the size. The binarization step is done using the scripts provided with *Moses* toolkit [39].

For our six different systems, four language models are created. These are created respectively from the lemmas, morphological tags, surface forms and the data itself. Some examples from these four language models are shown in Figure 4.8. The n-gram counts for the language models are shown in Table 4.3.

```
Surface Forms 5-Gram LM

-0.00116351   için+Postp+PCNom önlem+Noun+A3pl+Pnon+Nom al+Verb^DB+Verb+Pass+Pos+Neces+A3sg+Cop .+Punc </s>
-0.265813     için+Postp+PCNom inşa+Noun+A3sg+Pnon+Nom et+Verb^DB+Verb+Pass+Pos+Narr+A3sg ve+Conj kamu+Noun+A3sg+Pnon+Nom
-0.00757507   için+Postp+PCNom henüz+Adverb çok+Adverb erken+Adj^DB+Verb+Zero+Pres+A3sg+Cop .+Punc
-0.112539     için+Postp+PCNom henüz+Adverb erken+Adj^DB+Verb+Zero+Pres+A3sg+Cop fakat+Conj bu+Det
-0.163052     için+Postp+PCNom makroekonomik+Adj istikrar+Noun+A3sg+Pnon+Nom ortam+Noun+A3sg+P3sg+Gen devam+Noun+A3sg+P3sg+Nom


Morphological Tags 5-Gram LM

-0.0363788    Noun+A3sg+P3sg+Loc Verb+Pos^DB+Adj+PastPart+P3sg Noun+A3sg+Pnon+Nom Verb^DB+Verb+Pass+Pos+Neces+A3sg+Cop Punc
-0.591635     Noun+A3sg+P3sg+Loc Verb+Pos^DB+Adj+PastPart+P3sg Det Noun+A3sg+Pnon+Nom Postp+PCNom
-0.0780051    Noun+A3sg+P3sg+Loc Verb+Pos^DB+Adj+PastPart+P3sg Det Adj Noun+A3sg+Pnon+Nom
-0.221735     Noun+A3sg+P3sg+Loc Verb+Pos^DB+Adj+PastPart+P3sg Conj Conj Noun+A3sg+Pnon+Nom
-0.149146     Noun+A3sg+P3sg+Loc Verb+Pos^DB+Adj+PastPart+P3sg Conj Verb+Pos^DB+Adj+PastPart+P3sg Punc


Lemmas 5-Gram LM

-0.252155     paylaş değer çerçeve güç esas
-0.276159     paylaş Azerbaycan yakın bir ortaklık
-0.251981     makul ol bu sonuç var
-0.0745453    makul ol şekil müdahale et
-0.228644     makul ol neden ver imtina


Original Data 5-Gram LM

-0.141865     uzman desteği ile gerçekleştirilen altı
-0.253692     uzman gönderilmesi ve eğitimi gibi
-0.180547     kullanımına sunulan bütün borç senetleri
-0.417881     kullanımına ilişkin idari kapasite geliştirilecektir
-0.00593203   kullanımına yönelik onay ya da
```

Figure 4.8: Example segments from language models

Table 4.3: N-gram counts in the Turkish language model

N-Gram Order	Surface Forms LM N-Gram Count	Morpheme Tags LM N-Gram Count	Lemmas LM N-Gram Count	Original Data LM N-Gram Count
1	57290	3024	14419	66519
2	356189	41445	218539	408779
3	97780	65766	108190	115316
4	75567	103508	85042	89769
5	58803	101263	65250	70607

4.6.2 Training Translation Systems

Training is the most important step where the systems learn how to translate by estimating the parameters of the statistical translation model with the use of parallel corpus. During this step, translation models are trained by computing word-alignments, extracting and scoring phrase tables and creating reordering tables. The training data for each dataset is selected randomly from the parallel corpus, so translation models are unique and recreated for each dataset. Moses toolkit [39], is used to train the systems and create the translation models.

The first step in training, is to create vocabulary files for both languages, these contain word identifiers (integers), words and word counts. With the help of these vocabulary files, the parallel corpus is digitalized (each word is represented with its integer identifier) and a sentence-aligned corpus file is created with the word identifiers. Figure 4.9 and Figure 4.10 are examples of vocabulary files and sentence-aligned corpus file. In the vocabulary files, the first column is the word identifier, the second column is the word itself and the third column is the count of that word in the corpus. In sentence-aligned corpus file, a sentence pair is shown in three lines where first is the frequency of the sentence, second and third are the sentences from both languages where words are represented with their identifiers from the vocabulary files.

```
tr.vcb
--------------------------
1       UNK         0
2       ,           50015
3       .           42109
4       ve          33432
5       bir         12678
6       )           7525
7       (           7008
8       ile         6749
9       olarak      6186
10      bu          6095
11      için        5688
12      -           5346
13      ilgili      3914
14      veya        3864
15      ilişkin     3623
```

```
en.vcb
-------------------
1       UNK     0
2       the     87380
3       of      52345
4       ,       44763
5       .       39517
6       and     36053
7       to      29312
8       in      27021
9       a       12468
10      on      10931
11      -       10784
12      for     10633
13      be      9454
14      is      8378
15      )       8232
```

Figure 4.9: Example segments from vocabulary files

In the second step, *mGiza* [24] is run to find the word alignments. *mGiza* is a tool based on *GIZA++* [55], extended to support multi-threading and incremental training. *mGiza* is run bidirectionally and takes the intersection of these two runs to find the correct word alignments. An example word-alignment diagram is shown in Table

```
tr-en-int-train.snt
-----------------------------------------------------------------------------------------------
1
7421 3306 1457 1264
501 2342 266 5137 932 117 1579 8518
1
2 893 512 610 395 18 2 90 10 841 300 126 4 32 137 18 2 526 379 10 384 480 126 8 3293 5
700 204 134 64 102 22 935 37318 1481 501 2342 295 593 932 2 935 26092 2908 11719 31 144 3
1
8 276 3 24 11 39 568 4 27 1284 14 120 3 2 222 1787 1305 18 2 10096 3 2 1308 67 71 682 8 293 144 5
10 1147 2 16 12 23 1895 305 2 156 148 1185 23 92 4 192 3743 2913 9546 304 34 10999 6423 3
1
2 893 512 610 63 5447 20 2 1057 5862 16 2 144 1308 67 71 25 532 21 1203 5
501 2342 295 593 932 2 156 1185 2913 8 1006 250 4302 3213 37 29790 3086 17 5 40212 3
1
2 1284 440 28 2448 1787 20 1537 535 3828 5
765 11553 2 54333 3984 5301 656 3
```

Figure 4.10: Example segment from digitalized sentence-aligned corpus

4.4.

Table 4.4: Example word-alignment

	Adopt	a	programme	for	transposition	of	the	transport	acquis	.
Ulaştırma								■		
konusundaki							■			
müktesabatın									■	
üstlenilmesi					■	■				
için				■						
bir		■								
program			■							
benimsenmesi	■									
.										■

In the third step, phrases are extracted from these word alignments, which is done for both languages. Neighbouring words, that occur together in the data, are extracted as phrases. Here we use the default maximum phrase length for Moses, which is seven. Each phrase pair is assigned alignment points, showing the number of word-alignments in this pair. Then these pairs are sorted for both languages and scored with the calculated probability of that translation. At the end, a translation table is created containing all the extracted phrase pairs and their scores. A small portion from the translation table is shown in Figure 4.11. The four different scores seen at the end of every line are; inverse phrase translation probability, inverse lexical weighting, direct phrase translation probability and direct lexical weighting.

```
in compliance with the EU acquis ||| AB kuralları ile uyumlu bir ||| 0.5 1.8307e-06 0.333333 3.40816e-08
in compliance with the EU acquis ||| AB kuralları ile uyumlu ||| 0.5 1.8307e-06 0.333333 1.36449e-06
in compliance with the EU acquis ||| AB mevzuatına uygun ||| 0.2 1.20453e-05 0.333333 0.00135892
in compliance with the EU criteria ||| , AB kriterleriyle uyumlu ||| 0.5 8.0725e-07 0.5 2.25346e-05
in compliance with the EU criteria ||| AB kriterleriyle uyumlu ||| 0.5 8.0725e-07 0.5 0.000226142
in compliance with the EU legislation ||| AB mevzuatına uygun olarak ||| 1 3.4027e-06 1 0.000250681
```

Figure 4.11: A portion from the phrase translation table

At the end of training step, a configuration file is created for the Moses decoder. This file specifies, the factors used, the translation tables, the reordering models, the generation models, language models used and the weights of all these models.

4.6.2.1 Factored Training

In factored translation systems, multiple phrase translation tables are created according to the translation factors given to Moses toolkit. Moses accepts a set of factor lists as translation factors and creates a translation table for each element in the set. For example, if the translation factors are given as `0-0+1-1`, then translation tables will be created from source factor 0 to target factor 0 and from source factor 1 to target factor 1. This means, from the factored training data, a translation table will be created for the first factors and another will be created for the second factors. In this study, translation factors are given as `0-0+1-1+2-2`, which means separate translation tables are created for surface forms (factor 0), lemmas (factor 1) and POS/Morpheme tags (factor 2). An illustrative example of how translation factors work is shown in Figure 4.12.

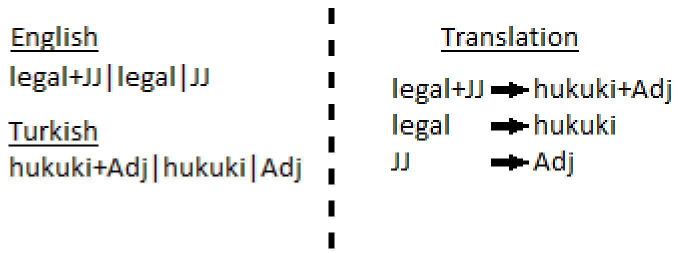

Figure 4.12: Translation factors

Factored systems are also used to train word alignments and create generation models for factors. In this study, only lemmas are used to train the word alignments as they are the smallest counterparts of the words and they occur more general than surface forms or POS/Morpheme tags in the parallel corpus. Generation models are used to decide which target side factors will be used to generate other target side factors. In this study, `0-1,2+1,2-0` is given as generation factors, which means two generation models are created, one for generating a lemma and morpheme tags from a surface form and one for generating a surface form from a pair of lemma and morpheme tags. Figure 4.13 shows an illustrative example of how generation factors work.

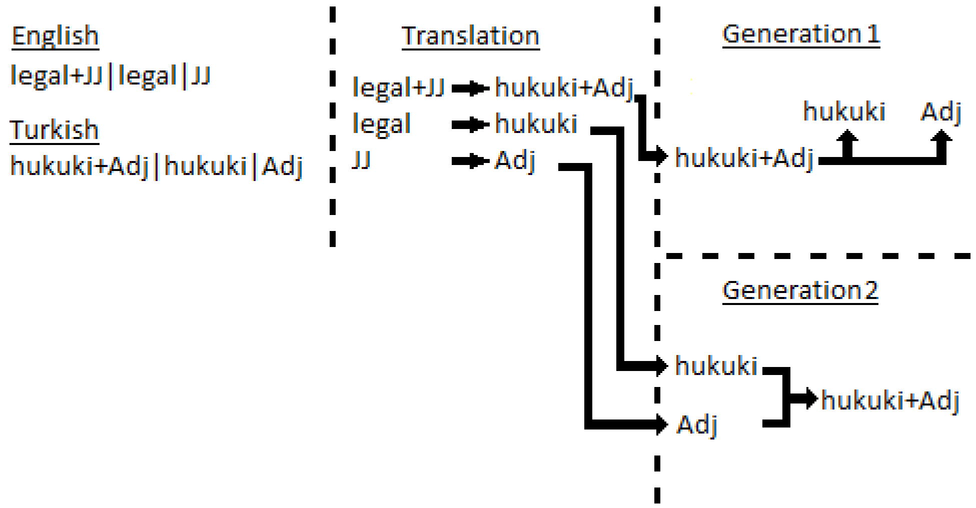

Figure 4.13: Generation factors

4.6.3 Tuning

Moses uses the created models with different weights. These weights are recorded in the configuration file which is created after the training step. However, in the configuration file, these weights have default values and they need to be optimized. This optimization is done at the tuning step in order to find better rates and achieve high quality translations. In this study, minimum error rate training (*MERT*) is used to tune the systems [54], [9]. MERT runs the Moses decoder on the same tuning data several times and finds the best set of weights that provide the best quality translations in terms of BLEU scores. In this study, the number of maximum iterations of MERT is limited to 10 for all systems because tuning is the slowest step of the process.

Tuning of factored systems is similar to tuning of the phrase-based systems. The configuration file contains input and output factors and the system is tuned according to these factors. In non-factored phrase-based systems, the input and output factors in the configuration file are both 0, which means there is only one factor used.

4.6.4 Decoding

Decoding is simply creating the translation hypothesis from the input text. Using the translation and language models, the decoder machine, namely the Moses toolkit decoder, looks for the best translation of an input text. Moses uses beam search algorithm which allows to keep all the hypotheses in stack data structures according to their translated word counts. Each time a hypothesis is placed into a stack, the stack may need pruning which means hypotheses with lower scores are removed and then new hypotheses are created from that hypothesis. This is done in iterations until all the input words are translated into target language.

4.6.4.1 Factored Decoding

Factored systems allow us to have multiple decoding paths. In this study, we use two decoding paths:

1. `t0,g0`
2. `t1,t2,g1`

`t0` is the translation step between surface forms, `t1` between lemmas, `t2` between POS-Morpheme tags. `g0` is the generation step where a lemma and the morpheme tags are generated from a surface form and `g1` is the generation step where a surface form is generated from a lemma and the morpheme tags. The first decoding path uses `t0` and `g0` and the second one uses `t1,t2` and `g1`. The two paths are alternatives to each other. During decoding, translation options are generated from each and the one with the higher probability is used. These decoding paths are illustrated in Figure 4.14.

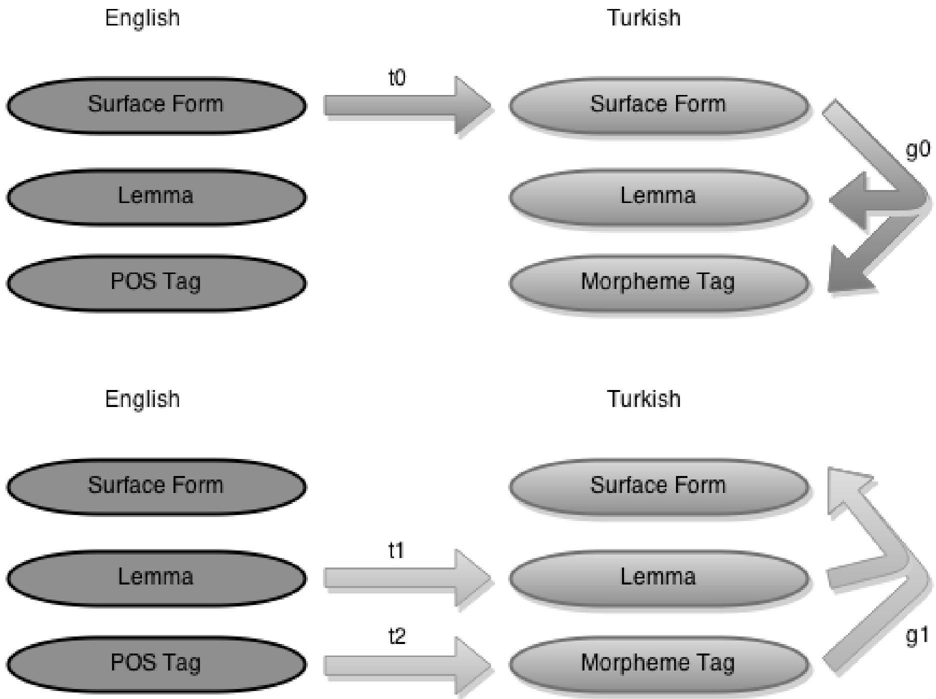

Figure 4.14: Decoding paths for the factored system

4.6.5 Evaluation

The automatic evaluation of machine translation systems has always been challenging throughout the history of the task itself as one sentence can have multiple valid translations in the target language. In this study, for evaluating the systems built and measuring the quality of translations of test data, BLEU metric is used [59]. BLEU algorithm takes the n-grams from the candidate sentence and tries to match them in the reference sentence. The more n-grams matched, the higher the candidate sentence's score is. The score of a candidate translation is the maximum number of matched n-grams from the candidate divided by the total number of n-grams in the candidate.

In order to find the BLEU score, first the geometric average of the modified n-gram precisions, p_n, is computed using n-grams up to length N and positive weights w_n summing to one [59]. c is the length of the candidate translation and r is the effective reference corpus length. The brevity penalty BP is computed as shown in Formula

4.1.

$$BP = \begin{cases} 1, & \text{if } c > r \\ e^{(1-r/c)}, & \text{if } c \leq r \end{cases} \tag{4.1}$$

Finding Brevity Penalty for BLEU Score

Then, BLEU score is found as shown in Formula 4.2.

$$BLEU = BP \cdot exp(\sum_{n=1}^{N} w_n log p_n) \tag{4.2}$$

Finding BLEU Score

CHAPTER 5

EXPERIMENTS

In this study, we build six English to Turkish statistical machine translation systems. These systems are:

1. Baseline System
2. Baseline-Reordered System
3. Tagged-Baseline System
4. Tagged-Reordered System
5. Factored-Baseline System
6. Factored-Reordered System

We perform experiments with ten data sets for each system. These data sets are created from Oflazer's corpus [80], as explained in Section 4.1. We also experiment our systems with SETimes corpus [70]. We present the comparable results for each system over all data sets, error analysis and discussion over the results in this chapter.

5.1 Baseline System

The baseline system is built with the original data, without applying any pre-processing steps. All datasets are trained, tuned and tested with Moses and the BLEU scores obtained are shown in Table 5.1.

Table 5.1: BLEU scores of baseline system

Dataset	Baseline System
0	20.12
1	19.21
2	20.91
3	20.16
4	19.39
5	19.32
6	19.27
7	18.30
8	20.23
9	19.31
Average	**19.62**

5.2 Baseline-Reordered System

To build the baseline-reordered system, English data is head-finalized as mentioned in Section 4.2. The Turkish side of the parallel corpus is left as it is. Reordering is done over all ten datasets. The BLEU scores and comparisons with the baseline system are shown in Table 5.2.

Table 5.2: BLEU scores of baseline-reordered system

Dataset	Baseline System	Baseline-Reordered System
0	20.12	20.73
1	19.21	21.02
2	20.91	22.00
3	20.16	21.06
4	19.39	20.82
5	19.32	20.28
6	19.27	19,85
7	18.30	19,63
8	20.23	21,63
9	19.31	19,84
Average	**19.62**	**20.69**

As seen from the results, with *Head Finalization* of English data, the average BLEU score has increased by **1.07** points which corresponds to an improvement of **5.42%**.

5.3 Tagged-Baseline System

In this system, the Turkish portion of the parallel corpus is first morphologically analyzed and disambiguated. The words are represented as `lemma+MORPH`. On the other hand, the English side of the data is POS tagged and represented as `lemma+POS`. These representations are called the *surface forms* for both languages. These pre-processing steps are applied to all ten datasets. The comparative BLEU scores with the previous systems are shown in Table 5.3

Table 5.3: BLEU scores of tagged-baseline system

Dataset	Baseline System	Baseline-Reordered System	Tagged-Baseline System
0	20.12	20.73	24.37
1	19.21	21.02	24.24
2	20.91	22.00	25.64
3	20.16	21.06	25.65
4	19.39	20.82	26.20
5	19.32	20.28	24.97
6	19.27	19,85	24.16
7	18.30	19,63	23.77
8	20.23	21,63	25.94
9	19.31	19,84	25.64
Average	**19.62**	**20.69**	**25.06**

As we do not apply any reordering in this system, it would be better to compare the results with the baseline system, and we can see that on average the BLEU score has increased by **5.44** points which means an improvement of **27.70%**.

5.4 Tagged-Reordered System

This system is the combination of tagged-baseline system and baseline-reordered system. Both sides of the parallel data is tagged and the English side is head-finalized and reordered. The comparable BLEU scores are shown in Table 5.4

Tagged-reordered system increased the average BLEU score of tagged-baseline system by **2.39** points, which is **9.54%**. Furthermore, if we compare the results of

Table 5.4: BLEU scores tagged-reordered system

Dataset	Baseline-Reordered System	Tagged-Baseline System	Tagged-Reordered System
0	20.73	24.37	26.26
1	21.02	24.24	27.51
2	22.00	25.64	28.07
3	21.06	25.65	27.98
4	20.82	26.20	27.91
5	20.28	24.97	27.08
6	19.85	24.16	26.00
7	19.63	23.77	26.93
8	21.63	25.94	28.25
9	19.84	25.64	28.50
Average	**20.69**	**25.06**	**27.45**

baseline-reordered system and tagged-reordered system, we see the effect of morphological and POS tagging with reordering, which is on average **6.76** points (**32.67%**) in BLEU metric.

5.5 Factored-Baseline System

In this system, Turkish data is represented as `Surface_Form|Lemma|Morpheme_Tags` and English data as `Surface_Form|Lemma|POS_Tags`. Word-alignment is done between `Lemmas`. Two decoding paths, as described in Section 4.6.4.1, are given to the system; one is to translate surface forms and generate lemmas and tags, the other is to translate lemmas and tags and generate surface forms. The results that are comparable to the other non-reordered systems are shown in Table 5.5.

From the results, we see that factored-baseline system has slightly improved the BLEU scores of tagged-baseline system with **1.43** points (**5.71%**).

5.6 Factored-Reordered System

The only difference of this system from factored-baseline system is that it uses head-finalized, factored English data. The results are shown in Table 5.6.

Table 5.5: BLEU scores of factored-baseline system

Dataset	Baseline System	Tagged-Baseline System	Factored-Baseline System
0	20.12	24.37	26.09
1	19.21	24.24	25.89
2	20.91	25.64	27.43
3	20.16	25.65	26.85
4	19.39	26.20	26.75
5	19.32	24.97	26.16
6	19.27	24.16	25.53
7	18.30	23.77	25.38
8	20.23	25.94	27.15
9	19.31	25.64	27.64
Average	**19.62**	**25.06**	**26.49**

The results show that factorizing the tagged and reordered data had an improvement of **3.48** points (**12.68%**) in BLEU score over non-factored tagged-reordered system and an improvement of **4.44** points (**16.76%**) over factored-baseline system. The comparative BLEU scores of all systems over all data sets are shown in Figure 5.1.

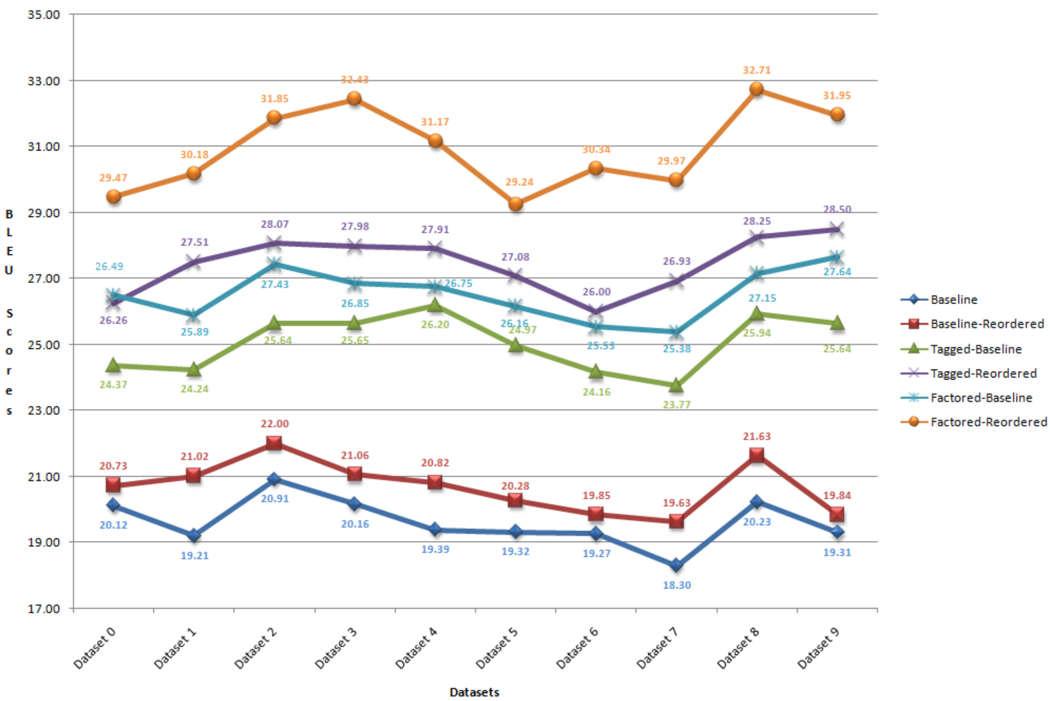

Figure 5.1: BLEU scores of all built systems

Table 5.6: BLEU scores of factored-reordered system

Dataset	Tagged-Reordered System	Factored-Baseline System	Factored-Reordered System
0	26.26	26.09	29.47
1	27.51	25.89	30.18
2	28.07	27.43	31.85
3	27.98	26.85	32.43
4	27.91	26.75	31.17
5	27.08	26.16	29.24
6	26.00	25.53	30.34
7	26.93	25.38	29.97
8	28.25	27.15	32.71
9	28.50	27.64	31.95
Average	**27.45**	**26.49**	**30.93**

5.7 Experiment with SETimes Corpus

In addition to the experiments carried out with Oflazer's corpus [80], we run our systems on the SETimes corpus [70] in order to see the effects of morphology and head-finalization with a larger corpus. The SETimes corpus contains 207,674 parallel sentences. We remove the lines which are too short or too long (we limit the sentence length to three to fifty words). Additionally, we also remove the lines which cannot be successfully parsed by the Enju Parser on the English side and which cannot be morphologically parsed on the Turkish side. After those pre-processing steps, 168,331 parallel lines were left, which was more than three times of Oflazer's corpus. The same experimental setup we create with Oflazer's corpus, is created with the 168K SETimes corpus. The average results for each system are shown in Table 5.7.

Table 5.7: BLEU scores of all systems with SETimes corpus

SMT System	BLEU Score
Baseline	24.41
Baseline-Reordered	30.22
Tagged-Baseline	22.87
Tagged-Reordered	28.73
Factored-Baseline	22.39
Factored-Reordered	30.49

These results show that the effect of morphological analysis decreases as the data gets larger. We infer that the tags complicate the translation process and have a negative effect on the BLEU scores. On the other hand, the contribution of head-finalization is more visible. With a larger corpus, head-finalization has more impact than it has with a small corpus. We can see that baseline-reordered system increases the BLEU score of baseline system by **5.81** points (**23.80%**). Tagged-reordered system has an improvement of **5.86** points (**25.62%**) over tagged-baseline system. And factored-reordered system boosts the score of factored-baseline system by **8.10** points which corresponds to **36.17%**.

5.8 N-Gram Language Model Experiments

We produce 3-gram, 4-gram, 5-gram and 6-gram language models and performed experiments with one of the test data on the factored-reordered system with each language model. The results are given in Table 5.8.

Table 5.8: BLEU scores with different language models

Language Model	**BLEU Score (%)**
3-gram LM	30.07
4-gram LM	31.00
5-gram LM	31.43
6-gram LM	30.21

The results show that, 5-gram language model outperforms the others. We use 5-gram language models in all experiments in this study.

5.9 Error Analysis and Discussion

In this study, experiment results are reviewed and some of them are checked manually to see if the translations are done correctly and as expected. This check is done on the same input sentences that are chosen randomly for all systems. Errors were analyzed in order to find the reason and the erroneous system was retrained with the corrected data if necessary. One of the most common errors we encounter while training the systems was about factorizing the punctuation marks. It is not viable to go over all training data to find the mistyped or faulty punctuation marks, thus we make numerous corrections at each training step. Figure 5.2 shows an example of those errors.

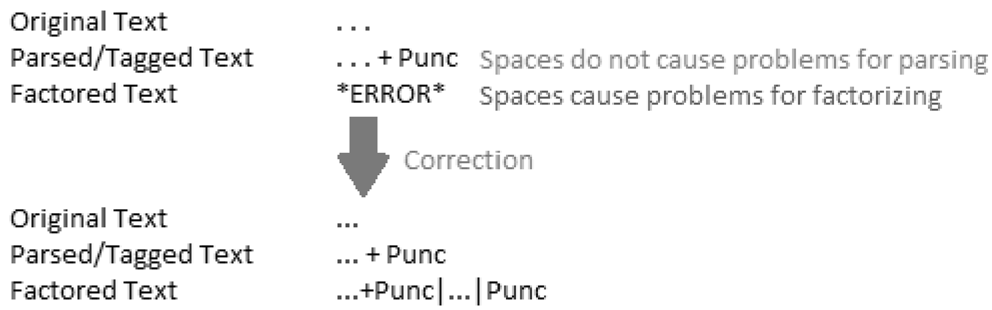

Figure 5.2: An example of punctuation errors

We analyzed the output of the testing steps to see where the systems are producing faulty translations. Below, in Table 5.9 and 5.10, two example input sentences, the expected outputs and the outputs of all systems are given. In order to gain space, the following abbreviations are used in the examples:

Inp. : Input Sentence

Exp. Out. : Expected Output Sentence

Bas. : Baseline System Output

Bas. Reo. : Baseline-Reordered System Output

Tag. : Tagged-Baseline System Output

Tag. Reo. : Tagged-Reordered System Output

Fac. : Factored-Baseline System Output

Fac. Reo. : Factored-Reordered System Output

In the example shown in Table 5.9, we must note that even the expected output is not a perfect translation. The expected sentence has the word " `kriter`", however, the input sentence does not contain a word " `criterion`" which is the translation of this word. Thus, we expect to see that the system outputs do not contain that word. Nevertheless, the baseline system output contains " `kriter`" in a false positive way. When we check the phrase-table of baseline system, we see that " `sufficient amount of`" is translated as " `kriter, yeterli`" with a probability of 0.67 which explains the wrong translation. We believe that this problem may be solved with

Table 5.9: Example outputs of all systems

Inp.	It also requires a sufficient amount of human and physical capital, including infrastructure.
Exp. Out.	Ayrıca bu kriter, yeterli bir beşeri ve altyapıyı da içeren bir fiziki sermayenin mevcudiyetini gerektirmektedir.
Bas.	Güçlü bir kararlılık gerektirir ayrıca kriter, yeterli beşeri ve altyapıyı da içeren fiziki sermayeyi.
Bas. Reo.	, Bu aynı zamanda, altyapı dahil beşeri ve maddi sermayesinin yeterli miktarda gerektirmektedir.
Tag.	Botaş'ı ayrıca bu kriter, yeterli bir beşeri ve alyapıyı da içeren bir fiziki sermayenin.
Tag. Reo.	Bu, ayrıca, altyapı dahil yeterli bir beşeri ve bir fiziki sermayenin mevcudiyetini gerektirmektedir.
Fac.	Ayrıca, gerektirmektedir yeterli bir beşeri ve altyapıyı da içeren bir fiziki sermayenin.
Fac. Reo.	Bu, ayrıca, altyapı dahil yeterli bir beşeri ve fiziki sermaye tutarı gerektirmektedir.

a larger and well-organized parallel corpus. Tagged-baseline system has a similar problem. The output contains the word " `Botaş'ı`" while there is no sign of " `Botaş`" in the input. Other than these examples, we observe that all baseline systems perform badly regarding the word ordering. Reordered systems seem to perform better in this situation. For example, in baseline output and factored-baseline output, we see that the translation of the verb " `requires`", which " `gerektir-`", is placed in the wrong place according to the expected output. This supports our claim evident with BLEU scores that the reordered systems outperform the baseline ones. We conlclude that head-finalization works well for the word order problem.

When we look at the example shown in Table 5.10, the first error we see is in the baseline system output. Although the input does not contain the word " `Romania`", the output contains " `Romanya'ya`" which means " `to Romania`". We analyzed the phrase-table and saw that the phrase " `to conclude readmission agreements with`" was translated into " `Romanya'ya geri kabul anlaşması`" which was caused by a long training sentence. And as Moses prioritize longer phrases, it did not continue with shorter chunks of this phrase when it found the faulty translation. This is one of the examples that shows the effects of insufficient data. Limiting the maximum phrase

Table 5.10: Example outputs of all systems

Inp.	Negotiations to conclude readmission agreements with Bulgaria and Russia have continued.
Exp. Out.	Bulgaristan ve Rusya ile geri kabul anlaşmaları imzalanmasına yönelik müzakereler sürmüştür.
Bas.	Müzakerelerin Romanya'ya geri kabul anlaşması Bulgaristan ve Rusya devam edilmiştir.
Bas. Reo.	İle Rusya'yı Bulgaristan ve geri kabul anlaşmaları akdetme müzakereler devam etmiştir.
Tag.	Geri kabul anlaşması müzakereleri Bulgaristan ve Rusya ile devam etmektedir.
Tag. Reo.	Bulgaristan ve Rusya ile geri kabul anlaşması müzakereleri devam etmiştir.
Fac.	Müzakerelerin yeniden kabul anlaşmaları yapmak ile Bulgaristan ve Rusya devam etmiştir.
Fac. Reo.	Bulgaristan ve Rusya ile geri kabul anlaşmaları akdedilmesine yönelik müzakereler devam etmektedir.

length of Moses is a viable solution. We use the default phrase length which is seven, but smaller values can be used to see the effects. The second significant error is in the baseline-reordered system output. The phrase " `with Bulgaria and Russia`" is translated to " `ile Rusya'yı Bulgaristan ve`", which is a completely wrong translation. At first sight, this looks like a reordering problem. However, when we check the reordered output sentence, we see that the input sentence is correctly reordered to " `Bulgaria and Russia with readmission agreements conclude to Negotiations continued have.`", which means this is not because of our head-finalization reordering. The reason is that Moses cannot find the complete phrase in the phrase-table and searches for smaller pieces of it. Unfortunately, the phrase-table contains the phrases " `Russia with`" and " `Bulgaria and`" separately. However, Moses decoder puts " `Russia with`" in front of " `Bulgaria and`" while decoding according to the reordering model computed from the training data. This also shows the effects of insufficient data. The results of the factored-baseline system output shows how important reordering is. The rest of the outputs are all acceptable in terms of meaning. However, we see that the factored-reordered system output is almost identical to the expected output. The only difference is the translation of two words (" `imzalanmasına`" and " `sürmüştür`") but this can be ignored as synonyms of the translations are used (" `akdedilmesine`" and " `devam etmektedir`").

Generally, the erroneous translations are caused by the lack of a large parallel corpus. Out-of-vocabulary words and unseen phrases reduce the accuracy as they are not translated and left as they are. Another important point to check is the contribution of morphological analysis. Although morphological analysis improves the accuracy when the smaller data set is used, on SETimes corpus which is three times larger than Oflazer's corpus, we see that the effects of morphological analysis reduces significantly. The baseline and baseline-reordered systems produce better results than the tagged and tagged-reordered counterparts when trained with the larger SETimes corpus. This result complies with the findings of Tatlıcıoğlu (2013) where he uses the same data set [68]. Another issue is that, we do not segment the morphological tags from the base forms and use them as one word which still means an English phrase can be aligned to a single Turkish word. However, segmentation of morphological tags can also be used to improve the effects of morphological analysis [20]. Cakici (2012) provides a morphological segmentation data for METU-Sabancı Turkish Treebank and suggests that the use of lexical forms for morphemes will solve some of the sparse data problems [4]. This idea can also be applied to our work to see more concrete effects of morhological segmentation. Cakici (2008) explores the idea of treating morpheme groups as separate nodes in dependency trees which evokes the use of factored models in this study [3].

The slight decrease may also be caused by the accuracy of the morphological disambiguator. Neither the morphological analyzer [57] nor the averaged perceptron-based morphological disambiguator [62], always give the correct output although they are the state-of-the-art tools used in the literature for Turkish language.

Regardless of morphological analysis and disambiguation, we see that head-finalization works well and boosts the BLEU scores in all systems. The effect is more obvious with larger data sets.

CHAPTER 6

CONCLUSIONS

In this study, a new approach is introduced to the task of English to Turkish factored phrase-based statistical machine translation. With this new approach, we try *Head Finalization* for this language pair and additionally exploited morphology by means of a factored model in order to improve the quality of translations. Experiments were carried out with two different data sets, one with Oflazer's corpus containing 54,391 parallel sentences and the other with SETimes corpus containing 168,331 parallel sentences. The results show that using *Head Finalization* on the English side, improves the BLEU scores of the translations significantly whether or not we use morphology on the Turkish side. The best results we have with Oflazer's corpus surpass what Yeniterzi and Oflazer (2010) [80] obtained with the same data, increasing the BLEU score from 23.78 to 30.93. We also improve the accuracy up to 30.49 BLEU score with SETimes corpus, whereas Tatlıcıoğlu (2013) [68] obtained 26.22 BLEU score with the same data.

While testing the systems with the SETimes data set, we see that the baseline system was slightly more accurate than the factored-baseline system. In the factored models, we use the morphologically tagged representation of Turkish words as surface forms. When we investigate further into this, we see that the tagged systems also perform worse that the baseline systems. This is because of the negative effect of using morphological analysis with a large data set. In order to obtain better results with factored models, an additional experiment can be done with adding the original data as the fourth factor and an extra decoding step for using that to make use of the already known translations of the original data. We believe that this can improve the results as the baseline systems are more successful than the tagged systems, we will be experimenting this in our future studies.

Moreover, feature reduction segmentation of morphological tags may also improve the results. These methods are introduced in the literature and have been applied successfully in previous research.

The techniques used and suggested throughout this study can be applied for other head-final languages, especially members of the Turkic language family. We have a strong belief that this study can form a basis for future research in the field.

To summarize, this study introduces an approach of applying *Head Finalization* on the English data and *Morphological Analysis and Disambiguation* on the Turkish data in order to build an English to Turkish statistical machine translation system. The results of this study are compared to the benchmark systems which were built with the same data sets and are found to be significantly higher than those.

REFERENCES

[1] Web Inventory of Transcribed and Translated Talks. `https://wit3.fbk.eu/score.php?release=2014-01`. Accessed: 2015-04-30.

[2] *Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)*, Lake Tahoe, December 4th and 5th 2014.

[3] R. Çakıcı. *Wide-Coverage Parsing for Turkish*. PhD thesis, University of Edinburgh, Edinburgh, Scotland, 2008.

[4] R. Çakıcı. Morpheme Segmentation in the METU-Sabanc&Inodot; Turkish Treebank. In *Proceedings of the Sixth Linguistic Annotation Workshop*, LAW VI '12, pages 144–148, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

[5] M. T. Çakmak, S. Acar, and G. Eryiğit. Word Alignment for English-Turkish Language Pair. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA).

[6] D. Arnold, L. Balkan, R. L. Humphreys, S. Meijer, and L. Sadler. *Machine Translation: An Introductory Guide*. Blackwell Pub, January 1994.

[7] A. Axelrod. Factored Language Model for Statistical Machine Translation. Master's thesis, Edinburgh University, Edinburgh, Scotland, 2006.

[8] Y. Bar-Hillel. The present state of research on mechanical translation. In *American Documentation 2*, pages 229–237, 1951.

[9] N. Bertoldi, B. Haddow, and J. Fouet. Improved minimum error rate training in Moses. *The Prague Bulletin of Mathematical Linguistics*, 2009.

[10] A. Bisazza and M. Federico. Morphological pre-processing for Turkish to English statistical machine translation. In *Proceedings of the International Workshop on Spoken Language Translation, Tokyo, Japan, December*, 2009.

[11] P. Brown, J. Cocke, S. D. Pietra, V. D. Pietra, F. Jelinek, R. Mercer, and P. Roossin. A Statistical Approach to Language Translation. In *Proceedings*

of the 12th Conference on Computational Linguistics - Volume 1, COLING '88, pages 71–76, Stroudsburg, PA, USA, 1988. Association for Computational Linguistics.

[12] P. F. Brown, J. Cocke, S. A. D. Pietra, V. J. D. Pietra, F. Jelinek, J. D. Lafferty, R. L. Mercer, and P. S. Roossin. A Statistical Approach to Machine Translation. *Comput. Linguist.*, 16(2):79–85, June 1990.

[13] P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Comput. Linguist.*, 19(2):263–311, June 1993.

[14] S. F. Chen and J. Goodman. An Empirical Study of Smoothing Techniques for Language Modeling. In *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics*, ACL '96, pages 310–318, Stroudsburg, PA, USA, 1996. Association for Computational Linguistics.

[15] D. Chiang. A Hierarchical Phrase-based Model for Statistical Machine Translation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 263–270, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.

[16] B. J. Dorr, E. H. Hovy, and L. S. Levin. Machine Translation: Interlingual Methods.

[17] İ. Durgar El-Kahlout and K. Oflazer. Aligning Turkish and English Parallel Texts for Statistical Machine Translation. In pInar Yolum, Tunga Güngör, Fikret Gürgen, and Can Özturan, editors, *Computer and Information Sciences - ISCIS 2005*, volume 3733 of *Lecture Notes in Computer Science*, pages 616–625. Springer Berlin Heidelberg, 2005.

[18] İ. Durgar El-Kahlout and K. Oflazer. Initial Explorations in English to Turkish Statistical Machine Translation. In *Proceedings on the Workshop on Statistical Machine Translation*, pages 7–14, New York City, June 2006. Association for Computational Linguistics.

[19] I. Durgar El-Kahlout, C. Mermer, and M. U. Dogan. Recent Improvements In Statistical Machine Translation Between Turkish and English. In *Multilingual Processing in Eastern and Southern EU Languages.- Lowresourced Technologies and Translation*. Cambridge Scholars Publishing, 2012.

[20] I. Durgar El-Kahlout and K. Oflazer. Exploiting Morphology and Local Word Reordering in English-to-Turkish Phrase-Based Statistical Machine Translation. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(6):1313–1322, Aug 2010.

[21] A. Eisele, C. Federmann, H. Saint-Amand, M. Jellinghaus, T. Herrmann, and Y. Chen. Using Moses to Integrate Multiple Rule-based Machine Translation Engines into a Hybrid System. In *Proceedings of the Third Workshop on Statistical Machine Translation*, StatMT '08, pages 179–182, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[22] M. Federico and M. Cettolo. Efficient Handling of N-gram Language Models for Statistical Machine Translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, StatMT '07, pages 88–95, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.

[23] L. Formiga, A. Hernandez, J. B. Mari, and E. Monte. Improving English to Spanish out-of-domain translations by morphology generalization and generation. In *AMTA Workshop on Monolingual Machine Translation*, San Diego, United States, 2012.

[24] Q. Gao and S. Vogel. Parallel Implementations of Word Alignment Tool. In *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, SETQA-NLP '08, pages 49–57, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[25] O. Gorgun and O.T. Yildiz. Using morphology in English-Turkish statistical machine translation. In *Signal Processing and Communications Applications Conference (SIU), 2012 20th*, pages 1–4, April 2012.

[26] I. Goto, M. Utiyama, and E. Sumita. Post-ordering by Parsing for Japanese-English Statistical Machine Translation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 311–316. Association for Computational Linguistics, 2012.

[27] Y. Graham, J. van Genabith, and A. Bryl. F-Structure Transfer-Based Statistical Machine Translation. In Miriam Butt and Tracy Holloway King, editors, *Proceedings of the LFG09 Conference*. CSLI Publications, 2009.

[28] D. Groves and A. Way. Hybrid data-driven models of machine translation. *Machine Translation*, 19(3-4):301–323, 2005.

[29] B. Hoffman. *Computational Analysis of the Syntax and Interpretation of 'Free' Word-order in Turkish*. PhD thesis, University of Pennsylvania, Philadelphia, PA, USA, 1995. IRCS Report 95-17.

[30] G. Hong, S. Lee, and H. Rim. Bridging Morpho-syntactic Gap Between Source and Target Sentences for English-Korean Statistical Machine Translation. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 233–236, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[31] W. J. Hutchins. *Machine Translation: Past, Present, Future*. John Wiley & Sons, Inc., New York, NY, USA, 1986.

[32] W. J. Hutchins. The Georgetown-IBM experiment demonstrated in January 1954. In *R. E. Frederking and K. Taylor, editors, AMTA, volume 3265 of Lecture Notes in Computer Science*, pages 102–114. Springer, 2004.

[33] H. Isozaki, K. Sudoh, H. Tsukada, and K. Duh. Head Finalization: A Simple Reordering Rule for SOV Languages. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, WMT '10, pages 244–251, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[34] F. Jelinek and R. L. Mercer. Interpolated estimation of Markov source parameters from sparse data. In *In Proceedings of the Workshop on Pattern Recognition in Practice*, pages 381–397, Amsterdam, The Netherlands: North-Holland, May 1980.

[35] R. Johnson. Contemporary Perspectives in Machine Translation. In *Human translation, machine translation. Papers from 10th annual conference on computational Linguistics, Odense, Denmark, 22-23 November 1979*, pages 133–147. ed. Suzanne Hanon and Viggo Hjørnager Pedersen, 1979.

[36] D. "Jurafsky and H. J." Martin. *SPEECH and LANGUAGE PROCESSING An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, May 2008.

[37] R. Kneser and H. Ney. Improved backing-off for M-gram language modeling. In *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, volume 1, pages 181–184 vol.1, May 1995.

[38] P. Koehn and H. Hoang. Factored translation models. In *In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 868–876, 2007.

[39] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 177–180, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.

[40] P. Koehn, F. J. Och, and D. Marcu. Statistical Phrase-based Translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03, pages 48–54, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.

[41] Tsujii Laboratory. Enju — A practical HPSG parser, 2007.

[42] A. Lavie, S. Vogel, L. Levin, E. Peterson, K. Probst, A. F. Llitjós, R. Reynolds, J. Carbonell, and R. Cohen. Experiments with a Hindi-to-English Transfer-based MT System Under a Miserly Data Scenario. 2(2):143–163, June 2003.

[43] J. Lee, D. Lee, and G. G. Lee. Improving Phrase-based Korean-English Statistical Machine Translation at INTERSPEECH, 2006.

[44] C. Li, M. Li, D. Zhang, M. Li, M. Zhou, and Y. Guan. A Probabilistic Approach to Syntax-based Reordering for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 720–727. Association for Computational Linguistics, 2007.

[45] Cettolo M., C. Girardi, and M. Federico. WIT^3: Web Inventory of Transcribed and Translated Talks. In *Proceedings of the 16th Conference of the European Association for Machine Translation (EAMT)*, pages 261–268, Trento, Italy, May 2012.

[46] D. Marcu and W. Wong. A Phrase-based, Joint Probability Model for Statistical Machine Translation. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 133–139, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

[47] C. Mermer, H. Kaya, and M. U. Dogan. The TÜBİTAK-UEKAE Statistical Machine Translation System for IWSLT 2009. In *Proceedings of the 6th International Workshop on Spoken Language Translation, IWSLT'09*, Tokyo, Japan, December 2009.

[48] C. Mermer, H. Kaya, and M. U. Dogan. The TÜBİTAK-UEKAE Statistical Machine Translation System for IWSLT 2010. In *Proceedings of the 7th International Workshop on Spoken Language Translation, IWSLT'10*, Paris, France, December 2010.

[49] C. Mermer, H. Kaya, I. D. El-Kahlout, and M. U. Dogan. The TÜBİTAK-UEKAE Statistical Machine Translation System for IWSLT 2012. In *Proceedings of the 9th International Workshop on Spoken Language Translation, IWSLT'12*, Hong-Kong, December 2012.

[50] C. Mermer, M. Saraclar, and R. Sarikaya. Improving Statistical Machine Translation Using Bayesian Word Alignment and Gibbs Sampling. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21(5):1090–1101, May 2013.

[51] Márton Miháltz and Gábor Pohl. Exploiting Parallel Corpora for Supervised Word-Sense Disambiguation in English-Hungarian Machine Translation, 2006.

[52] U. Muegge. An Excellent Application for Crummy Machine Translation: Automatic Translation of a Large Database. In Elisabeth Gräfe, editor, *Proceedings*

of the Annual Conference of the German Society of Technical Communicators, pages 18–21, Stuttgart: tekom, 2006.

[53] M. Nagao. A Framework of a Mechanical Translation Between Japanese and English by Analogy Principle. In *Proc. Of the International NATO Symposium on Artificial and Human Intelligence*, pages 173–180, New York, NY, USA, 1984. Elsevier North-Holland, Inc.

[54] F. J. Och. Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 160–167, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.

[55] F. J. Och and H. Ney. Improved Statistical Alignment Models. In *ACL00*, pages 440–447, Hongkong, China, October 2000.

[56] S. Oepen, E. Velldal, J. T. Lønning, P. Meurer, V. Rosén, and D. Flickinger. Towards hybrid quality-oriented machine translation. On linguistics and probabilities in MT. In *In Proceedings of the 11th Conference on Theoretical and Methodological Issues in Machine Translation*, pages 144–153, 2007.

[57] K. Oflazer. Two-level Description of Turkish Morphology. In *Proceedings of the Sixth Conference on European Chapter of the Association for Computational Linguistics*, EACL '93, pages 472–472, Stroudsburg, PA, USA, 1993. Association for Computational Linguistics.

[58] K. Oflazer and İ. Durgar El-Kahlout. Exploring different representational units in English-to-Turkish statistical machine translation. In *Proceedings of the Statistical Machine Translation Workshop at ACL 2007*, pages 25–32, 2007.

[59] K. Papineni, S. Roukos, T. Ward, and W. Zhu. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 311–318, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

[60] M. V. Reddy and M. Hanumanthappa. Interlingual Machine Translation. In In Chief Sahil Seth, editor, *Proc. of the Intl. Conf. on Advances in Computer Science and Electronics Engineering*. Universal Association of Computer and Electronics Engineers, 2012.

[61] R. Rosenfeld. Two decades of statistical language modeling: where do we go from here? *Proceedings of the IEEE*, 88(8):1270–1278, Aug 2000.

[62] H. Sak, T. Güngör, and M. Saraçlar. Morphological Disambiguation of Turkish Text with Perceptron Algorithm. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, volume 4394 of *Lecture Notes in Computer Science*, pages 107–118. Springer Berlin Heidelberg, 2007.

[63] F. Sanchez-Martinez, M. L. Forcada, and W. Andy. Hybrid Rule-Based-Example-Based MT: Feeding Apertium with Sub-sentential Translation Units. In *Proceedings of the 3rd Workshop on Example-Based Machine Translation*, pages 11–18, Dublin, 2009.

[64] S. Sato and M. Nagao. Toward Memory-based Translation. In *Proceedings of the 13th Conference on Computational Linguistics - Volume 3*, COLING '90, pages 247–252, Stroudsburg, PA, USA, 1990. Association for Computational Linguistics.

[65] H. Sawaf, M. Shihadah, and M. Yaghi. Hybrid machine translation, July 15 2010. US Patent App. 12/606,110.

[66] C. Schmidt, Vilar D., and H Ney. Using a Bilingual Context in Word-Based Statistical Machine Translation, 2008.

[67] M. Simard, C. Goutte, and P. Isabelle. Statistical phrase-based post-editing. In *In Proceedings of NAACL*, 2007.

[68] M. Tatlıcıoglu. Frequency-Driven Late Fusion-Based Word Decomposition Approach On The Phrase-Based Statistical Machine Translation Systems. Master's thesis, Middle East Technical University, Ankara, Turkey, 2013.

[69] J. Tiedemann. Parallel Data, Tools and Interfaces in OPUS. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA).

[70] F. M. Tyers and M. S. Alperen. South-east european times: A parallel corpus of the balkan languages. In *In Proceedings of the Workshop on Exploitation of multilingual resources and tools for Central and (South) Eastern European Languages, LREC'10*, 2010.

[71] B. Vauquois. A Survey of Formal Grammars and Algorithms for Recognition and Transformation in Machine Translation. In *Proceedings of the IFIP Congress-68*, pages 254–260, 1968.

[72] S. Virpioja, J. J. Väyrynen, M. Creutz, and M. Sadeniemi. Morphology-aware statistical machine translation based on morphs induced in an unsupervised manner. In *PROC. OF MT SUMMIT XI*, pages 491–498, 2007.

[73] H. Watanabe. A Similarity-Driven Transfer System. In *Proceedings of COLING-92*, pages 23–28, Nantes, 1992.

[74] H. Watanabe and K. Takeda. A Pattern-based Machine Translation System Extended by Example-based Processing. In *Proceedings of the 17th International Conference on Computational Linguistics - Volume 2*, COLING '98, pages 1369–1373, Stroudsburg, PA, USA, 1998. Association for Computational Linguistics.

[75] W. Weaver. Translation 1949. In *Machine Translation of Languages: fourteen essays.* MIT Press,Cambridge,MA, 1955.

[76] F. Xia and M. McCord. Improving a Statistical MT System with Automatically Learned Rewrite Patterns. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.

[77] P. Xu, J. Kang, M. Ringgaard, and F. Och. Using a Dependency Parser to Improve SMT for Subject-object-verb Languages. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, pages 245–253, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[78] K. Yamada and K. Knight. A Syntax-based Statistical Translation Model. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, ACL '01, pages 523–530, Stroudsburg, PA, USA, 2001. Association for Computational Linguistics.

[79] M. Yang and K. Kirchhoff. Phrase-Based Backoff Models for Machine Translation of Highly Inflected Languages. In *Proceedings of the 21st International Conference on Computational Linguistics*, pages 1017–1020, 2006.

[80] R. Yeniterzi and K. Oflazer. Syntax-to-morphology Mapping in Factored Phrase-based Statistical Machine Translation from English to Turkish. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 454–464, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[81] E. Yılmaz, I. D. El-Kahlout, B. Aydın, Z. S. Ozil, and C. Mermer. TÜBİTAK Turkish-English Submissions for IWSLT 2013. In *Proceedings of the 10^{th} International Workshop on Spoken Language Translation Heidelberg, IWSLT'13*, Heidelberg, Germany, December 2013.

APPENDIX A

SAMPLE TRANSLATIONS FROM SMT SYSTEMS

Table A.1: Sample translations from all built systems

Source:	The aim of this reform is to ensure the establishment of internal labour inspection mechanisms .
Reference:	Bu reformun amacı , işletme içi iş teftiş mekanizmaları oluşturmaktır .
Baseline:	Amacı bu reformun , kurulması . denetleme mekanizmaları iç iş
Baseline-Reordered:	İç bu reformun amacı işgücü denetleme mekanizmaları kurulması sağlamaktır .
Tagged-Baseline:	Bu reformun amacıyla , iç belirtilenlerden denetleme mekanizmaları kurulması .
Tagged-Reordered:	Bu reformun amacı , kurgulanmış belirtilenlerden denetleme mekanizmaları kurulması sağlanmasıdır .
Factored-Baseline:	Bu reform sağlamak üzere , iç denetim mekanizmalarının oluşturulması iş .
Factored-Reordered:	Bu reformun amacı , iş iç denetim mekanizmalarının kavuşturulması sağlanmasıdır .

Table A.1 Continued: Sample translations from all built systems

Source:	He was taken out of the room for interrogation on several occasions , when he received electric shocks and was beaten .
Reference:	Birçok kez sorgulama için odadan çıkarılmış , elektrik şoku verilmiş ve dövülmüştür .
Baseline:	Öldüğü sırada veya tutukevinden odadan sorgulama amacıyla birçok vesilelerle , , Konseyi'nden elektrik şoku ve zaman dövülmüş ve .
Baseline-Reordered:	Başvurucu gelemeyişini elektrik şoku ve bir zaman dövülmüş ve , birkaç kez Tutuklunun sorgulama amacıyla odası grubunda ilişkin alınmıştır .
Tagged-Baseline:	Birlikte tutuldukları odadan alınmış olan bu ifade alma odası için , çeşitli vesilelerle ve elektrik şoku verildiğini söylediği yüzü alınmıştır .
Tagged-Reordered:	Kendisinin ve elektrik şoku verildiğini söylediği dövülmüş , çeşitli vesilelerle yapılanmadaki ifade alma odası ile ilgili tedbirler alınmıştır .
Factored-Baseline:	Başvurucu , bu ifade alma odası için , çeşitli vesilelerle , o zaman dövülmüş ve elektrik şoku edilmiştir .
Factored-Reordered:	Kendisine elektrik şoku ve dövülmüş halde, birçok kez sorgulama amacıyla alınmıştır .
Source:	Religious textbooks have been redrafted in order to address the concerns of Christian minorities .
Reference:	Hıristiyan azınlıkların kaygılarının giderilmesi için din dersi kitapları yeniden yazılmıştır .
Baseline:	Din kitapları , redrafted edilmiştir ele alınması amacıyla , Hıristiyan azınlıkların .
Baseline-Reordered:	Din kitapları , Hıristiyan azınlıkların endişelerine yanıt vermek amacıyla yeniden kaleme alınmıştır .
Tagged-Baseline:	Din kitapları endişelerine yanıt vermek amacıyla yeniden kaleme alınmıştır , Hıristiyan azınlıkların .
Tagged-Reordered:	Din kitapları , Hıristiyan azınlıkların endişelerine yanıt vermek amacıyla yeniden kaleme alınmıştır .
Factored-Baseline:	Din kitapları , yeniden amacıyla Hıristiyan azınlıkların endişelerine yanıt .
Factored-Reordered:	Din kitapları , Hıristiyan azınlıkların endişelerine yanıt vermek amacıyla yeniden kaleme alınmıştır .

Table A.1 Continued: Sample translations from all built systems

Source:	A number of transitional arrangements under the Customs Union expired in December 2000 .
Reference:	Gümrük birliği kapsamındaki bir çok geçici düzenlemenin süresi Aralık 2000'de dolmuştur .
Baseline:	Bir dizi Gümrük Birliği kapsamında yer alan Aralık 2000'de sona erdiğinden .
Baseline-Reordered:	Gümrük Birliği kapsamındaki bir dizi geçici düzenlemenin süresi Aralık 2000'de sona ermiştir .
Tagged-Baseline:	Bir dizi Gümrük Birliği , Aralık 2000'de sonra ermiştir .
Tagged-Reordered:	Gümrük Birliği kapsamında bir dizi geçici düzenlemenin süresi Aralık 2000'de sona ermiştir .
Factored-Baseline:	Bir dizi geçici düzenlemenin süresi Gümrük Birliği çerçevesinde , Aralık 2000'de sona ermiştir .
Factored-Reordered:	Gümrük Birliği kapsamında bir dizi geçici düzenlemenin süresi Aralık 2000'de sona ermiştir .
Source:	The judiciary plays an important role in the implementation of political reforms .
Reference:	Yargı , siyasi reformların uygulanmasında önemli bir rol oynamaktadır .
Baseline:	Yargı önemli bir rol oynar siyasi reformların uygulanmasında .
Baseline-Reordered:	Yargı siyasi reformların uygulanmasında önemli bir rol oynamaktadır .
Tagged-Baseline:	Yargının önemli bir rol oynar siyasi reformların uygulanmasında .
Tagged-Reordered:	Yargı siyasi reformların uygulanmasında önemli bir rol oynamaktadır .
Factored-Baseline:	Yargının önemli bir rol oynayan siyasi reformların uygulanmasında .
Factored-Reordered:	Yargı siyasi reformların uygulanmasında önemli bir rol oynamaktadır .