A Hybrid Translation System from Turkish Spoken Language to Turkish Sign Language

Dilek Kayahan
Computer Engineering
Boğaziçi University
İstanbul, Turkey
dilek_kayahan@hotmail.com

Tunga Güngör
Computer Engineering
Boğaziçi University
İstanbul, Turkey
gungort@boun.edu.tr

***Abstract*—Sign language is the primary tool of communication for deaf and mute people. It employs hand gestures, facial expressions, and body movements to state a word or a phrase. Like spoken languages, sign languages also vary among the regions and the cultures. The aim of this study is to implement a machine translation system to convert Turkish spoken language into Turkish Sign Language (TİD). The advantages of rule-based and statistical machine translation techniques are combined into a hybrid translation system.**

***Index Terms*—sign language, TİD, hybrid translation, rule-based translation, statistical translation**

I. Introduction

Sign language is a type of natural language that is emerged to communicate visually. Contrary to the popular opinion, sign languages are not derived from spoken languages. Each country or region has its own sign language and embodies different grammatical rules and lexicons.

In this work, a hybrid translation system to translate Turkish spoken language into Turkish Sign Language (TİD) is proposed. The system comprises of rule-based and statistical translation components. Turkish text is first fed into rule-based translation component which applies predefined Turkish-to-TİD grammatical rules. Then intermediate translation results are processed by the statistical translation component and the final TİD translation is generated. Gloss representation is used to typify the TİD.

The main obstacle of the proposed translation system is the lack of information about TİD since it is still under development. There is also no written form of the sign language which makes it more difficult to analyze. In order to create a Turkish-to-TİD bilingual dataset, the online dictionary which is published by The Ministry of Family and Social Policies was parsed, and 3561 sentence pairs were extracted.

II. Related Work

Sign languages have four main components and additional non-manual markers to articulate a sign [1]. The main components are hand-shape, orientation, location, and movement. Hand-shape is the form of the hand, while orientation is the direction of the palm. Location is the signing position referenced to the body, such as chest or shoulders and movement is the action of the hand-shapes such as circling or touching. Non-manual markers are extra expressions such as eye gaze, head tilting and shoulder raising that are used to support the hand sign. In order to sign the words which have special meaning in the spoken language but lack a sign in the sign language, finger spelling is used. It simply expresses the word by signing the letters of the word individually. Each sign language has its own manual alphabet.

In order to typify sign languages, several notation systems were introduced such as Stokoe Notation, HamNoSys, SignWriting and Gloss representation.

HamNoSys [2] is a common notation system for all sign languages. It contains approximately 210 symbols. By the combination of these symbols it is possible to model any visual sign. It divides a sign into four main parts; hand-shape, hand position, location, and movement. Each part in the HamNoSys notation represents the relevant part of the visual sign. For example, hand is positioned according to the "hand position" part in the HamNoSys notation. Gesture realization tools interpret HamNoSys notation and visualize the correspondent gesture with avatars.

Gloss representation does not involve any hint about the gesture of the signs. Simply, they work as labels for the signs and they are the capitalized forms of the correspondent word translation of a sign.

Hernandez *et al.* [3] propose a Spanish speech to Spanish Sign Language (LSE) translation system for assisting deaf people with identity card applying or renewal process. The system converts officer's speech into sign language in real time. It has three components; speech recognizer, natural language translator, and 3D avatar animation. The speech recognizer component translates the spoken utterance into word sequences. Then, the natural language translator converts these sequences into LSE glosses by implementing rule-based and statistical methods separately. Finally, the resulting LSE sequences are matched with the predefined HamNoSys notations of the signs and fed into eSIGN editor for avatar animation. The rule-based translator comprises 153 translation rules and achieved 0.578 BLEU score while the statistical translator scores 0.4941. The statistical translator is trained with 266 sentence pairs and tested with 150 sentences. It is important to note that the system was designed for a particular domain and the dataset contains only sentence pairs from this domain.

Manzano [4] introduces a neural machine translation (NMT) system to translate English text into American Sign Language (ASL). The proposed system is used as a natural language translation component of the Speech2signs project. This project interprets input video and extracts the speech, then converts the speech into text. Then, it translates the English text into ASL and realizes the ASL signs by virtual avatar. ASLG-PC12 [5] dataset is used as parallel corpus. The train dataset contains 83618 sentence pairs, the development dataset has 2045 and the test dataset has 2046 sentences. The BLUE score of the system is denoted as 17.73.

Stoll *et al.* [6] implement a system that converts spoken language into sign language video. Unlike the aforementioned studies, it does not rely on the virtual avatars, instead implements its own sign video generation component with generative adversarial networks. The natural language translation component translates text into glosses. It is trained with a German dataset and it is evaluated in terms of the cumulative BLEU scores. The PHOENIX14T dataset containing 8257 German to German Sign Language (DGS) sentences is used to train the component. This component achieves 50.67 BLUE-1, 32.25 BLUE-2, 21.54 BLUE-3, and 15.26 BLUE-4 scores.

III. Dataset

Sign languages use visual expressions and they do not have any written form. That makes it challenging to generate a large dataset. The ambiguity of available sign language data and the lack of strict grammatical rules also make it harder.

The Turkey Ministry of Family and Social Policies built an online Turkish to TİD dictionary [7] containing video and gloss representations of the TİD signs. It also introduces Turkish to TİD sample sentences with relevant glosses. In this study, gloss representation is used to typify the signs in the dataset and the official online TİD dictionary is used to acquire reliable, Turkish to TİD translations.

In this study, we need a sentence-aligned, bilingual corpus for statistical translation. To do so, sample sentences for each word translation are used to compose the Turkish to TİD parallel corpus. The online TİD dictionary comprises 2000 words which are grouped alphabetically and it would be challenging to extract the sample sentences by hand. In order to automate the sample sentence extraction task, a website crawler is implemented. 3561 sentence pairs are retrieved and saved as the bilingual parallel corpus.

The generated corpus is then randomly split into the test, train, and development corpora for different components of the system. Approximately 80% of the corpus is reserved as train corpus while the remaining 20% is divided between test and development corpora. Among the 3561 sentence pairs, 2851 randomly selected ones are added to the train corpus, 363 are assigned to the test corpus, and 346 to the development corpus.

IV. Methodology

The Turkish to Turkish Sign Language hybrid translation system combines the advantages of the rule-based and statistical machine translation techniques. It consists of three components: rule-based translation component, preprocessor, and statistical translation component. These components are implemented with python programming language and corpus is stored in plain text format.

Turkish sentence is first processed by the rule-based translation component and the intermediate sign language translation of the input sentence is generated. Then the preprocessor fine-tunes the intermediate results for the statistical translation component. The statistical translation component applies the phrase-based statistical translation model using the Moses Decoder [8]. Fig. 1 illustrates an overview of the proposed system.

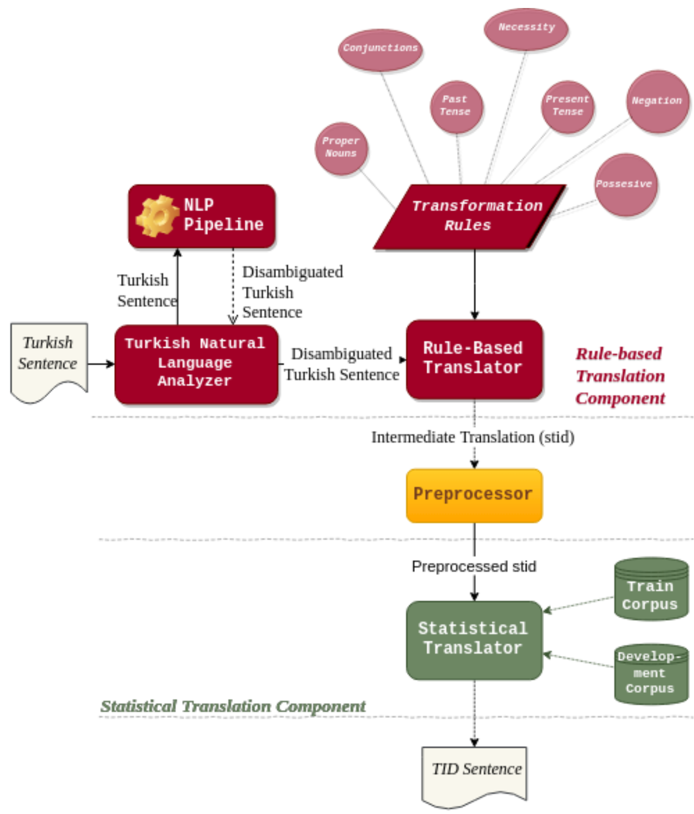

Fig. 1. Hybrid Translation System From Turkish Spoken Language to Turkish Sign Language Architecture

A. Rule-Based Translation Component

This component first analyzes the Turkish input sentence morphologically by the Boun Morphological Analyzer then applies the predefined Turkish to TİD translation rules. In this study, 13 Turkish to TİD translation rules are defined and explained in detail below accompanied with example sentences.

1) Infinitive Verb Inflection: Turkish Sign Language does not embody any suffixes. Instead, verbs are represented in infinitive forms while nouns are in nominative forms. TİD fills this gap by employing non-manual markers such as head tilt, eye gaze, and mouthings to convey the additional meanings or implications. This rule omits the suffixes of each word in the

Turkish sentence and translates stems of the Turkish words into the correspondent TİD glosses. Stems other than the verbs are translated as they are, while verb stems are inflected for their infinitive forms (-mek or -mak form). The infinitive inflection rule is simply performed by inspecting the last vowel in the verb stem. If the last vowel in the verb stem is a front vowel it is conjugated with “-mek", otherwise “-mak" suffix is applied. On the other hand, passive and causative verbs are exceptions for this rule since they derive new words from the stems. In order to eliminate this problem, passive and causative verb stems are regenerated by appending the derivative suffixes to the root stem.

Turkish Sentence:

Piknik	için	plan	yapmıştık	.
|	|	|	|	|
Picnic	for	plan	have-done	.

(We had a plan for picnic.)

Disambiguator result:

```
Piknik      piknik[Noun]+[A3sg]+[Pnon]+[Nom]
için        için[Postp]+[PCNom]
plan        plan[Noun]+[A3sg]+[Pnon]+[Nom]
yapmıştık   yap[Verb]+[Pos]+mHş[Narr]+YDH[Past]+
            +k[A1pl]
,           ,[Punc]
```

TİD Sentence: PİKNİK İÇİN PLAN YAPMAK

2) Punctuation Marks: Punctuation Marks in Turkish input sentence are eliminated since they are not used in TİD.

3) Conjunctions: If “-de" connector follows a verb in the Turkish input sentence, the verb is reduplicated in TİD.

Turkish Sentence:

Ben	de	sustum	.
|	|	|	|
I	also	quieted-down	.

(I also quieted down.)

Disambiguator result:

```
ben      ben[Pron]+[Pers]+[A1sg]+[Pnon]+[Nom]
de       de[Conj]
sustum   sus[Verb]+[Pos]+DH[Past]+m[A1sg]
```

TİD Sentence: BEN SUSMAK SUSMAK

“-ki" connector (relative pronoun) in Turkish input sentence is omitted since it is nonfunctional in TİD.

Other conjunctions like “ve" (“and"), “ama" (“but") and “ile" (“with") are translated from Turkish into TİD as they are.

4) Person Agreement: This rule is only applied to the verbs in the sentence to extract person information. If a verb has person agreement, the corresponding personal pronoun is added to the beginning of the TİD sentence.

Turkish Sentence:

Hemen	hastaneye	gittik	.
|	|	|	|
Immediately	to-hospital	we-went	.

(We went to hospital immediately.)

Disambiguator result:

```
hemen       hemen[Adv]
hastaneye   hastane[Noun]+[A3sg]+[Pnon]+YA[Dat]
gittik      git[Verb]+[Pos]+DH[Past]+k[A1pl]
```

TİD Sentence: BİZ HEMEN HASTANE GİTMEK

5) Present Tense Rule: This rule is defined to convey the time information. If any verb in the Turkish input sentence has only progressive feature as the time indicator and also has the first single person agreement, “ŞİMDİ" (“NOW") gloss is added to the head of the TİD sentence as the time adverb.

Turkish Sentence:

Çok	üzülüyorum	.
|	|	|
Very	I-am-sorry	.

(I'am very sorry.)

Disambiguator result:

```
Çok          çok[Adv]
üzülüyorum   üz[Verb]-Hl[Verb+Pass]+[Pos]
             +Hyor[Prog1]+YHm[A1sg]
.            .[Punc]
```

TİD Sentence: BEN ŞİMDİ ÇOK ÜZÜLMEK

6) Past Tense Rule: This rule is defined to convey the time information. If a verb in the Turkish sentence has past tense inflection along with progressive feature, “BİTTİ" (like “END") gloss is added to the end of the TİD sentence as the time adverb.

Turkish Sentence:

Eve	gidiyordum	.
|	|	|
To-home	I-was-going	.

(I was going to home.)

Disambiguator result:

```
eve          ev[Noun]+[A3sg]+[Pnon]+YA[Dat]
gidiyordum   git[Verb]+[Pos]+Hyor[Prog1]+
             +YDH[Past]+m[A1sg]
.            .[Punc]
```

TİD Sentence: BEN EV GİTMEK BİTTİ

7) Future Tense Rule: Turkish Sign Language does not employ future tense.

8) Necessity Rule: Necessitative which is relayed with "-meli", "-malı" suffixes in Turkish language, is transferred to TİD by "LAZIM" (like "REQUIRED") gloss. It is concatenated to the infinite form of the word stem.

Turkish Sentence:

Cam	su	şişelerinden	almalısınız	.
|	|	|	|	|
Glass	water	bottles	should-buy	.

(You should buy glass water bottles.)

Disambiguator result:

```
cam            cam[Noun]+[A3sg]+[Pnon]+[Nom]
su             su[Noun]+[A3sg]+[Pnon]+[Nom]
şişelerinden   şişe[Noun]+lAr[A3pl]+
               +SH[P3sg]+NDAn[Abl]
almalısınız    al[Verb]+[Pos]+mAlH[Neces]+
               +sHnHz[A2pl]
.              .[Punc]
```

TİD Sentence: SİZ PLASTİK ŞİŞE SAĞLIK ZARAR CAM SU ŞİŞE ALMAK LAZIM

9) Negation Rule: Privative affixes "-ma", "-me" and "-madan", "-meden" convey negation meaning in Turkish, while "DEĞİL" gloss is used in TİD. If a verb has privative affix in the Turkish input sentence, "DEĞİL" ("NOT") gloss is attached to the infinitive form of the word stem.

Turkish Sentence:

Müdür	beğenmedi	.
|	|	|
Manager	he/she-didn't-like	.

(Manager didn't like.)

Disambiguator result:

```
müdür       müdür[Noun]+[A3sg]+[Pnon]+[Nom]
beğenmedi   beğen[Verb]+mA[Neg]+DH[Past]+[A3sg]
.           .[Punc]
```

TİD Sentence: MÜDÜR BEĞENMEKˆDEĞİL

10) Possessive Rule: The possessive suffix in Turkish is translated into possessive pronoun in TİD and it is prepended to the relevant word stem.

Turkish Sentence:

Arabam	var	.
|	|	|
My-car	have	.

(I have a car.)

Disambiguator result:

```
arabam   araba[Noun]+[A3sg]+Hm[P1sg]+[Nom]
var      var[Adj]
.        .[Punc]
```

TİD Sentence: BENİM ARABA VAR

11) Locative Rule: The locative meaning in Turkish is transferred to TİD by utilizing "İÇİNDE" gloss. If a noun is inflected with locative suffix and followed by a verb in Turkish sentence, it is translated to TİD by appending "İÇİNDE" ("INSIDE") gloss to its stem.

Turkish Sentence:

Doğum günü partimi	evde	yapmayı	düşünüyordum
|	|	|	|
Birthday my-party	at-home	to-make	I-was-thinking

(I was thinking to make my birthday party at home.)

Disambiguator result:

```
Doğum          doğum[Noun]+[A3sg]+[Pnon]+[Nom]
günü           gün[Noun]+[A3sg]+SH[P3sg]+[Nom]
partimi        parti[Noun]+[A3sg]+Hm[P1sg]+NH[Acc]
evde           ev[Noun]+[A3sg]+[Pnon]+DA[Loc]
yapmayı        yap[Verb]+[Pos]-mA[Noun+Inf2]+
               +[A3sg]+[Pnon]+YH[Acc]
düşünüyordum   düşün[Verb]+[Pos]+Hyor[Prog1]+
               +YDH[Past]+m[A1sg]
.              .[Punc]
```

TİD Sentence: BİZ DOĞUM GÜN BENİM PARTİ EV İÇİNDE YAPMAK DÜŞÜNMEK

12) Ablative Rule: The ablative suffixes in Turkish sentence are omitted since they are not used in TİD.

13) Proper Nouns: Fingerspelling is the representation of each letter of a word by hand movements in sign languages. If there is a proper noun in the Turkish sentence, "fingerspell" mark is appended to its translation in TİD.

Rule-Based Translator

The rule-based translator is a python based application that implements the aforementioned rules by utilizing the Boun Morphological Analyzer. It gets input sentences as a file and executes morphological parser, morphological disambiguator and translation rules consecutively. The resulting translations are then saved into the given output file.

The rule-based translator also fine-tunes the translation results by extra enhancements. It first trims the sentence then eliminates the rule collisions such as possessive and personal pronoun conflictions, as exemplified below.

Turkish Sentence:

Ailemden	ayrı	yaşıyorum	.
|	|	|	|
My-family	apart-from	I-live	.

(I live apart from my family.)

Applied transformation rules:

```
Possesive Rule:          ailemden -> BENİM AİLE
Person Agreement Rule:   yaşıyorum -> BEN YAŞAMAK
Translation:             BEN BENİM AİLE AYRI YAŞAMAK
```

The rule-based translator detects the collision in the above sentence and subtracts the redundant "BENİM" possessive pronoun. So it converts the final translation into "BEN AİLE AYRI YAŞAMAK" TİD sequence.

B. Preprocessor

The preprocessing stage is required to reduce data sparsity for the evaluation process and statistical machine translation components. In order to calculate consistent BLEU scores during the system evaluation, the translated output and the correspondent test sentence should be well aligned in terms of the punctuation, case sensitivity, and sentence length. These divergences mislead the training and tuning phases of the machine translation component. In order to overcome the language-specific concerns, custom Turkish and TİD preprocessors are implemented.

Custom Turkish Preprocessor

Custom Turkish preprocessor first eliminates the expressions in the parentheses, then converts all characters into lowercase with Turkish encoding. Then, it deletes "ki" and "de" conjunctions since they do not have individual representations in TİD. Lastly, it removes all punctuations, empty lines and trims the redundant whitespaces.

Custom TİD Preprocessor

Unlike Turkish, expressions in the parentheses deliver significant information in TİD rather than extra information. So these expressions are not omitted. Instead, they are treated as standard expressions. The custom TİD preprocessor first extracts the expressions in the parentheses, then removes the punctuations.

In TİD sentence, "ˆ" circumflex accent is used to sign negations such as BEĞENMEKˆDEĞİL and multi-word expressions such as "GİTMEKˆGELMEK". If it is used to convey negation, the preprocessor deletes it and concatenates the negation marker "DEĞİL" to the former word. On the other hand, if it is used to express multi-words, preprocessor splits these words by replacing the circumflex accent with whitespace.

Finally, the preprocessor removes the fingerspell marker "ˆFS" and converts all characters into lowercase with Turkish encoding.

C. Statistical Translation Component

Statistical Translation Component implements statistical machine translation (SMT) techniques to translate the Turkish Spoken Language into the TİD. SMT approach is a state-of-the-art translation methodology which relies on the statistical models that are extracted from the parallel data.

This component takes the advantage of the Moses Decoder [8] to perform statistical machine translation. The Moses Decoder has two main components: a training pipeline which is a collection of tools for generating language models and a decoder to translate the input sentence. Language modeling and tuning are also significant parts of the translation system.

The tuning process improves the translation quality of the translation model which is generated by the training pipeline. A parallel corpus other than the training corpus is used to fine-tune the translation model's output by comparing the target sentence in the development corpus with the target sentence that is generated by the translation model for the same source sentence.

In this study, statistical translation component is trained with the outputs of the rule-based translation system.

V. Evaluation

There are two main approaches to measure the accuracy of machine translation systems; human evaluation and automated scoring metrics. These two natural language oriented approaches are also applicable to the sign languages.

The human evaluation method has bottlenecks such as subjectiveness, time consumption, and non-reproducibility, for the evaluation of the spoken language translations. In addition, it has a major drawback for the sign languages; most of the native signers have trouble to express and interpret sign languages in written forms. The reason is that they generally learn the sign languages visually from their family and they do not have a theoretical background about it. In the case of TİD, most of the grammatical rules are not well defined yet and it could be misleading to rely on the evaluation of non-signers. Due to the aforementioned obstacles, automated scoring method is used for the system evaluation rather than the human evaluation method.

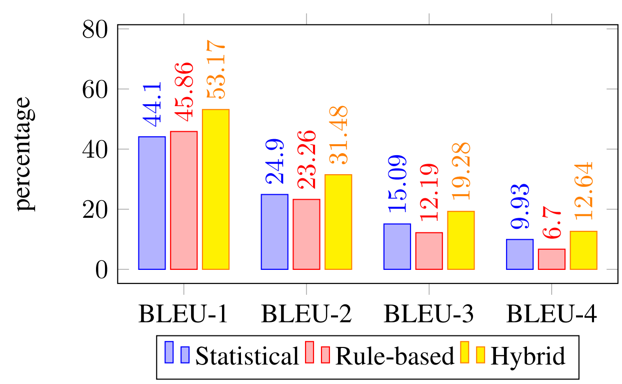

Fig. 2. Comparision of the hybrid translation system, statistical translation component and rule-based translation component.

The proposed system's performance is directly proportional to the performance of the translation components. For this reason, performance of the rule-based and statistical translation components are measured individually and compared to the hybrid translation system as shown in Fig.2.

VI. Comparision and Discussion

The proposed Hybrid Translation System is compared to the studies which are proposed by Hernandez *et al.* [3], Manzano [4] and Stoll *et al.* [6]. In order to facilitate the naming, they are called as System-1, System-2 and System-3 respectively. These studies are described in detail in Section 2. These systems are compared in terms of the BLEU scores as shown in Fig. 3. System-1 and System-2 only calculate the BLEU-4 scores for the evaluation. This is why BLEU-3, BLEU-2, and BLEU-1 scores are marked as 0.

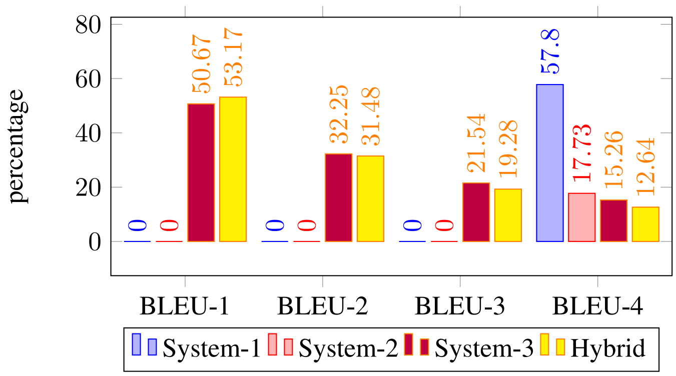

Fig. 3. Comparision of the hybrid translation system with the related studies.

System-1 achieves the best score among the others by 57.8%. This system employs 153 translation rules and limits its translation domain to utterances which are used in identity card office. It is obvious that applying rules to a specific domain will have high performance.

TABLE I
DATASET COMPARISON OF THE SYSTEMS

	System-2	*System-3*	*Hybrid*
Train	83618	Unknown	2851
Develop	2045	Unknown	346
Test	2046	Unknown	363
Overall	87709	8257	3561

System-2 and System-3 are neural machine translation (NMT) based systems, therefore, their performance depends on the dataset size. The Hybrid Translation System is also affected greatly by the dataset size. So the dataset sizes are compared in Table 1. Although having the smallest dataset among these systems, the Hybrid Translation System scores well.

The main contribution of the proposed system is the rule-based translation component. Compared to the related studies, it is obvious that language specific translation rules increase the overall system performance. In order to measure the effect of a rule, it is omitted from the system and the whole system has trained from the scratch. According to the BLEU scores, Negation rule decreases the overall performance by %0.66 while Present Tense rule decreases by %0.38. The difference between the effects of the rules does not give an insight about the importance of the rule. Instead, it indicates that the occurrence frequency of the Negation rule in the test data set is more than the Present Tense rule. In the same manner, bigger test data will increase the performance impact of the rules.

On the other hand Hybrid Translation System interprets the input sentence only morphologically. In order to increase translation accuracy, it should be analyzed semantically as well, by introducing new rules.

VII. CONCLUSION

This study introduces a hybrid translation system to convert Turkish text into Turkish Sign Language. Rule-based and statistical translation approaches are combined and we obtained satisfactory results.

The Turkish input sentence is first analyzed morphologically by The Boun Morphological Analyzer. According to the parser results, the rule-based translator applies the predefined Turkish to TİD transformation rules. Each rule first interprets the Turkish input sentence in various aspects such as tense, person agreement, possessiveness, and conjunctions, then defines the appropriate TİD translation. The rule-based translation component comprises 13 rules. The output of the rule-based translation component is then fed into the statistical translation component in order to enhance the translation quality. The Moses Decoder is used to implement statistical machine translation.

Translation accuracy is evaluated by the cumulative BLEU scoring metric. The proposed hybrid translation system has achieved %12.64 BLEU-4, %19.28 BLEU-3, %31.48 BLEU-2 and %53.17 BLEU-1 scores. Rule-based and statistical translation components of the system are also evaluated individually. Evaluation results demonstrate that the combination of the rule-based and statistical machine translation techniques increases the overall system performance.

As future work, translation output should be fed into a virtual avatar tool to realize the gestures of the sign language.

REFERENCES

[1] Baker, A., B. van den Bogaerde, R. Pfau and T. Schermer, “The Linguistics of Sign Languages: An introduction", John Benjamins Publishing Company, 2016

[2] Hanke, T., “HamNoSys—Representing sign language data in language resources and language processing contexts", *LREC 2004, Workshop Proceedings: Representation and Processing of Sign Languages. Paris: ELRA*, pp. 1–6, 2004.

[3] Hernandez, R., R. Barra-Chicote, R. Cordoba, L. D'Haro, F. Fernández-Martínez, J. Ferreiros, J. Lucas, J. Macias-Guarasa, J. Montero and J. Pardo, “Speech to sign language translation system for Spanish", *Speech Communication*, Vol. 50, pp. 1009–1020, 2008.

[4] Manzano, D., “English to Asl Translator for Speech2signs", 2018.

[5] Othman, A. and M. Jemni, “English-ASL Gloss Parallel Corpus 2012: ASLG-PC12", 2012.

[6] Stoll, S., N. C. Camgöz, S. Hadfield and R. Bowden, “Sign Language Production using Neural Machine Translation and Generative Adversarial Networks", *BMVC*, 2018.

[7] Makaroğlu, B. and H. Dikyuva, *Güncel Türk İşaret Dili Sözlüğü [The Contemporary Turkish Sign Language Dictionary]*, 01 2017.

[8] Koehn, P., *MOSES Statistical Machine Translation System: User Manual and Code Guide*, 2018.

[9] Okan Kubus, A. H., The phonetics and phonology of TİD (Turkish Sign Language) bimanual alphabet, *Formational units in sign languages*, 2017.

[10] Zeshan, U., Aspects of Türk İşaret Dili (Turkish Sign Language), *Sign Language & Linguistics*, Vol. 6, 01 2003.

[11] Smith, R., *HamNoSys 4.0 User Guide edited by Robert Smith*, Ireland, 2013.

[12] Kaur, K. and P. Kumar, HamNoSys to SiGML Conversion System for Sign Language Automation, *Procedia Computer Science*, Vol. 89, pp. 794–803, 12 2016.