Statistical Machine Translation for Greek to Greek Sign Language Using Parallel Corpora Produced via Rule-Based Machine Translation

Dimitrios Kouremenos[1], Klimis Ntalianis[2], Giorgos Siolas[3] and Andreas Stafylopatis[4]

[1] School of Electrical & Computer Engineering
National Technical University of Athens
15780 Athens, Greece
dkourem@gmail.com
[2] Department of Business Administration
University of West Attica
Athens, Greece
kdal75@gmail.com
[3] School of Electrical & Computer Engineering
National Technical University of Athens
15780 Athens, Greece
gsiolas@islab.ntua.gr
[4] School of Electrical & Computer Engineering
National Technical University of Athens
15780 Athens, Greece
andreas@cs.ntua.gr

Abstract. One of the objectives of Assistive Technologies is to help people with disabilities communicate with others and provide means of access to information. As an aid to Deaf people, we present in this work a novel prototype Rule-Based Machine Translation (RBMT) system for the creation of large quality written Greek text to Greek Sign Language (GSL) glossed corpora. In particular, the proposed RBMT system supports the professional translator of GSL to produce high quality parallel Greek text - GSL glossed corpus, which is then used as training data by the Statistical Machine Translation (SMT) MOSES [1] application system. It should be noted that the whole process is robust and flexible, since it does not demand deep grammar knowledge of GSL. With this work we manage to overcome the two biggest obstacles in Natural Processing Language (NLP) of GSL. Firstly, the lack of written system and secondly the lack of grammar and finally we have been able to lay the foundations for an autonomous translation system of Greek text to GSL. Evaluation of the proposed scheme is carried out in the weather reports domain, where 20,284 tokens and 1,000 sentences have been produced. By using the BiLingual Evaluation Understudy (BLEU) metric score, our prototyped MT system achieves a relative average score of 60.53% and 85.1%/65.5%/53.8%/44.8% for for 1-gram/2-gram/3-gram/4-gram evaluation.

Keywords: machine translation, Greek, Greek Sign Language, GSL, Deaf people communication, SMT, Moses, Phrase model.

1 Introduction

Translation helps people to communicate across linguistic and cultural barriers. However, according to Isabelle and Foster [2], translation is too expensive, and its cost is unlikely to fall substantially enough, to constitute it as a practical solution to the everyday needs of ordinary people. Machine translation can help break linguistic barriers and make translation affordable to many people. This situation is especially important for Deaf people, since translation supports the communication between Deaf and hearing communities and provides Deaf people with the same opportunities to access information as everyone else [3].

1.1 Sign Languages – The Greek Sign Language

Sign languages (SLs) exploit a different physical medium from the oral-aural system of spoken languages. SLs are gestural-visual languages, and this difference in modality causes SLs to constitute another branch within the typology of languages. However, there are still many myths around SLs. One of the most common and enduring myths is that the SL is universal; however, in reality, each country generally has its own, native sign language [4, 5].

This paper focuses on the Greek Sign Language (GSL), which is a complete language using the same grammar mechanisms incorporated by the oral language[1]. According to the Greek law no. 2817/2000[2], GSL is the official language of the Greek Deaf community[3], while in 2013 the Greek Deaf Federation has published a formal announcement demanding the institutional recognition of GSL[4]. Currently more than 40,000[5] people use GSL. Additionally, another common myth is that there is a correlation between the Greek spoken language and GSL. However SLs do not derive from spoken languages, but, as natural languages, they are influenced by their contact to other languages, allowing the development of dialects and varieties [6].

1.2 Problems of SLs

According to Porta et. al. [6] regarding the fundamental problems of SLs, most contemporary works on SLs have adopted language theories created for the spoken language instead of developing new theories. From the point of view of natural language processing, SLs are still under-resourced or low-density languages – that is to say, little or no specific technology is available for these languages, and computerized linguistic resources, such as corpora or lexicons, are very scarce.

Additionally, another major problem of SLs is the lack of a writing system. Strictly speaking, the only way to represent SLs is by using video and this is why there is lack

---

1 https://goo.gl/pAemOJ, https://en.wikipedia.org/wiki/Greek_Sign_Language
2 https://goo.gl/oItdK0
3 https://goo.gl/GGPIUo
4 http://www.omke.gr/anakoinwseis/diakirixi-syntagmatiki-anagnwrish-eng/
5 https://goo.gl/OZPAX5

of large corpora. The limitations in composing, editing and reusing SL utterances as well as their consequences for Deaf education and communication have been systematically mentioned in the SL studies literature since the second half of the twentieth century [7]. However, several notational systems exist. The most important include Stokoe [8], SignWriting [9], HamNoSys [10] and Neidle [11]. SignWriting was conceived primarily as a writing system, and has its roots in DanceWriting [12], a notation for reading and writing dance movements. HamNoSys was conceived as a phonological transcription system for SLs, with the same objective as the International Phonetic Alphabet (IPA) for spoken languages. A very promising system is SiGML [13], which represents the 3-D properties of SLs. Last but not least, the “si5s” writing system [14] has been proposed for the American Sign Language (ASL).

Furthermore, regarding GSL and to the best of the authors’ knowledge, currently no Language Model exists. To confront the aforementioned problems, in this paper an innovative RBMT system is proposed, which quickly produces high quality large glossed GSL corpus. In particular, the focus is primarily on syntax, so glosses are used instead of phonological notation. Glossing is a commonly used system for explaining or representing the meaning of signs and the grammatical structure of signed phrases and sentences in a text, written in another language. However, glossing is not a writing system that could be understood by SL users. For this reason, a novel gloss system is proposed based on the Berkley system (for the ASL), which is also decorated with Non Manual Component Sign (NmCs) tag features. The proposed scheme also enables the production of a simpler version of gloss without NmCs tags, adopted from the Deaf Community and especially from the bilingual deaf people who use a similar written Greek system in the Social Media.

- To sum up the main innovations of the proposed scheme include:
- The implemented GSL MT System is based on open source Toolkits.
- The overall scheme, with the help of a professional translator, can produce different kinds of large quality GSL Glossed Corpus that can be used for several purposes.
- The performance of the proposed GSL scheme is evaluated by the BLEU metric score [15].

The rest of this paper is organized as follows: in Section 2 we present a sketch of GSL and presets a review of Rule-based SL MT Systems. In Section 3 the related work is analyzed and we describe how our prototyped RBMT system produces a parallel Greek text with GSL glossed corpus and finally train the SMT Moses system. In section 4 we evaluate the proposed SMT MOSES system. Finally, in section 5 concludes this paper, providing also some directions for future work.

2 Literature review of SL MT systems

2.1 Background

Machine Translation (MT) of spoken languages has its roots in the 1940s, with a significant expansion of interest in the late 70s and 80s [16]. A similar level of development cannot be said for SL MT. Widespread research in this area did not emerge until the 1990s, where linguistic analysis of SLs has appeared [17]. Despite this late venture, the development of SL MT systems has roughly followed that of spoken language MT from 'second generation' rule-based approaches towards data-driven approaches. The 'second generation' or rule-based approaches to MT, emerged in the 1970s/1980s with the development of systems such as Meteo [18, 19] and Systran [20]. These systems are examples of the first commercially adopted MT systems to successfully translate spoken languages.

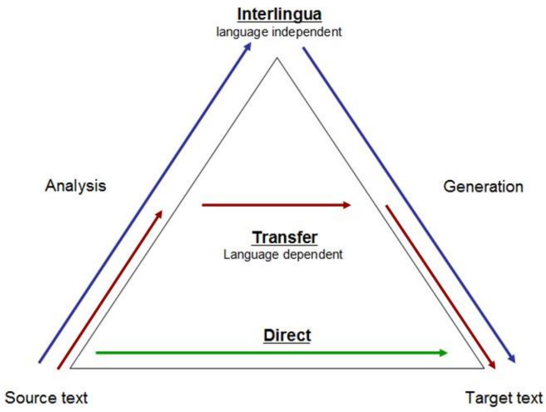

Fig. 1. The Vauquois Pyramid

Rule-based approaches may be sub-classified into transfer– and interlingua–based methodologies. The Vauquois Pyramid, shown in Fig. 1. [21], is widely used in MT circles to demonstrate the relative effort involved in translation processes. Transfer approaches, being language-dependent, need to know the source and target languages. Interlingua approaches tend to enact a deeper analysis of the source language sentence that creates structures of a more semantic nature. Both methods have their advantages and disadvantages.

2.2 Sketch of GSL

The most important documentation for a language is a reference grammar, which documents the principles governing the construction of words and all kinds of grammatical structures found in a language. Currently and regarding GSL, there are some attempts to gather resources, create a dictionary and annotated corpora and analyze a set of signers' data deriving from the annotated corpora [22, 23]. Additionally another interesting initiative to develop the blueprint for SL grammars is carried out by the SignGram COST Action[6].

2.3 Rule-based SL MT Systems

All MT systems for SLs published up to 2003 were just works in progress or simple demonstrators [24]. However, some systems were particularly distinguished, including the ZARDOZ system [25], the ViSiCAST Translator [26], the ASL Workbench [27], the SL translation via DRT and HPSG Safar et al. [28] and the TEAM project Zhao et al. [29]. All these systems were rule-based and made use of transfer-based or interlingua-based approaches. The only approach dealing with classifier predicates was that of Huenerfauth [24], who proposed a multi-path approach combining interlingua, transfer and direct approaches as a whole.

For Spanish to Spanish Sign Language (LSE), Baldassarri and Royo-Santas [30] described a rule-based demonstrator. Spanish is analyzed using FreeLing dependency analysis [31]. The dependency analysis through grammatical rules is transformed into a series of glosses. The system was tested with 92 sentences containing a total of 561 words. Appropriate dictionary entries were created for the evaluation, with very satisfactory results: 96% of the words were correctly translated, and 93.7% of them were in correct order. Another interesting Spanish SL MT system is the rule-based Spanish-to-LSE MT system based on Apertium, a free/open-source platform [32]. There are no published results on this system but it is available online .

Now regarding GSL, Kouremenos et. Al [33], presented a prototype Greek text to GSL conversion system. In that work, the detailed implementation of the language-processing component is provided, focusing upon the inherent problems of knowledge elicitation of sign language (SL) grammar and its implementation within a parser framework. Recently Efthimiou et. al. [7] presented the implementation of a post-processing stage to a grammar-based machine translation (MT) system from written Greek to GSL.

2.4 Data-Driven Based SL MT Systems

Lately, example-based machine translation (EBMT), statistical machine translation (SMT) and other types of data-driven machine translation systems have replaced the earlier RBMT approaches. However, data-driven approaches estimate their parameters from an aligned bilingual corpus, and their accuracy depends heavily on the quality and size of this corpus. Unfortunately, corpora for SLs are still very far from reaching the state-of-art of those for spoken languages. Additionally, the problem of modality and the lack of a standardized writing system make data acquisition for SLs a time-consuming and expensive task. Despite the lack of parallel corpora, the success

---

6 SignGram COSTS Action IS-1006 ''A blueprint for sign language grammars—unravelling the grammars of European sign languages: pathways to full citizenship of deaf signers and to the protection of their linguistic heritage'' (www.signgram.eu).

of data-driven approaches to MT between spoken languages, has led to the application of the same techniques to SLs. However, according to Morrissey [34] , most research in SL MT has emanated from sporadic and short-term projects as opposed to long term research investment. Some works are still worth mentioning: the Thai-to-Thai SL machine translation system [35] presents a direct translation system with reordering rules. The system for Thai reaches an F-score of about 97% for a set of 297 test sentences. Bauer et. al. [36] presented the first statistical approach to SL MT for German. In their paper they report that for 52 signs they achieve a recognition accuracy of 94% and a score of 91.6% for 100 signs. Morrissey [17] presented exhaustive experiments on the MaTrEx, a hybrid approach combining EBMT and SMT [37]. Results of MaTrEx on the ATIS corpus reached 0.39 BLEU for English-to-Irish Sign Language translation, and about 50% for German to German Sign Language (DGS) translation. Recently, Morrissey and Way [38] exploited the bidirectionality of the MaTrEx system, demonstrating how additional modules, such as recognition and SL animation, can potentially build a full SL MT model for spoken and SL communication.

2.5 Overall Discussion and Focus of the Proposed Scheme

This paper attempts to solve a very serious problem of the GSL, the lack of large GSL corpora. Towards this direction, a processing methodology is proposed for creating large quality parallel data for SLs by a human professional translator. The translator uses a simple rule-based system based on Python, open source tools which incorporate a transfer module in case of interlingua approaches and a robust grammar tree transfer parser. Next we feed the parallel corpus for training the Moses system, an Open source toolkit for statistical machine translation [1].

All aforementioned components (except the open source tools) have been fully developed and extensively tested by the authors.

3 The Proposed MT System for Greek-to-GSL Translation

The proposed MT system has taken into consideration the Basic Unification Grammar principles [1, 10, 39, 40]. For its overall development, different tools and technologies have been combined for the prototype RBMT system : (a) AUEB’s POS Parser [41], (b) the NLTK (Natural Language Toolkit) 3.0 suite , which is a free, open source, community-driven, leading platform for building Python programs to work with human language data, (c) Java and (d) Perl scripts. And for the ST system we finally use Moses, an open-source toolkit for statistical machine translation [1].

Additionally, translation by RBMT system is supervised by a professional translator, so that output texts are corrected and new transfer rules and lexicon mapping data are added to the RBMT, so that any newly appearing cases (linguistic phenomena) are covered.

3.1 Overall Architecture

The whole procedure of our system is divided into two main stages, (Fig. 1). Firstly, we use our RMBT system to produce parallel corpora of Greek text and GSL gloss text. At RBMT system we perform analysis actions separating by POS parsing is carried out by AUEB's Greek POS Parser [41] and chunk partial parsing (Fig. 3). Table 1 provides a list of the fine most frequently appearing morphological tags of the Parole standard. Chunk Partial Parser use the chunk parser and regular grammar from Python's NLTK Toolkit, Partial Chunking is accomplished and sentences are divided into sub-sentences as constituency tree structure. Chunk Partial Parser use the chunk parser and regular grammar from Python's NLTK Toolkit, Partial Chunking is accomplished and sentences are divided into sub-sentences as constituency tree structure. Next we have the transfer action separating by chunk transfer and word transfer. The Chunk transfer module incorporates a bilingual lexicon and specific knowledge from the language pair-specific rule database to transfer the Greek constituency tree structure into the corresponding GSL constituency tree structure. Then Gloss sequence and Gloss synthesis are performed to complement the structure, so that the final sentence is formed.

The transfer module incorporates a bilingual lexicon and specific knowledge from the language pair-specific rule database to transfer the Greek constituency tree structure into the corresponding GSL constituency tree structure. Word ordering and morphological rules are applied to the transferred constituency tree, so that the output of the generation stage of RBMT system is a sequence of written glosses with morphological and non-manual components' indications. The proposed written GSL glosses system uses the code style of BERKLEY Gloss System [42, 43] as a transcribing system, which abstracts away the phonological representation of signs (Fig. 4). Details of the different stages of the MT strategy are provided in the following subsections (Fig. 2)

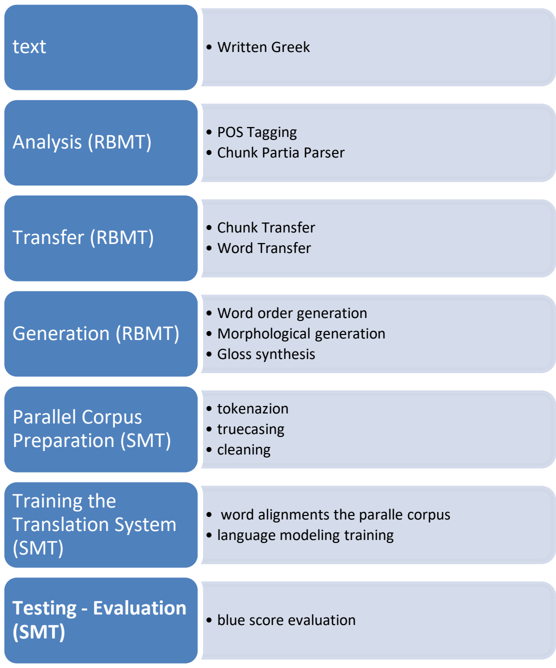

Fig. 2. Architecture of the system

```
S
 (NP Βροχές/NoCmFePlAc και/CjCo καταιγίδες/NoCmFePlAc)
 (VB θα/PtFu εκδηλωθούν/VbMnIdXx03PlXxPePvXx)
 (NP
   κατά/AsPpSp
   τόπους/NoCmMaPlAc
   στη/AsPpPaFeSgAc
   Δυτική/AjBaFeSgAc
   Ελλάδα/NoPrFeSgAc)
 (NP-CM Τα/AtDfNePlNm Χριστούγεννα/NoPrNePlAc)
)
```

Fig. 3. POS Parsed and Chunked Sentence

Table 1. TABLE I. Five most frequently appearing morphological tags of the parole standard.

POS	Example	Comment
No	Βροχές/tag=NoCmFePlAc (rain)	Ουσιαστικό/Noun (No), γένους θηλυκού/feminine (Fe) στον πληθυντικό/plural (Pl)
Aj	Δυτική/AjBaFeSgAc (Western)	Επίθετο/Adjective(Aj), γένους θηλυκού/feminine (Fe) στον ενικό/in singular (Sg)

As	στα/ AsPpPaNePlAc (at)	Adposition (= Preposition)[7]
At	τα/AtDfNePlNm (the)	Άρθρο/Article (At), γένος ουδέτερο/gender neutral (Ne) στον πληθυντικό/plural (Pl)
Vb	εκδηλωθούν/VbMnIdXx03PlXxPe PvXx (occurs)	Ρήμα/Verb (Vb), παθητικής φωνής/passive voice (Pv), πληθυντικός/plural (Pl)

The RBMT system, generates the sequence of GSL glosses decorated with non-manual component tags, using code types of the BERKLEY Gloss system [42, 43] and next after making corpus preparation actions, we have the parallel corpus (Fig. 4) in order to train the Moses SMT system [1] at the last stage.

Written Greek (Source)
Βροχές και καταιγίδες θα εκδηλωθούν κατά τόπους στη Δυτική Ελλάδα τα Χριστούγεννα . (Rain and thunderstorms will occur locally in western Greece at Christmas.)
GSL Gloss text (export)
ΧΡΙΣΤΟΥΓΕΝΝΑ/CHRISTMAS/NoAcNePlXx
ΜΕΤΑ/after/Pt/ΧΛ(ΜΕΤΑ) ΓΙΝΕΙ/occur/Vb ΒΡΟΧΗ/rain/NoAcFePlXx
ΚΑΙ/and/Cj ΚΑΤΑΙΓΙΔΑ/thunderstorms/NoAcFePlXx/ΜΧ(ΕΝΤΑΣΗ/INTENSITY) /ΜΓΛ(ΦΟΥΣΚΩΜΕΝΑ/BOOKED) ANT_3/there/PreDict /ΜΤ(ΑΝΟΙΧΤΑ/OPEN) ΤΟΠΟΣ/LOCALY/ΤΠΘ(X1)-ΤΟΠΟΣ/LOCALY/ΤΠΘ(X2)/No
ANT_3/THERE/PreDict/MT(ΑΝΟΙΧΤΑ/OPEN)
ΕΛΛΑΔΑ/GREECE/NoAcFeSgXx ΔΥΤΙΚΟΣ/WESTERN/AjAcFeSgXx
./PTERM_P

Fig. 4. Written Greek - Gloss Text

The parallel sentences, of RBMT system, are then word-aligned, typically using GIZA++3, which implements a set of statistical models developed at IBM in the 80s. These word alignments are used to extract phrase-phrase translations, or hierarchical rules as required, and corpus-wide statistics on these rules are used to estimate probabilities. Phrase-Based Models translate phrases as atomic units. The phrase-based statistical machine translation model we present here was defined by Koehn et al. [44]. See also the description by Zens [45].

An important part of the translation system is the language model, a statistical model built using monolingual data in the target language and used by the decoder to try to ensure the fluency of the output.

To estimate the phrase translation probability φ(e|f) we proceed as follows: First, the extract file is sorted. This ensures that all English phrase translations for a foreign phrase are next to each other in the file. Thus, we can process the file, one foreign phrase at a time, collect counts and compute φ(e|f) for that foreign phrase f. To esti-

[7] http://nlp.ilsp.gr/nlp/tagset_examples/tagset_en/adposition.html

mate φ(f|e), the inverted file is sorted, and then φ(f|e) is estimated for an GSL Gloss phrase at a time (Fig. 5). By default, only a distance-based reordering model is included in final configuration. This model gives a cost linear to the reordering distance.

```
$ grep 'πληροφορίες |' ./phrase-table | sort -nrk 7 -t\ | head
πληροφορίες ||| ΠΛΗΡΟΦΟΡΙΑ ||| 1 1 1 1 ||| 0-0 ||| 1 1 1 ||| |||
περισσότερες πληροφορίες ||| ΠΛΗΡΟΦΟΡΙΑ ΠΟΛΥΣ ||| 1 0.625 1 1 ||| 1-0
0-1 ||| 1 1 1 ||| |||
! περισσότερες πληροφορίες ||| ! ΠΛΗΡΟΦΟΡΙΑ ΠΟΛΥΣ ||| 1 0.511364 1
0.9 ||| 0-0 2-1 1-2 ||| 1 1 1 ||| |||
! ! περισσότερες πληροφορίες ||| ! ! ΠΛΗΡΟΦΟΡΙΑ ΠΟΛΥΣ ||| 1 0.418388
1 0.81 ||| 0-0 1-1 3-2 2-3 ||| 1 1 1 ||| |||
```

Fig. 5. Score Phrases

4 Evaluation of the MT System

Human evaluation is fundamental and remains crucial to proper assessment of the quality of MT systems. When the output of an MT system is evaluated, however, the accuracy of translation process is taken into account.

Initially, by performing text mining from several weather-related web pages[8], we have created a large parallel written Greek – GSL Gloss language corpus, consisting of 1,015 sentences and 20,287 tokens. Next the corpus was divided into 2 subcorpuses (one of 100 sentences for evaluation and one large of 900 sentences for training the SMT system). The whole procession translation by RBMT system is supervised by a professional translator, so that output texts are corrected.

For measuring the translation accuracy of the proposed MT system, the Bleu Score [15] for 1 to 4-gram is used.

Our prototyped MT system achieves a relative average score of 60.53% and 85.1%/65.5%/53.8%/44.8% for for 1-gram/2-gram/3-gram/4-gram evaluation. Here it should also be mentioned that the larger the n-gram the better the quality of translation. Nevertheless, we expect in the future to try to improve performance rates by extending to larger corpora sizes and alternative algorithms of Moses Suite.

On the other hand, and for comparison reasons, it is worth noting that similar experiments can be found in the literature. Kanis [46] in his work, the training set consisted of 12,616 sentences, regarding Czech to Czech Sign Language. In these experiments the proposed system reached a BLEU score of 0.81, a WER of 13.14% and a PER of 11.64%. Similarly, in [47] and in case of German to German Sign Language two experiments have been performed. In these cases, the BLEU and PER obtained were 0.021 and 85.7% for the first experiment and 0.026 and 81.1% for the second experiment respectively. However, the reported baseline with the open source toolkit for statistical machine translation Moses [1] was 0.181 BLEU and a 71.0% TER with a training set of 2,565 sentences and a test set of 512 sentences. By combining several systems, they finally reached a BLEU of 0.234 and a TER of 65.5%. Here it should be noted that the disparity between these results is because Czech and Czech Sign Lan-

8 http://www.deltiokairou.gr/, http://www.weather.gr/, http://meteo.gr/

guage have the same surface order, but German and German Sign Language do not. Furthermore, results confirm that data scarcity and domain sparseness lead the data-based approaches to perform worse than the rule-based systems. Providing bilingual lexical resources has a positive effect in data-based approaches. We think that this result should not be interpreted as domain independence. Instead, we consider that data are not still enough to measure the out-of-domain effect. We think that this result should not mean that GSL and Greek have similar word orders or that the order generated by the system is not valid. We consider that GSL order admits some degree of freedom and that the order of signs in the learning corpus is also valid for the purpose of communication. At this point, deeper and more extensive experiments, measuring human understanding, should be performed to draw further conclusions.

5 Conclusions and future work

The choice of a particular type of technology to process a language is greatly influenced by the density of the language, i.e., the availability of digitally stored resources. Commercial research and development have concentrated on high-density languages. Today GSL, like any other sign language, is a low-density or under-resourced language. Because of modality, acquisition of sign language data is a time consuming and expensive task, compared to the acquisition of spoken or written data. Currently is maybe one of the first attempts of creating parallel corpus of sufficient size for written Greek - GSL, which could enable data-driven approaches to machine translation in non-restricted domains. Additionally, the few existing works on the area of creating and analyzing GSL Corpus are copyrighted and thus not open to the researchers or the Deaf communities.

On the other hand, GSL, as all other SLs in the world, is not standardized, and GSL's full grammar has not been published yet. Only some recent works point out important grammar points, lines and references [7, 33, 48]. All these problems make the development of a RBMT system "supervised by a professional translator" the only viable solution. In this case the translator will be enabled to create large, parallel, quality, Greek to GSL corpus, without the need of grammar.

Finally, many other important aspects have not been addressed in this paper, and there is still a great deal of work to do. In particular, the proposed system should be tested using: (a) the factored Translation Model of Moses [1], (b) in other thematic areas, by gathering large relevant corpus, and (c) in the field of SL synthesis (animation), using animation technologies and motion captures technologies in order to have exports to a realistic animation motion of SL and speed up the creation of multimedia dictionary database.

References

1. Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., Herbst, E.: Moses: Open Source Toolkit for Statistical Machine Translation. In: Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions. pp. 177–180. Association for Computational Linguistics, Stroudsburg, PA, USA (2007).
2. Isabelle, P., Foster, G.: Machine Translation: Overview. In: Brown, K. (ed.) Encyclopedia of Language and Linguistics, 2-nd edition. pp. 404–422. Elsevier: Gatineau (2006).
3. Porta, J., López-Colino, F., Tejedor, J., Colás, J.: A Rule-based Translation from Written Spanish to Spanish Sign Language Glosses. Comput. Speech Lang. 28, 788–811 (2014).
4. Klima, E., Bellugi, U.: The Signs of Language, Harvard University Press, Cambridge, Massachusetts. (1979).
5. Edward, T.: Hall, The Silent Language. Garden City, Nueva York. (1959).
6. Stokoe Jr, W.C.: Sign Language Diglossia. (1969).
7. Efthimiou, E., Fotinea, S.-E., Dimou, A.-L., Goulas, T., Kouremenos, D.: From grammar-based MT to post-processed SL representations. Universal Access in the Information Society. 15, 499–511 (2016).
8. Stokoe, J., William C.: Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. The Journal of Deaf Studies and Deaf Education. 10, 3–37 (2005).
9. Sutton, V.: Lessons in Sign Writing: Textbook. SignWriting, La Jolla. (1995).
10. Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Henning, J.: An introductory guide to HamNoSys Version 2.0: Hamburg notation system for Sign Languages. International Studies on Sign Language and Communication of the Deaf, Vol. 5, Signum Press, Hamburg, Germany. (1989).
11. Neidle, C., Sclaroff, S., Athitsos, V.: SignStream: A tool for linguistic and computer vision research on visual-gestural language data. Behavior Research Methods, Instruments, & Computers. 33, 311–320 (2001).
12. Sutton, V.: Sutton Movement Shorthand Dance Writing. Cambridge, MA: The Movement Shorthand Society Press. (1978).
13. Elliott, R., Glauert, J., Jennings, V., Kennaway, J.: An overview of the SiGML notation and SiGML Signing software system., (2004).
14. Augustus, R.A., Ritchie, E., Stecker, S.: The Official American Sign Language Writing Textbook. Los Angeles, CA: ASLized. (2013).
15. Papineni, K., Roukos, S., Ward, T., Zhu, W.-J.: BLEU: A Method for Automatic Evaluation of Machine Translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. pp. 311–318. Association for Computational Linguistics, Stroudsburg, PA, USA (2002).
16. Trujillo, A.: Transfer Machine Translation. In: Trujillo, A. (ed.) Translation Engines: Techniques for Machine Translation. pp. 121–166. Springer London, London (1999).

17. Morrissey, S.: Data-driven machine translation for sign languages, http://doras.dcu.ie/570/, (2008).
18. Chandioux, J.: METEO: an operational system, for the translation of public weather forecasts. In: FBIS Seminar on Machine Translation. American Journal of Computational Linguistics, microfiche. pp. 27–36 (1976).
19. Chandioux, J.: Météo: 100 million words later. In: American Translators Association Conference 1989: Coming of Age, Learned Information, Medford, NJ. pp. 449–453 (1989).
20. Toma, P.: Systran as a multilingual machine translation system. In: Proceedings of the Third European Congress on Information Systems and Networks, Overcoming the language barrier. pp. 569–581 (1977).
21. Hutchins, W.J., Somers, H.L.: An introduction to machine translation. Academic Press London (1992).
22. Efthimiou, E., Fotinea, S.-E., Hanke, T., Glauert, J., Bowden, R., Braffort, A., Maragos, P., Lefebvre-Albaret, F.: Sign Language technologies and resources of the Dicta-Sign project. In: Proc. of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. Satellite Workshop to the eighth International Conference on Language Resources and Evaluation (LREC-2012). pp. 37–44. , Istanbul, Turkey (2012).
23. Efthimiou, E., Fotinea, S.-E.: GSLC: Creation and Annotation of a Greek Sign Language Corpus for HCI. In: Proceedings of the 4th International Conference on Universal Access in Human Computer Interaction: Coping with Diversity. pp. 657–666. Springer-Verlag, Berlin, Heidelberg (2007).
24. Huenerfauth, M.: Generating American Sign Language Classifier Predicates for English-to-asl Machine Translation, (2006).
25. Veale, T., Conway, A.: Cross Modal Comprehension in ZARDOZ an English to Sign-language Translation System. In: Proceedings of the Seventh International Workshop on Natural Language Generation. pp. 249–252. Association for Computational Linguistics, Stroudsburg, PA, USA (1994).
26. Bangham, J.A., Cox, S.J., Elliott, R., Glauert, J.R.W., Marshall, I., Rankov, S., Wells, M.: Virtual signing: Capture, animation, storage and transmission - an overview of the visicast project. In: IEEE Seminar on Speech and language processing for disabled and elderly people (2000).
27. Speers, D.L.: Representation of American Sign Language for Machine Translation, (2002).
28. Sáfár, É., Marshall, I.: Sign Language Translation via DRT and HPSG. In: Gelbukh, A. (ed.) Computational Linguistics and Intelligent Text Processing: Third International Conference, CICLing 2002 Mexico City, Mexico, February 17–23, 2002 Proceedings. pp. 58–68. Springer Berlin Heidelberg, Berlin, Heidelberg (2002).
29. Zhao, L., Kipper, K., Schuler, W., Vogler, C., Badler, N., Palmer, M.: A Machine Translation System from English to American Sign Language. In: White, J.S. (ed.) Envisioning Machine Translation in the Information Future: 4th Conference of the Association for Machine Translation in the Americas, AMTA 2000 Cuernavaca,

Mexico, October 10–14, 2000 Proceedings. pp. 54–67. Springer Berlin Heidelberg, Berlin, Heidelberg (2000).
30. Baldassarri, S., Royo-Santas, F.: An automatic rule-based translation system to Spanish Sign Language (LSE). In: New Trends on Human–Computer Interaction. pp. 1–11. Springer London (2009).
31. Atserias, J., Casas, B., Comelles, E., González, M., Padró, L., Padró, M.: FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. In: Proceedings of LREC. pp. 48–55 (2006).
32. Forcada, M.L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J.A., Sánchez-Martínez, F., Ramírez-Sánchez, G., Tyers, F.M.: Apertium: A Free/Open-source Platform for Rule-based Machine Translation. Machine Translation. 25, 127–144 (2011).
33. Kouremenos, D., Fotinea, S.-E., Efthimiou, E., Ntalianis, K.: A prototype Greek text to Greek Sign Language conversion system. Behaviour & Information Technology. 29, 467–481 (2010).
34. Morrissey, S.: Assessing three representation methods for sign language machine translation and evaluation. In: Proceedings of the 15th annual meeting of the European Association for Machine Translation (EAMT 2011), Leuven, Belgium. pp. 137–144 (2011).
35. Dangsaart, S., Naruedomkul, K., Cercone, N., Sirinaovakul, B.: Intelligent Thai text–Thai sign translation for language learning. Computers & Education. 51, 1125–1141 (2008).
36. Bauer, B., Nießen, S., Hienz, H.: Towards an Automatic Sign Language Translation System. In: 1st International Workshop on Physicality and Tangibility in Interaction: Towards New Paradigms for Interaction Beyond the Desktop. , Siena (1999).
37. Stroppa, N., Way, A.: MaTrEx: DCU Machine translation system for IWSLT 2006. In: Proceedings of the International Workshop on Spoken Language Translation (IWSLT). pp. 31–36. , Kyoto, Japan (2006).
38. Morrissey, S., Way, A.: Manual labour: tackling machine translation for sign languages. Machine Translation. 27, 25–64 (2013).
39. Carpenter, R.: Logic of Typed Feature Structures, The (Cambridge Tracts in Theoretical Computer Science). Cambridge University Press, New York, NY, USA (2005).
40. Carpenter, R.: The logic of typed feature structures. Cambridge University Press, Cambridge, UK (1992).
41. Koleli, E.: A new Greek part-of-speech tagger, based on a maximum entropy classifier, (2011).
42. Hoiting, N., Slobin, D.I.: Transcription as a tool for understanding: The Berkeley Transcription System for sign language research (BTS). G. Morgan & B. Woll (Eds.). Directions in sign language acquisition (pp. 55-75), Amsterdam/Philadelphia:John Benjamins. (2002).
43. Slobin, D.I., Hoiting, N., Anthony, M., Biederman, Y., Kuntze, M., Lindert, R., Pyers, J., Thumann, H., Weinberg, A.: Sign language transcription at the level of

meaning components: The Berkeley Transcription System (BTS). Sign Language & Linguistics. 4, 63–104 (2001).
44. Koehn, P.: Pharaoh: a beam search decoder for phrase-based statistical machine translation models. Machine translation: From real users to research. 115–124 (2004).
45. Zens, R., Och, F.J., Ney, H.: Phrase-based statistical machine translation. In: Annual Conference on Artificial Intelligence. pp. 18–32. Springer (2002).
46. Kanis, J., Müller, L.: Automatic Czech – Sign Speech Translation. In: Matoušek, V. and Mautner, P. (eds.) Text, Speech and Dialogue: 10th International Conference, TSD 2007, Pilsen, Czech Republic, September 3-7, 2007. Proceedings. pp. 488–495. Springer Berlin Heidelberg, Berlin, Heidelberg (2007).
47. Stein, D., Schmidt, C., Ney, H.: Sign language machine translation overkill. In: Federico M, Lane I, Paul M, Yvon F (eds) International workshop on spoken language translation. pp. 337–344. , Paris, France (2010).
48. Fotinea, S.-E., Efthimiou, E., Kouremenos, D.: Generating linguistic content for Greek to GSL conversion. In: Proc. of the HERCMA-2005 Conference (The 7th Hellenic European Conference on Computer Mathematics & its Applications). , Athens, Greece (2005).