A GSL Continuous Phrase Corpus: Design and Acquisition

Athanasia–Lida Dimou*, Vassilis Pitsikalis, Theodor Goulas*, Stavros Theodorakis**, Panagiotis Karioris*, Michalis Pissaris*, Stavroula–Evita Fotinea*, Eleni Efthimiou*, Petros Maragos****

*Institute for Language and Speech Processing – R.C. "Athena",
Artemidos 6 & Epidavrou, GR 151 25 Maroussi, Greece
**NTUA, School of E.C.E., National Technical University of Athens, Greece
Heroon Polytechneiou 9, GR-15780 Zografou Campus
E-mail: ndimou@ilsp.gr, pitsik@cs.ntua.gr, tgoulas@ilsp.gr, sth@atrion.gr, pkarior@ilsp.athena-innovation.gr, mpissaris@gmail.com, evita@ilsp.gr, eleni_e@ilsp.gr, petros.maragos@cs.ntua.gr

Abstract

In this study, we present a newly built Greek Sign Language (GSL) corpus. The procedures followed during its implementation, consists of the linguistic design and validation, the studio and hardware acquisition configuration, the implementation and supervision of the acquisition itself, and the post-processing of the annotations for the release of accompanying linguistic/annotation resources. The reported GSL phrase corpus forms the basis for machine learning and training to serve continuous sign language processing and recognition.

Keywords: GSL, Kinect Camera, SL resources, sign recognition, continuous sign recognition

1. Introduction

The corpus presented in this article is composed of a limited number of Greek Sign Language (GSL) sentences and was created in order to provide additional data to the already obtained corpus during the first year of the Dicta-Sign project (Matthes et al., 2010). More specifically this corpus intended to serve as the ground upon which a significant part of the recognition process would be tested and evaluated, more precisely, the continuous sign language recognition algorithms developed in the project.

Given the targeted nature of this corpus we next present step by step the constraints as well as the procedure followed in order to obtain it.

2. Methodology

Methodologically, the creation of the GSL phrase corpus made re-use of already existing data that were acquired within the Dicta-Sign project as part of the project's "parallel" corpus and lexicon resources. However, the initially acquired data were characterised by a number of restrictions relating to their specific discourse content and, most significantly, not incorporating the whole range of parameters required for running the continuous sign language recognition experiments.

Limitations noticed in respect to the content of the semi-spontaneous Dicta-Sign corpus, were noticed in relation to the variety of sign formation parameters, significant for recognition processing, such as location and handshape. Such parameters had limited a lot the volume of useful segment of the initially acquired corpus. The GSL phrase corpus presented in the rest of this paper, contains GSL phrases of simple to modest complexity.

To meet sign language recognition experimentation requirements, the phrases that formed the corpus were selected to be significant for Sign Language linguistic analysis and also for their employment in automatic recognition tasks, given that in sign language recognition terms, simple phrases constitute a task of intermediate complexity when compared to a more open and unconstrained continuous corpus[1].

2.1 The GSL Dicta-Sign Corpus

The GSL phrase corpus is directly related to parts of the previously acquired Dicta-Sign parallel corpora (Matthes et al., 2010).

The existing Dicta-Sign GSL corpus has been exploited in order to extract from it small in length and simple in structure phrases that would become the stimuli for the GSL phrase corpus.

The Dicta-Sign corpus was built upon using real life discourse situations between native GSL signers e.g. being at the airport, travelling etc. Those real life situations were divided into tasks that the signers had to reproduce in front of the camera after being instructed thoroughly on how to do so. The complete GSL Dicta Sign corpus was created with the use of nine different elicitation tasks that were performed by eight pairs of native adult Deaf signers.

2.1.1 Task 4 of the GSL Dicta-Sign Corpus

The GSL phrase corpus covers lexically one of the nine elicitation tasks of the Dicta-Sign parallel corpus, namely the fourth one. This task (Dicta-Sign Task 4) treats a specific topic; the situation in the Airport entailing communications on issues of check-in, luggage depository, boarding, safety instructions, meals, take off and landing. There are many reasons for choosing this

[1] A first exploitation of the newly acquired Continuous Recognition Training Phrase Corpus will be seen via the Dicta-Sign project Demonstrator (currently available at: http://signwiki.cmp.uea.ac.uk/dictasign).

specific task, the most important of which is that in all four project sign languages, this was the task that presented the strongest similarities in terms of common topics addressed and resulted in the most comparable parts of the Dicta-Sign corpus among the involved four languages. For this reason this task was also given priority in the manual annotation procedure, the latter being a time-consuming and labour-intensive process.

2.1.2 Annotation of the GSL Dicta-Sign Corpus

Availability of annotations for the complete Task 4 segment was another reason to select this specific task. The annotation in the GSL segment of the Dicta-Sign corpus included the following four tiers:

- GSL Lemmas (creation of the lexicon entries in Greek that would most suitably fit the GSL lemmas),
- Clause boundaries (the continuous signing was cut into clauses),
- English translation of the clauses,
- HamNoSys transcription of the GSL lemmas.

2.2 The "list of 1000 Common Concepts"

In order to obtain the lexical entries that the phrases of the GSL phrase corpus were going to be composed of, we compared the lexical entries found in Task 4 of the Dicta-Sign corpus to a list of concepts assigned lexical entries in all four project languages, that served during the first two years of the Dicta-Sign project as the basic common lexicon among the project sign languages.

Throughout the course of the project these lexical entries, consisted a means of exchanging data upon a common ground, which henceforth will be referred to as "The list of 1000 Common concepts".

These "1000 common concepts" apart from being a reference point and a visible outcome of the project[2], became the object of another recording that took place in year 2 of the project for GSL and German Sign Language (DGS). This recording aimed at obtaining data on the handshape, and the movement (trajectory, orientation etc) that is effectuated during the performance of each one of these signs. For this reason extensive footage sessions took place that tracked with a 3D camera the trajectories the signers performed for each one of the 1000 corresponding signs in each language. In notation terms, the whole set of these lexical entries are represented via HamNoSys notations.

The representation of signing of the concepts was transformed computationally, in a way to relate the signing with Postures, Detentions, Transitions and Steady Shifts (PDTS). PDTS is a sequential model proposed by Johnson & Liddell (2011) to capture and label the sequential structure in sign language at the level of linguistic phonetic units.

Every sign is further represented with HamNoSys notation. This is a phonetic transcription for signing language lemmas (Hanke, 2004). The first step in the adopted procedure was to transform the HamNoSys representation into structured sequences of labels in the form of Gestural SiGML (Signing Gesture Markup Language). This is a form of representing gestures in a structured sequence, better understandable by human readers than the HamNoSys form. The next step was to convert the results into segmented SiGML and finally summarise the segmentation in a set of PDTS labels. These sets of labels were used as a basis for the training of the algorithm treating sign location and movement. Combining this information with skeleton tracking information it was possible to align sequences of structured labels with visual data segments (Pitsikalis et al., 2011)

The above mentioned procedure as well as the measurements that took place were essential and allowed to support the recognition processing the linguistic treatment, as well as the synthesis procedures maintaining the ability to represent lemmas by means of HamNoSys notations.

2.3 Comparison of the two lists

The comparison of the two lists of sign lemmas, deriving from the Dicta-Sign resources, resulted to an overlapping between the two lists that consisted of 113 GSL lemmas. For each of these lemmas the following information is available: a. Gloss (written in Greek), b. HamNoSys transcription, c. English translation, d. kinect based skeleton tracking information. These 113 GSL lemmas from the Dicta-Sign corpus became the repository upon which the GSL phrase corpus was based.

3. The GSL phrase corpus

The 113 GSL lemmas that came out of the comparison of the two different lists, were initially examined in terms of the linguistic resources to be considered, namely: the vocabulary units in relation to the various lexical types and the frequency of occurrence of the considered lemmas. The methodology adopted in order to obtain the GSL phase corpus can be divided into two parts, which are presented in subsections 3.1 and 3.2.

3.1 Original Phrases from the Corpus

The 113 GSL lemmas were located within the transcripts of the recordings of Task 4, as they were originally uttered by the GSL signers of the Dicta-Sign Corpus.

Given that Task 4 was one of the Tasks that were fully annotated very early in the timeline of the project, ,we were able to locate the 113 GSL lemmas within the sentence-level annotations.

The number of phrases that contained one or more of the 113 lemmas within the eight[3] transcripts of Task 4 was

[2] http://www.sign-lang.uni-hamburg.de/dicta-sign/consign/demo/cs/cs_51.html

[3] Task 4 was a task that was performed only from one of the two signers who participated in each session of recordings for the Dicta-Sign corpus. So in this case the 113 lemmas were crosschecked across the eight transcripts, one signer per pair.

more than 300; nonetheless, our goal was to obtain those phrases that contained the maximum number of items from the 113 lemmas set in each phrase. With this criterion a selection was made leaving out those phrases that contained only 1 or only 2 lemmas from the 113.
After checking vocabulary coverage, a qualitative criterion was applied: many of the retrieved phrases were significantly long and could not be performed by the signer.. By excluding the complex and long phrases we reached the basic set of 137 simple ones that constituted the pilot part of our corpus.

3.1.1 Creating the videos for the elicitation material

The original videos of the phrases have been cut into new separate video files according to the video time codes found in the related transcripts.
The software that has been used for the annotation of the GSL sign language corpora is iLex (Hanke & Storz,.2008).
This procedure, even though it may seem trivial, entailed some major difficulties as, for example, the correct time boundaries extraction out of the iLEX transcript files.
Using information from transcript files, the phrases of interest were located and a list of them was created. According to this list and the transcription file, where information on video time codes was available, the original video files were chopped in smaller ones so that every phrase of interest was entailed into a separate video file.

3.1.2 Elicitation procedure – Recording sessions

The video files of the selected phrases were employed to construct the elicitation material to be presented to the GSL native signer. The signer was asked to repeat the phrase that he saw on the monitor, as close to the original production as possible. The signer was allowed as many repetitions as he wished.
Unfortunately, in naturally uttered signed speech, signers very rarely perform pieces of language that can be reproduced with the minimum set of instructions by other signers. This is mainly the reason why the original phrases in which the 113 lemmas were found served only as a pilot study for the data acquisition process of the set of formally defined simple phrases presented in 3.2.

3.2 Formally Defined Simple Phrases

The above mentioned procedure as well as the repository of the 137 original phrases of Task 4 functioned as the cast upon which 56 formally defined phrases were produced. These are 56 phrases that were put together by means of gloss ordering, which combined the 113 lemmas into phrases that obey the grammar rules of GSL.
The 56 phrases were evaluated by a native GSL signer and they are partitioned in the following sets:

a) Phrases Set I (PS1): Contains a set of 20 simple continuous phrases;

b) Phrases Set II (PS2): Contains a set of 28 of slightly more complex continuous phrases;

c) Supplementary Phrase Set III (PS3): Contains a set of 8 phrases to lexically cover for missing signs.

3.2.1 Elicitation procedure – Recording sessions

Since the formally defined phrases are phrases that were put together by ordering the involved glosses, there were no available video to show to the signer.
During the recordings a GSL interpreter performed each phrase and, if needed, explained each phrase to the Deaf adult native signer. There were no limitations in the number of repetitions to be performed other than the fatigue of the signer.

3.3 The signers

Four signers in total participated in these recording sessions. Only two out of the four signers performed all the above phrase sets (I, II & III) as well as the list of the 113 lemmas (Lexicon (L)) in isolated mode, three times each. Their recordings served as a pilot test bed upon which the final recordings were based.
The final recordings took place with two signers, the Official Signer A" and the "Official Signer B" who performed multiple times the 56 phrases as well as the list of the 113 lemmas of the Lexicon (L). Their acquired data served as the database used in experimentation.

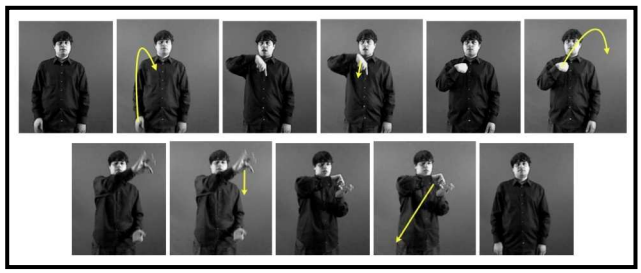

Figure 1: Sample of continuous signing utterance: "TOMORROW I ARRIVE ATHENS" from the GSL Continuous Phrases Dataset (arrows indicate each transition)

4. Data acquisition

The data acquired consisted of:

- High Definition (HD) appearance data employing a High Definition camera with frame rate 25 fps
- Depth and appearance data employing a Kinect sensor and
- Skeleton tracking as obtained utilizing the depth data from the Kinect sensor.
- To collect as much data as possible a second Kinect sensor was used to record also the interpreter.

Signer and interpreter were place opposite one another and the Kinect sensors were placed in the middle, one facing the interpreter and the other facing the signer.
Each camera/sensor was controlled by a different person.
The acquisition was supported by a moderator who supervised the whole procedure, annotated mistakes,

stated the need of extra repetitions, and marked transcriptions updates.

Figure 2: Setup studio setting for the GSL Phrase Corpus

5. Discussion

Herewith presented newly acquired corpus features:

1. topic specific linguistic content
2. structure that simulates simple phrases
3. sharing of linguistic content/vocabulary with a larger and more complex continuous corpus, which can be employed in parallel
4. acquisition of High Definition video data
5. parallel acquisition with the recent high-tech Kinect sensor accounting for both Depth and Skeleton Tracking.

This data set served the purpose of experimentation towards development of continuous sign language recognition algorithms.

Although the corpus is of limited scale, the above features render these data a highly appealing test-bed for interdisciplinary research in the domain of Sign Language and Gesture technology.

6. Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/ 2007-2013) under grant agreement No 231135. We would also like to thank Helen Cooper in University of Surrey for providing the Kinect acquisition software.

7. References

Hanke, T. (2002). iLex - A tool for sign language lexicography and corpus analysis. In Proceedings of the 3rd International Conference on Language Resources and Evaluation. Las Palmas de Gran Canaria, Spain. ELRA Paris pp. 923—926.

Hanke, T. (2004). HamNoSys – representing sign language data in language resources and language processing contexts. In LREC 2004, Workshop proceedings: Representation and processing of sign languages. ELRA Paris, 2004, pp. 1-6.

Hanke, T. iLex: http://www.sign-lang.uni-hamburg.de/ilex/

Hanke, T., Storz, J. (LREC 2008) iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In LREC 2008, Work shop proceeding of the 3rd Workshop on the Representation and Processing of Sign Languages, Maroko 2008.

Johnson, R. E. and Liddell, S. K. (2011). A segmental framework for representing signs phonetically. sign language studies. 11(3).

Matthes, S., Hanke, T., Storz, J., Efthimiou, E., Dimou, A-L. Karioris, P., Braffort, A., Choisier, A., Pelhate, J. and Safar, E. (2010). Elicitation Tasks and Materials designed for Dicta-Sign's Multi-lingual Corpus. In Dreuw, P. et al., (eds.), *LREC-2010, Proceedings of 4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pp. 158-163.

Pitsikalis, V., Theodorakis, S., Vogler, C., Maragos P. (2011) Advances in Phonetics-based Sub-Unit Modeling for Transcription Alignment and Sign Language Recognition, Workshop on Gesture Recognition and launching of a Benchmark program (CVPR2011), 2011 Colorado, Springs, USA.