SL-REDU GSL: A LARGE GREEK SIGN LANGUAGE RECOGNITION CORPUS

Katerina Papadimitriou [1], *Galini Sapountzaki* [2], *Kyriaki Vasilaki* [3], *Eleni Efthimiou* [3], *Stavroula-Evita Fotinea* [3], *Gerasimos Potamianos* [1]

[1] Department of Electrical & Computer Engineering, University of Thessaly, Volos, Greece
[2] Department of Special Education, University of Thessaly, Volos, Greece
[3] Institute for Language & Speech Processing, Athena Research & Innovation Center, Athens, Greece

Emails: aipapadimitriou@uth.gr, gsapountz@sed.uth.gr, {kvasilaki, eleni_e, evita}@athenarc.gr, gpotam@ieee.org

ABSTRACT

We present a large multi-signer video corpus for the Greek Sign Language (GSL), suitable for the development and evaluation of GSL recognition algorithms. The database has been collected as part of the "SL-ReDu" project that focuses on the education use-case of systematic teaching of GSL as a second language (L2). The project aims to assist this process by allowing self-monitoring and objective assessment of GSL learners' productions through the use of recognition technology, thus requiring suitable data resources relevant to the aforementioned use-case. To this end, we present the SL-ReDu GSL corpus, an extensive RGB+D video collection of 21 informants with a duration of 36 hours, recorded under studio conditions, consisting of: (i) isolated signs; (ii) continuous signing (annotated at the sentence level); and (iii) fingerspelling of words. We provide a detailed description of the design and acquisition methods used to develop it, along with corpus statistics and a comparison to existing sign language datasets. The SL-ReDu GSL corpus, as well as proposed frameworks for recognition experiments on it, are publicly available at https://www.sl-redu.e-ce.uth.gr/corpus.

***Index Terms*—** Greek sign language recognition, SLR datasets, SL-ReDu, GSL, sign language translation

1. INTRODUCTION

Automatic sign language recognition (SLR) is an important human-computer interaction technology, critical to accessibility and inclusion of the deaf and hard-of-hearing population. To this end, the SL-ReDu project [1, 2] aims to develop such technology for GSL using novel deep-learning algorithms, with the goal of supporting the standardized teaching of GSL as L2 by automating the process of student self-monitoring and assessment. Achieving this requires the availability of appropriate GSL data resources relevant to the SL-ReDu education use-case.

Although a considerable amount of annotated GSL video data are available, such as the ITI-GSL dataset [3] and the Polytropon parallel corpus [4], the specific educational curriculum content required by the SL-ReDu use-case is not sufficiently covered in existing databases. For this reason, we collected the SL-ReDu GSL corpus to improve the field of GSL recognition, alignment, and translation by providing a sizable signer population and extended lexical coverage.

This paper provides an overview of the design and acquisition procedures used to create the SL-ReDu corpus, which covers the area of language education along with some general content. Regarding data acquisition, we describe the developed studio setup with two cameras, namely an RGB and a depth camera. In addition, we present details about the recording protocol used and subject recruitment strategy. Finally, we provide statistical data of the corpus and compare it against a number of existing SLR datasets in the literature.

The paper is structured as follows: Section 2 describes in detail the creation of the SL-ReDu GSL corpus; Section 3 presents corpus statistics and compares it with existing large datasets; finally, Section 4 concludes the paper.

2. CREATION OF THE SL-REDU CORPUS

In this section, we first describe the corpus content, followed by the data recording protocol, the informant recruitment strategy, the acquired data, and the studio setup used.

2.1. SL-ReDu Corpus Content

The content of the SL-ReDu GSL corpus derives from the definition of the language material for levels A0-A1 of the Common European Framework for Languages (CEFRL) for GSL as L2, as discussed in detail in [5, 6]. In particular, the corpus is organized in three subsets: (i) single word units; (ii) GSL phrases; and (iii) GSL alphabet fingerspelled words.

Concerning isolated signs, we consider a total of 369

This work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "1st Call for H.F.R.I. Research Projects to support Faculty Members & Researchers and the procurement of high-cost research equipment grant" (Project "SL-ReDu", Project Number 2456).

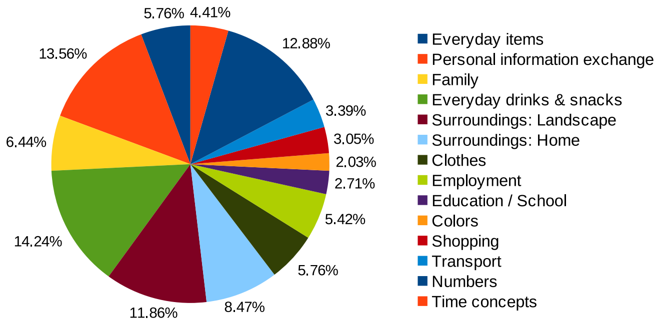

Fig. 1. Distribution of the 369 GSL signs of the SL-ReDu isolated-sign corpus part into 14 semantic categories.

unique lexical units. The semantic and structural organization criteria of the language material are based on morphological and articulatory features according to handshape, type of movement, single- or double-handedness, and symmetry of movements between the two hands. The content related to the core grammar and lexicon provides a well-defined paradigm of the generative capacity of the language with sufficient examples for the 3D articulation of GSL. In addition, the content expands to include SL-specific lexical categories expected for communication at this level. A distribution of the number of isolated signs per semantic category is shown in Fig. 1. Note that these also form the pool for extracting morphological and syntactic rules in GSL phrases and sentences.

The main phrase types included in the SL-ReDu GSL corpus are: (i) affirmative, interrogative and negative phrases; (ii) basic noun and verb phrases; (iii) common noun phrasal modifiers, including the use of numerals, quantifiers, qualifiers and classifiers; (iv) inflection for person and numerus with respect to all six different positions in the syntactic person marking space; (v) subject-comment constructions at the phrase and sentence level; (vi) agent/patient roles in singular and plural as inflected in each verb type; and (vii) non-manual articulators with linguistic value at the phrase level. Semantically, the content covers several important aspects of elementary communication skills involving personal information exchange, description of everyday objects, environment, numbers, time concepts, etc. Table 1 provides an overview of the typological categories of the corpus, along with examples of constructs corresponding to each type. In total, 799 phrases are included in the SL-ReDu corpus, with their annotation, as well as their translation into spoken Greek obtained by GSL experts.

Regarding fingerspelling data, all 24 Greek alphabet letters are included, as well as 926 GSL alphabet fingerspelled words composed of 3-6 letters. Fingerspelling is an often overlooked practical component of signing, commonly used for distinctive words that lack dedicated signs, such as names, technical terms, foreign words, or numerals [7]. More precisely, fingerspelling is a transcription system for representing letters of the alphabet by conventional hand gestures that exhibit features of both SL and the respective spoken language.

Table 1. Core phrase types in the SL-ReDu GSL corpus with examples that include linguistic notation (capitalized) and translation into spoken Greek (lower-case, but provided here in their English translation for readability purposes).

Phrase Types	Examples (annotation/translation)
Affirmative	IX3 APPLE WANT He/she wants
Interrogative	MONEY HAVE IX2 Do you have money?
Negative	IX3 SIBLINGS HAVE He/she doesn't have siblings
Noun phrase	WATER cl-vol: thick a lot of water
Verb phrase	1GIVE3 give-to-multiple-recipients
Numerals	IX1 SIBLINGS HAVE THREE I have three siblings
Quantifiers	WATER cl-vol: thin a little water
Qualifiers	JUMPER cl-vol: thick a thick jumper
Classifiers	PLATE cl-SASS: round a round plate
Inflection for person	IX3 CYCLE CANNOT He/she can't cycle
Inflection for number	THREE-HOURS three-hours
Topic-comment	IX3 HAVE-A-WALK LIKE NOT He/she doesn't like going-for-walks
Agent-patient	2ASK3a,3b, 3c You ask them
Expression	IX2 PAY Have you paid?
Shoulders	IX3 COFFEE LIKE Does he like coffee?

2.2. Recording Protocol and Informant Recruitment

The elicitation material for the corpus creation is video-based, corresponding to a list of items described in Section 2.1. After the list of necessary items to be included was formed, a contrastive search took place within the already available Polytropon database [4] and the Dicta-Sign corpus [8], in order to identify data that could be directly used for elicitation purposes. Additional captures were also recorded, in order to complete the lexical elicitation material. Note that this process concerns only the lexical units and the GSL phrases, while the elicitation for fingerspelling is text-based.

Afterwards, a recruitment strategy was developed, yielding a total of 21 informants of both genders (11 females and 10 males) and ages between 19 and 52 years old, sufficiently representative of the current state of GSL use. The informants were born deaf individuals with early GSL acquisition as first language, as well as bilingual GSL signers. Note that, prior to participating in the recording process, all subjects signed an informed consent form on handling personal information (approved by the AthenaRC Ethics Committee) and completed a brief demographic information questionnaire.

Fig. 2. Example RGB frames from the SL-ReDu GSL video data collection of 21 informants during continuous signing.

2.3. Acquired Data

In the case of isolated signs, video data from the 21 signers have been collected, where each signer articulated the 369 signs 3 consecutive times. Note that one signer performed only 184 isolated signs. In the case of continuous signing, the 799 phrases were divided into 3 different groups. Each of the 20 informants performed the content of only one group once, while one performed only 136 phrases and one performed all the 799 phrases. Concerning fingerspelling, each informant signed once the 24 GSL alphabet letters in isolation, as well as a set of 50 fingerspelled words composed of 3-6 letters, which was unique to each signer. Note that each signer signed each alphabet letter at least 4 times in total.

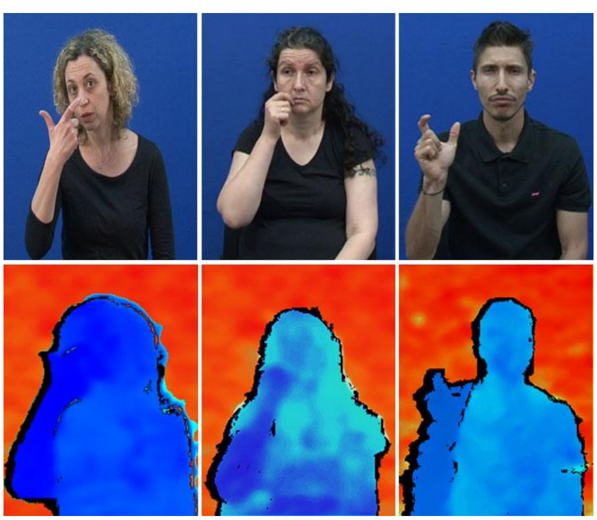

Fig. 3. Corpus example frames from the RGB camera (upper row) and the depth camera (bottom row).

2.4. Studio Setup for Data Collection

We next detail the setup used for the studio data collection process, namely the two cameras used for recording, as well as the tool developed for data collection. Studio video recordings took place inside the Institute for Language and Speech Processing (ILSP) at AthenaRC in Athens. The studio setup involves two cameras: (i) a Sony HVR-Z7N camera, allowing HD video capture at a 1920×1080-pixel resolution at 60 Hz (interlaced video) and (ii) an Intel RealSense D435i RGB+D camera at 30 Hz. In the latter case, only the depth stream was acquired, at a 848×480-pixel resolution with 16-bit depth format. In case of the RGB video stream, we have chosen video interlacing to be able to capture fine temporal information of SL articulation at some loss of vertical spatial resolution. To increase variability in the videos, some videos involve informants in a sitting position while others in a standing one. Example RGB frames are shown in Fig. 2, whereas depth frames

Fig. 4. Data recording setup, showing the informant in front of the two cameras and the video elicitation material display.

Table 2. SL-ReDu GSL recognition datasets statistics.

Task	Signers	Unique content	Vocab. size	Avg. units /video	Videos	Frames	Duration (hrs:mins)
Isolated	21	369 signs	369 signs	1 sign	22,632	2,715,840	25:15
Continuous	21	799 sentences	408 glosses	2.86 glosses	5,930	889,500	8:24
Fingerspelling	21	950 words	24 letters	4.55 letters	1,554	234,360	2:17
Total	21	—	—	—	30,116	3,839,700	35:56

Table 3. Summary of large-scale SLR datasets of isolated signing (Iso.), continuous signing (Con.), and fingerspelling (Fing.).

Dataset	Type	Language	Signers	Vocabulary	Videos	Duration (hrs:min)	Modalities
Signum [9]	Iso.	German	25	455	11,375	8:26	RGB
MSASL [10]	Iso.	English	222	1,000	25,513	24:39	RGB
Polytropon [4]	Iso.	Greek	1	2,703	3,517	8:19	RGB+D
ITI-GSL [3]	Iso.	Greek	7	310	40,785	6:26	RGB+D
AUTSL [11]	Iso.	Turkish	43	226	38,336	21:00	RGB+D
BOBSL[12]	Iso.	English	39	2,281	1,940	1,467:–	RGB
CSL [13]	Iso.	Chinese	50	500	125,000	67:45	RGB+D
Phoenix [14]	Con.	German	9	1,231	6,841	10:43	RGB
Phoenix-T [15]	Con.	German	9	1,231	8,257	10:32	RGB
CSL [16]	Con.	Chinese	50	178	25,000	100+:–	RGB+D
ITI-GSL [3]	Con.	Greek	7	310	10,295	9:35	RGB+D
Signum [9]	Con.	German	25	780	19,500	55:18	RGB
ChicagoFSWild [17]	Fing.	English	160	26	7,304	1:47	RGB
ChicagoFSWild+ [18]	Fing.	English	260	26	55,232	–	RGB
SL-ReDu	Iso.	Greek	21	369	22,632	25:15	RGB+D
	Con.	Greek	21	408	5,930	8:24	RGB+D
	Fing.	Greek	21	24	1,554	2:17	RGB+D

(with their corresponding RGB ones) are depicted in Fig. 3.

For the studio recording process we developed a tool that allows automated signing video data collection, where a monitor is placed in front of the informant to project both messages and the elicitation material (see also Fig. 4). Specifically, a signing video is displayed in the monitor to notify the informant of the sign to be performed. Note that our application provides the potential for watching the video as many times as the signer needs before signing. When the signer comprehends the sign to be expressed, a message to start signing is displayed. After the informant's production, a message is displayed asking if the informant wishes to repeat it, and, if not, the RGB and depth signing videos are saved. The resulting files follow a naming convention involving the signed content, signer id, sign repetition, and video type. To obtain synchronization of the two cameras, we adopt the use of multi-threading, employing two different threads, one for each camera, with their execution starting at the same time and lasting for the same time duration. It should be noted that cameras record for a specific time interval depending on the signing type, i.e., isolated, continuous, or fingerspelling.

3. SL-REDU CORPUS OVERVIEW

As shown in Table 2, the SL-ReDu corpus includes three different parts: (i) isolated signing; (ii) continuous signing; and (iii) fingerspelling. In particular, the corpus comprises 22,632 isolated sign videos with 369 unique signs, 5,930 continuous GSL videos with 799 unique sentences (408 vocabulary size and 2.86 glosses per sentence on average), and 1,554 fingerspelled videos with a vocabulary size of 24. The total duration of the SL-ReDu corpus is 35 hours and 56 minutes. Note that the recorded videos were checked by GSL experts, removing videos with incorrect signing content. As also shown in Table 3, our GSL corpus is the only one that includes isolated, continuous, and fingerspelling subsets, providing both RGB and depth modalities. It is also the largest GSL database available, both in terms of duration and number of signers.

4. CONCLUSION

In this paper, we introduce the SL-ReDu GSL corpus, an extensive RGB+D video collection suitable for GSL recognition, which covers the area of language education along with some general content. This database has been collected as part of the SL-ReDu project that focuses on the education use-case of systematic teaching of GSL as second language. Unlike most available SL corpora, it contains three distinct RGB+D video subsets: (i) isolated signs; (ii) continuous signing; and (iii) fingerspelling. The dataset is made publicly available, and our immediate plans include the release of experimental results of baseline SLR models on it.

5. REFERENCES

[1] G. Potamianos, K. Papadimitriou, E. Efthimiou, S. Fotinea, G. Sapountzaki, and P. Maragos, "SL-ReDu: Greek Sign Language recognition for educational applications. Project description and early results," in *Proc. ACM Int. Conf. on PErvasive Technologies Related to Assistive Environments (PETRA)*, 2020.

[2] E. Efthimiou, S.-E. Fotinea, C. Flouda, T. Goulas, G. Ametoglou, G. Sapountzaki, K. Papadimitriou, and G. Potamianos, "The SL-ReDu environment for self-monitoring and objective learner assessment in Greek sign language," in *Proc. Universal Access in Human-Computer Interaction. Access to Media, Learning and Assistive Environments*, 2021, pp. 72–81.

[3] N. Adaloglou, T. Chatzis, I. Papastratis, A. Stergioulas, G. Th. Papadopoulos, V. Zacharopoulou, G. J. Xydopoulos, K. Atzakas, D. Papazachariou, and P. Daras, "A comprehensive study on deep learning-based methods for sign language recognition," *IEEE Transactions on Multimedia*, vol. 24, pp. 1750–1762, 2022.

[4] E. Efthimiou, K. Vasilaki, S.-E. Fotinea, A. Vacalopoulou, T. Goulas, and A.-L. Dimou, "The POLYTROPON parallel corpus," in *Proc. Works. on the Representation and Processing of Sign Languages: Involving the Language Community (Satellite to LREC)*, 2018, pp. 39–44.

[5] G. Sapountzaki, E. Efthimiou, S.-E. Fotinea, K. Papadimitriou, and G. Potamianos, "Educational material organization in a platform for Greek Sign Language self monitoring and assessment," in *Proc. Int. Conf. on Education and New Learning Technologies (EDULEARN)*, 2021, pp. 3322–3331.

[6] G. Sapountzaki, E. Efthimiou, S.-E. Fotinea, K. Papadimitriou, and G. Potamianos, "3D Greek Sign Language classifiers as a learning object in the SL-ReDu online education platform," in *Proc. Int. Conf. on Education and New Learning Technologies (EDULEARN)*, 2022, pp. 6146–6153.

[7] K. Papadimitriou and G. Potamianos, "Multimodal sign language recognition via temporal deformable convolutional sequence learning," in *Proc. Interspeech*, 2020, pp. 2752–2756.

[8] S. Matthes, T. Hanke, A. Regen, J. Storz, S. Worseck, E. Efthimiou, A.-L. Dimou, A. Braffort, J. Glauert, and E. Safar, "Dicta-Sign – building a multilingual sign language corpus," in *Proc. Works. on the Representation and Processing of Sign Languages: Interactions Between Corpus and Lexicon (Satellite to LREC)*, 2012, pp. 117–122.

[9] U. von Agris, M. Knorr, and K. Kraiss, "The significance of facial features for automatic sign language recognition," in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG)*, 2008.

[10] H. R. V. Joze and O. Koller, "MS-ASL: A large-scale data set and benchmark for understanding American Sign Language," in *Proc. British Machine Vision Conf. (BMVC)*, 2019.

[11] O. M. Sincan and H. Y. Keles, "AUTSL: A large scale multi-modal Turkish sign language dataset and baseline methods," *IEEE Access*, vol. 8, pp. 181340–181355, 2020.

[12] S. Albanie, G. Varol, L. Momeni, H. Bull, T. Afouras, H. Chowdhury, N. Fox, B. Woll, R. Cooper, A. McParland, and A. Zisserman, "BOBSL: BBC-Oxford British sign language dataset," *CoRR*, arXiv:2111.03635, 2021.

[13] J. Pu, W. Zhou, and H. Li, "Sign language recognition with multi-modal features," in *Advances in Multimedia Information Processing - PCM 2016: 17th Pacific-Rim Conference on Multimedia*, E. Chen, Y. Gonh, and Y. Tie, Eds., pp. 252–261. Springer, 2016.

[14] J. Forster, C. Schmidt, O. Koller, M. Bellgardt, and H. Ney, "Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-weather," in *Proc. Int. Conf. on Language Resources and Evaluation (LREC)*, 2014, pp. 1911–1916.

[15] N. C. Camgöz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural sign language translation," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7784–7793.

[16] J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li, "Video-based sign language recognition without temporal segmentation," in *Proc. AAAI Conf. on Artificial Intelligence*, 2018, pp. 2257–2263.

[17] B. Shi, A. Martinez Del Rio, J. Keane, J. Michaux, D. Brentari, G. Shakhnarovich, and K. Livescu, "American sign language fingerspelling recognition in the wild," in *Proc. IEEE Spoken Language Technology Works. (SLT)*, 2018, pp. 145–152.

[18] B. Shi, A. Martinez Del Rio, J. Keane, D. Brentari, G. Shakhnarovich, and K. Livescu, "Fingerspelling recognition in the wild with iterative visual attention," in *Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, 2019, pp. 5399–5408.