Multimodal Corpus Lexicography: Compiling a Corpus-based Bilingual Modern Greek – Greek Sign Language Dictionary

Anna Vacalopoulou, Eleni Efthimiou, Kiki Vasilaki
ILSP-Institute for Language and Speech Processing / Athena RC
E-mail: avacalop@ilsp.gr, eleni_e@ilsp.gr, kvasilaki@ilsp.gr

Abstract

This paper describes the process of compiling NOEMA+, a bilingual dictionary of approximately 12,000 entries for the pair Greek Sign Language (GSL) - Modern Greek (MG) and of making it available openly online (http://sign.ilsp.gr/signilsp-site/index.php/el/noima/). The dictionary was based on several corpora that have been collected over the years, including information on compounding, GSL synonyms, classifiers, various lemma-related senses, semantic relationships, etc. These different corpora have been joined, normalized and translated into MG to form a parallel corpus of the language pair in question. In turn, this parallel corpus acted as the basis for the compilation of the bilingual dictionary described in this paper.

More specifically, among the issues to be discussed here are lemma identification, which proved far from intuitive for this particular language pair, lemma categorization, dictionary contents and structure, relations between entries as well as the corpus which was used for dictionary compilation. Finally, there will be a description of the different search choices offered, which cater for different user profiles and needs.

Keywords: bilingual lexicography, corpus-based lexicography, multimodal parallel corpus, sign language lexicography

1 Introduction and Background

In recent years, awareness raising efforts have placed the issue of equal rights at the top of national, European and international policy agendas. From the Convention on the Rights of Persons with Disabilities in 2006 (CRPD) to the EU Disability Strategy 2010-2020 (European Disability Strategy 2010-2020) there is an international movement towards providing all persons with the same rights, meaning, among other things, total accessibility to knowledge and information. In the case of deaf and hard-of-hearing (HoH) persons, several attempts have been made towards this goal. Among the many systematic attempts to compile European Sign Language (SL) corpora, are the British Sign Language Corpus (Schembri 2008), the Corpus NGT (Crasborn, O. & Zwitserlood 2008) of the Sign Language of the Netherlands, the e-LIS corpus of the Italian Sign Language (Vettori 2008), the German Sign Language Corpus (Konrad 2009), and the Swedish Sign Language Corpus (Mesch J. & Wallin 2015). At the same time, a number of lexicographic attempts have also been made. Indicative projects include, among others, dictionaries of Spanish SL (Fuertes et al. 2006), of Danish SL (Kristoffersen, & Troelsgård 2010), of Polish SL (Linde-Usiekniewicz et al. 2014), and of British SL (Fenlon 2014).

The dictionary described in this paper was first created as part of a specially designed workbench bringing together several Language Technology (LT) tools and technologies. The purpose of this venture was to provide better accessibility to an official educational content platform of the Greek Ministry of Education to deaf and HoH users, i.e. in their native language (Efthimiou et al. 2016). To this end, the workbench needed to provide all the necessary tools for the target group to make

full use of both the content in the platform in question and the Graphical User Interface (GUI) features.[1]

In order to approach this diverse material, which was designed for primary and secondary education levels, we brought together a combination of LT tools (Efthimiou et al. 2017) to handle both SL as well as written language, including a tagger and a lemmatizer. As far as resources are concerned, we made full use of already available monolingual GSL corpora as well as bilingual MG-GSL dictionaries. The result of that work was a fully functioning bilingual service, offering to the target audience bilingual and monolingual dictionary look-up, fingerspelling facilities as well as a dynamic SL synthesis environment.

As part of this project, NOEMA+, the dictionary described in this paper, underwent several stages of evaluation (ibid.) both internally and externally by professionals, GSL experts (some of whom were native signers), and actual end users. The results of this process, along with the availability of an updated version of the GSL corpus in greater detail, presented us with the opportunity to update and improve NOEMA+ as a separate lexical resource.

Briefly speaking, the tools and technologies that were incorporated in the workbench fall under those created for vocal or written language and those created for sign language.

As far as the first category is concerned, we used a language tool suite (Prokopidis et al. 2011) developed and continually undergoing adjustments by the Institute for Language and Speech Processing (ILSP). This includes a tool that segments character strings within text, a tagger that goes through the segmented units adding to them labels with grammatical information (i.e. morphological and syntactical categories), and a lemmatizer that links each of the segmented and tagged items with a particular lemma in the dictionary. The integration of an improved version of this suite proved to be especially valuable, as morphological complexity is one of the main characteristics of MG, which is a particularly inflectional language (Holton et al. 1997). Indeed, the numerous forms of lemmas that differ considerably or even completely from each other (i.e. irregular types) have been known to present serious barriers to the literacy of deaf users of around the age of the target group (Breadmore 2007). In the case of this platform, such cases would pose problems not only for text comprehension, but also for looking up lexical items in the dictionary.

As regards the SL part of the content, the tools and technologies that were incorporated in the workbench include the following: (a) the bilingual dictionary in question, which was linked with content uploaded in the educational platform, namely textbooks, offering different search possibilities; (b) a keyboard allowing virtual fingerspelling, which consists of gestural equivalents of the Greek alphabet characters and the digits 0-9, used mainly for visually spelling proper names and numbers to be found across all the subjects of the curriculum; and (c) a tool enabling the dynamic synthesis of sign phrases, in which a signing avatar performs GSL phrases typed in by users in real time (Efthimiou et al. 2017).

Finally, the workbench offers a series of tools enabling the use of the GUI features of the platform by deaf and HoH users. These comprise appropriately employed color code conventions and pop-up windows across the interface, help buttons of suitable shape and size, and extra help in the form of a video tooltip (Efthimiou et al. 2016).

1 The research leading to this output received funding from the POLYTROPON project (KRIPIS-GSRT, MIS: 448306) and was based on insights, technologies and language resources initially developed within the Dicta-Sign project (FP7-ICT, grant agreement n°: 231135).

2 Corpus Design and Contents

The POLYTROPON corpus is a bilingual parallel corpus for the language pair GSL-MG, which was designed with multiple applications in mind, as is often the case with parallel and comparable corpora (McEnery & Xiao 2007). More specifically, it was meant to serve as a "golden" corpus for the creation of both other lexical and terminological resources and sign language technologies, such as machine learning and machine translation from and into sign language. The creation of the corpus was based on data derived from several corpora that have been collected over the years by means of HD and Kinect cameras within the context of different projects, and it incorporates various data. The two main pre-existing resources were the list of entries of the bilingual GSL-MG multimedia dictionary project NOEMA (Efthimiou & Katsoyannou 2001) and a set of 2,000 lemmas extracted from the segmentation procedure of Dicta-Sign Corpus (Matthes et al. 2012). These two different corpora resulted in a rich set of GSL data, which includes information on signs, synonyms, compounds, classifier constructions,[2] different senses of each single lemma form, and so on.

The methodology followed to create the corpus was divided into three stages, as explicitly described in Efthimiou et al. (2018). In the first stage of the process, all lemmas found in the two pre-existing corpora (approximately 2,000), were reviewed by a working group of SL experts. During this phase, the status of each selected GSL entity had to be evaluated and validated as a simple sign, compound or classifier construction. In the second stage, examples of use for each GSL discussed lemma (classifiers excluded) were created and recorded in three repetitions by means of HD and Kinect cameras. This was mainly decided in order to make the lexicon resource exploitable in machine learning experiments targeting machine translation and sign recognition to illustrate articulation variation, as there can be no two identical video captures of the same token. This is also a key difference between sign language corpora and written language corpora, small or larger chunks of which can and will be repeated in the corpus numerous times. Moreover, the discussions among GSL experts added some 1,600 new lemmas to the corpus, which had not initially been included in the lexicon. This procedure led to the generation of new sentences which were connected to the new lemmas. In the third stage of the procedure, one out of every three recordings of the examples of use was annotated in iLex, a software tool for linguistics analysis of sign language data (Hanke & Storz 2008; Efthimiou et al. 2016). For that reason the annotation procedure of the example of use includes the following set of tiers:

- Clause: determines the clause boundaries of the signed clauses.
- Gloss: assigns an MG equivalent to each GSL identified entity.
- Greek equivalent clause: an MG translation is provided in each GSL example of use. The translation of the annotated clauses was completed in two phases: First, a strongly GSL-like translation was provided by the corpus annotator and, then, each example was reviewed in terms of naturalness and grammaticality by an expert in MG.
- Classifier: analyses classifiers occurring in the examples with respect to their morpho-phonological and semantic status. For dictionary compilation purposes, we were particularly interested in classifier constructions that combine with single signs to create compounds (Efthimiou 2010).
- S-type: marks GSL examples as main or subordinate constructions.
- S-category: marks each sentence type as declarative, negative, interrogative, imperative, or exclamatory.

For reasons of quality assurance and consistency across annotators, a peer-to-peer crosscheck procedure was followed along the lines of the earlier Dicta-Sign Corpus procedure (Dimou et al., 2011).

2 Classifiers are found in almost all SLs. They are morphemes expressed by meaningful hand configurations and often occur with verbs that express: i) motion through space, ii) the location or existence of a referent or iii) a referent that is being held (Efthimiou 2004).

This was essential during every new addition to the corpus in order to ensure that the kappa score for inter-annotator agreement was kept at a very high level.

3 Dictionary Compilation

The largest part of entries in NOEMA+ was based on an analysis of the annotated GSL corpus. As implied in the previous section, the identification of separate lemmas in such a newly developed lexical resource as a video SL corpus is far from intuitive. Lemma identification for this particular language pair proved to be a challenging task, as lexicographers had to select autonomous tokens representing separate concepts on their own. There were also frequent cases when a GSL lemma would not have a one-to-one equivalence with a lemma in MG or the other way around. This, of course, is far from surprising in bilingual lexicography, with an abundance of articles (Zgusta 1971, Snell-Hornby 1984, Piotrowski 1994) on similar problems having been published, proving that there is no one-to-one correspondence between lexical structures. As in this dictionary one of the languages presented is not an oral/written one, but rather a three-dimensional language, the problem of equivalence proved to be even more challenging.

In an attempt to get round this problem, lexicographers decided that the best solution would be to start by translating the signed examples of use first before getting to each lemma and its respective sense in the example. As expected, the lemmas were much easier to translate in the context of the examples, from which the lemma translations were then derived. For practical as well as illustrative reasons, some of the examples extracted from the corpus had to be shortened or simplified, as is the case in many lexicographic projects (Kilgarriff 2013). We soon discovered that this cannot be performed by processing an already existing caption, as this would result in a lack of naturalness in the GSL examples. This created the need for even more video captions. At the end of this process, a two-column table of equivalent sentences was created, which served as the basis for finalizing the list of lemmas/senses. Apart from classifier constructions, further items were excluded from this list based on the fact that they were not considered by lexicographers as an integral part of the core MG vocabulary. Such items were:

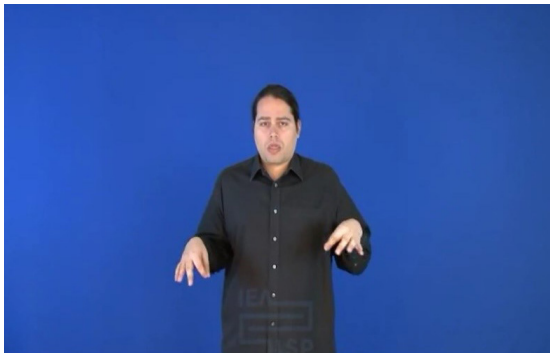

Figure 1: GSL expression “EMPTY-POCKETS” meaning “broke”.

- Signs representing brand names, football teams or other organizations, such as Ikea, Olympiacos, and Facebook.
- A set of GSL-specific expressions that have no direct equivalent in MG. For instance, the sign “EARS-DOWN”, which expresses the meaning of “obey” and that of “EMPTY-POCKETS” (Figure 1), which means “broke”.
- A set of gestures with semantic value equivalent to embodied signals occurring in oral communication in MG. These meaningful gestures add extra-linguistic information to utterances such as “I don’t know” (Figure 2) or “What can I say”, which also occur in GSL.

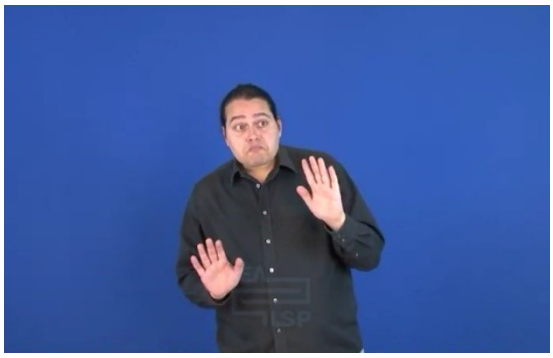

Figure 2: Embodied gestures found both in GSL and MG.

The outcome of this first compilation step was compared to existing MG dictionaries and glossaries, so that basic lemmas from different closed categories would not be left out. During this step, lexical items such as numbers, days of the week, months, seasons, units of measurement, geographical locations, etc., were included for reasons of comprehensiveness. As research on the basic vocabulary of Greek is still scarce (Vacalopoulou & Efthimiou 2015) lexicographers decided to include in it a number of lexical items that are particularly frequent in standard MG corpora. To this end, a frequency list from the Hellenic National Corpus (HNC, http://hnc.ilsp.gr/), a large, POS-tagged and lemmatized general-language MG corpus (Hatzigeorgiu et al. 2000), was cross-checked against the list of entries to locate more potential entries. However, a large number of top frequency items, among the first 2,000, had to be excluded automatically, as these typically are content-free function words, such as prepositions, articles or conjunctions, which do not correspond to any concrete sense but help in the formation of grammatical phrases and sentences in most vocal languages. Finally, extra care was taken in order to ensure that every sign (excluding those falling into the categories mentioned above) as well as every written word (excluding proper names and function words) that appear in the NOEMA+ examples is also a lemma in itself, following standard lexicographic practice. Finally, alternate orthographies of MG entries were also added to the list for reasons of comprehensiveness.

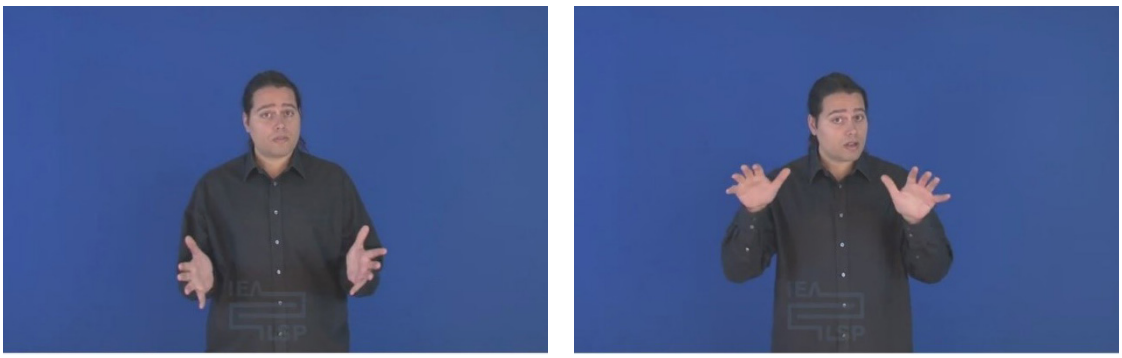

Figure 3: Alternate sign formations for the literal sense of the word “barrel”.

All these enhancement steps resulted in a new output, almost double the size of the original dictionary. As mentioned above, the new contents of the dictionary were piloted with different groups of informants, who identified possible adjustments and improvements so that the final product would have the widest possible acceptance of end users. As the source language of this bilingual dictionary is GSL, it was only natural that alternate sign formations (Figure 3) were also included at this stage by GSL experts so that there could be a balanced representation of variants in both languages. It is worth noting that this process helped, in turn, the enhancement of the original GSL corpus, as more lemmas and respective examples of use were being added.

The form of the GSL entries varies according to morphology. In particular, apart from basic, monomorphemic signs, several types of compounds are also formed as in other SLs (Brown 2010). In particular, the compounding options found as dictionary entries include the following combinations (Efthimiou et al. 2018):

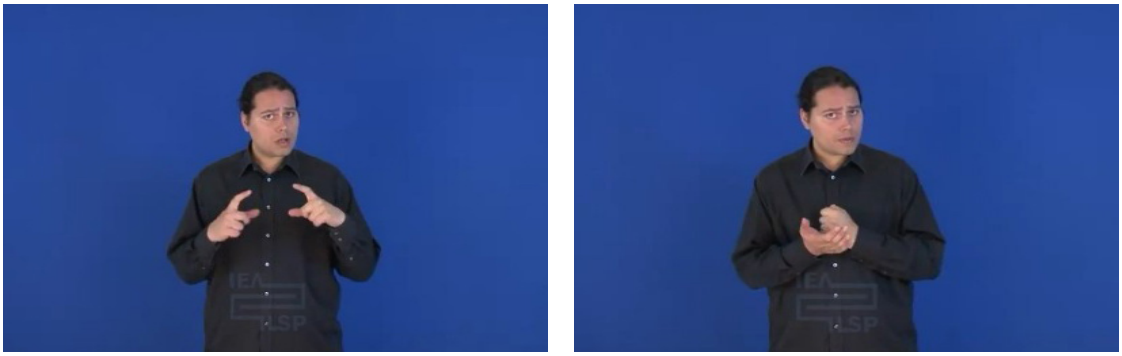

Figure 4: The compound “letter” formed by C1 (indicating shape) and the sign for “seal”.

- sign+sign: compounds such as “FLIGHT ATTENDANT” formed by two individual signs (i.e. AIRPLANE+ACCOMPANY).
- classifier+sign: compounds consisting of a classifier construction and a monomorphemic sign, such as C1+SEAL to form the equivalent of the word “letter” (Figure 4).
- classifier+classifier: compounds formed by the combination of two classifier constructions, such as C5+5 to represent the word “lighthouse” in the literal sense (Figure 5).

As far as dictionary structure is concerned, each GSL entry is accompanied by one or more MG equivalents for each sense it represents, by synonyms (if any) in either language, and by simple examples of use. When applicable, multi-word MG entries are linked to their respective single-word ones (excluding functional words) via cross-references. Apart from facilitating easy reference, this feature also has pedagogical added value, considering that most of the words which form these phrases are inflected types of other entries. It thus becomes easier for users to link each inflected type to the base form of the entry. A deliberate decision to exclude any metalinguistic information at this point was made in order to make NOEMA+ more user-friendly to the primary target audience, i.e. native signers. As the dictionary was based on the aforementioned corpus, the vast majority of the examples of use are authentic as opposed to constructed ones, whereas a small number of them were created ad hoc. As there was more than one video capture for each lemma and each example in the corpus, careful consideration was needed in order to select the best candidates for inclusion in the NOEMA+, meaning the most ‘representative’ ones. Again, the final decision for the most common and natural capture was made by groups of GSL experts, some of whom native signers.

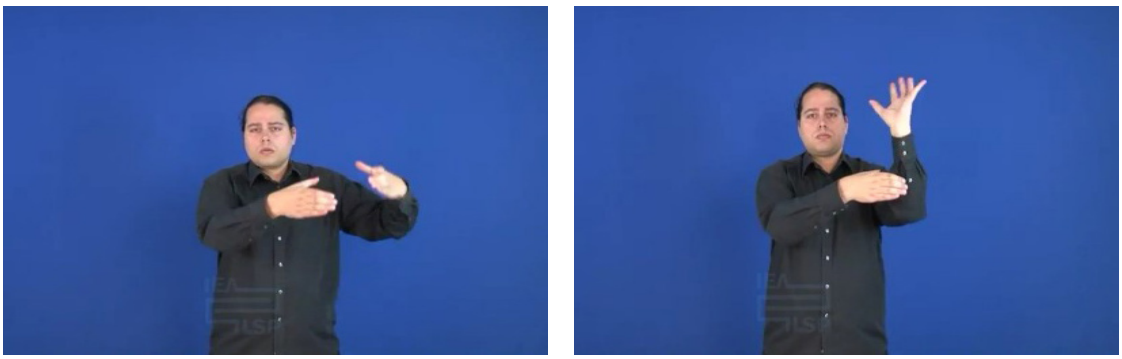

Figure 5: The compound “lighthouse” formed by two classifiers.

Finally, another convention adopted by lexicographers was to translate all GSL entries in the form they would be found as entries themselves in MG dictionaries for reasons of consistency and easy reference. Therefore, verbs appear in the first person singular present in the active voice; nouns appear in the singular nominative; adjectives and past participles appear in the nominative positive (in this case, in the masculine, feminine and neutral); adverbs appear in the positive form. As in MG dictionaries, the only exceptions occur when what is considered as the base form is either ungrammatical or particularly infrequent in the language. Indeed, this convention has proved particularly effective with regard to dictionary look-up, as explained in the next section of this paper.

4 Dictionary Look-up

One of the greatest challenges in this project relates to the presentation of the content rather than its compilation. This is based on the paradox that, although the source language of this electronic reference work is GSL, no search options are available in this language due to its three-dimensional nature. In other words, users who are native in the source language will have to perform searches in the target language. This barrier was also the reason behind several choices in the dictionary-making process. For instance, contrary to standard lexicographic practice, the addition of entries in NOEMA+ was made taking into account both the source language and the search language, which are not identical.

Although great care was taken to double-check every entry both ways, it was obvious that users, be they either beginner learners or regular users of GSL, would need to be presented with several alternative options, in order to make successful searches. Therefore, they are provided with three choices (Figure 6), i.e. type in search items, use a virtual fingerspelling keyboard, or select their search item from an alphabetical list of entries.

Figure 6: Looking up a word in NOEMA+.

5 Summary and Results

In this paper we present lexicographic work targeted at the development of a bilingual dictionary of approximately 12,000 entries for the pair GSL-MG. This reference work was initially developed as part of an official educational content platform of the Greek Ministry of Education offering open access to its content by deaf and HoH users in their native language. In addition to that, it has now grown to be a much richer dictionary, offered to end users as a standalone service.

The processes of dictionary compilation as well as the way users look up lexical items were particularly challenging, based on the paradox of the source language not coinciding with the search

language. We found, however, that apart from posing practical obstacles, this reality enabled us to apply two-way checks from source to target language and vice versa; not only did this help us avoid errors and misunderstandings (given that a large part of the GSL content had never been recorded in a dictionary before), but it also resulted in easier and quicker enhancement of the contents of both the dictionary and GSL corpus. In fact, because of the limited lexicographic work in GSL, we dare say that NOEMA+ is one among few such projects that cannot escape being corpus-driven. In the words of Tognini-Bonelli (2001: 84), the "corpus […] is seen as more than a repository of examples to back pre-existing theories […]."

References

Breadmore, H.L. (2008). Inflectional morphology in the literacy of deaf children, (PhD thesis, University of Birmingham, Birmingham, UK). Accessed at: http://etheses.bham.ac.uk/591/ [25/03/2018].

Brown, K. & Ogilvie, S. (eds.) (2006) *Concise Encyclopedia of Languages of the World*. P. 324. Elsevier, Amsterdam. 14 vols.

Crasborn, O. & Zwitserlood, I. (2008). The Corpus NGT: an online corpus for professionals and laymen. In O. A. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, E. Thoutenhooft (eds.) *Construction and Exploitation of Sign Language Corpora. (pp. 44-49)*. ELDA, Paris.

CRPD – The Convention on the Rights of Persons with Disabilities. Accessed at: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html [26/03/2018]

Dimou, A.L., Goulas T. & Efthimiou E. (2011) Grammar/Prosody Modelling in Greek Sing Language: Towards the definition of built-in sign synthesis rules. In E. Efthimiou, G. Kouroupetroglou, S.E. Fotinea (eds.) *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, vol.7206, pp. 183-193. Berlin, Heidelberg: Springer LNCS 7206.

Efthimiou, E., Vasilaki, K., Fotinea, SE., Vacalopoulou, A., Goulas, T. & Dimou, A.L. (2018) The POLYTROPON Parallel Corpus. In proceedings of the *8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community at LREC2018*, Miyazaki [to be published].

Efthimiou, E., Fotinea, S.E., Kakoulidis, P., Goulas, T., Dimou, A.L. & Vacalopoulou, A. (2017) Sign Search and Sign Synthesis Made Easy to End User: The Paradigm of Building a SL Oriented Interface for Accessing and Managing Educational Content. In: Antona M., Stephanidis C. (eds.) Universal Access in Human-Computer Interaction. Designing Novel Interactions. UAHCI 2017. Lecture Notes in Computer Science, vol. 10278. Springer, Cham.

Efthimiou, E., Fotinea, S.E., Goulas, T., Kakoulidis, P., Dimou, A.L. & Vacalopoulou, A. (2016) A Complete Environment for Deaf Learner Support in the Context of Mainstream Education. In *Proceedings of the Conference Universal Learning Design*, *vol. 5*. ISSN 1805-3947 Linz, 13-15 July, pp. 35-44.

Efthimiou, E., Fotinea, SE., Dimou, AL. & Kalimeris, C. (2010) Towards decoding Classifier function in GSL. *In the Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Satellite workshop of the LREC-2010 Conference, pp.76-79*, Valetta.

Efthimiou, E. & Katsoyannou M. (2001) Research issues on GSL: a study of vocabulary and lexicon creation. In *Studies in Greek Linguistics, Computational Linguistics 2:42-50* (in Greek).

European Disability Strategy 2010-2020. Accessed at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0636:FIN:en:PDF [26/03/2018].

Fenlon, J., Cormier, K.A., Rentelis, R, Schembri, A., Rowley, K., Adam. R. & Woll, B. (2014) BSL SignBank: A lexical database and dictionary of British Sign Language (1st Edition). Accessed at: http://bslsignbank.ucl.ac.uk/ [26/03/2018].

Fuertes, J.L., González, Á.L., Mariscal, G. & Ruiz, C. (2006) Bilingual Sign Language Dictionary. In K. Miesenberger, J. Klaus, W.L. Zagler, A.I. Karshmer (eds.) *Computers Helping People with Special Needs. ICCHP 2006. Lecture Notes in Computer Science, vol 4061*. Springer, Berlin, Heidelberg

Hanke, T. & Storz, J. (2008). "iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography". In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood & E. Thoutenhooft (eds.) *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*, pp. 64-67 ELRA, Paris.

Hatzigeorgiu, N., Gavrilidou, M., Piperidis, S., Carayannis, G., Papakostopoulou, A., Spiliotopoulou, A., Vacalopoulou, A., Labropoulou, P., Mantzari, E., Papageorgiou, H. & Demiros, I. (2000). Design and implementation of the online ILSP Greek Corpus. In *Proceedings of the second international conference of language resources and evaluation (LREC) (vol. 3, pp. 1737-1740)*, Athens.

Holton, D., Mackridge P. & Philippaki-Warburton I. (1997). *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.

Kilgarriff, A. (2013) Using corpora [and the web] as data sources for dictionaries. In H. Jackson (ed.) *The Bloomsbury Companion to Lexicography*, pp. 77-96. Bloomsbury Publishing, London.

Konrad, R. & Langer, G. (2009) Synergies between transcription and lexical database building: The case of German Sign Language (DGS). In M. Mahlberg, V. González-Díaz, C. Smith (eds.). *Proceedings of the Corpus Linguistics Conference (CL2009)*. University of Liverpool, UK, 20-23 July 2009. Accessed at: http://ucrel.lancs.ac.uk/publications/cl2009/#papers;%20Article%20#346 [25/03/2018].

Kristoffersen, J.H. & Troelsgård T. (2010). In Dykstra, A. & Schoonheim T. (eds.) *Proceedings of the 14th EURALEX International Congress, pp. 1549-1554*. Fryske Akademy, Leeuwarden/Ljouwert.

Linde-Usiekniewicz J., Czajkowska-Kisil M., Łacheta J. & Rutkowski P. (2014) A corpus-based dictionary of Polish Sign Language (PJM). In A. Abel, C. Vettori, N. Ralli (eds.) *Proceedings of the XVI EURALEX International Congress: The user in focus*, pp. 365-376. EURAC research, Bolzano/Bozen.

Matthes, S., Hanke, T., Regen, A., Storz, J., Worseck, S., Efthimiou, E., Dimou, A.L., Braffort, A., Glauert, J. & Safar, E. (2012). Dicta-Sign – Building a Multilingual Sign Language Corpus. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon (LREC 2012)*, Istanbul.

McEnery A.M. & Xiao R.Z. (2007). Parallel and comparable corpora: What are they up to? In Incorporating Corpora: Translation and the Linguist. Translating Europe. Multilingual Matters, Clevedon, UK.

Mesch J. & Wallin L. (2015) Gloss annotations in the Swedish Sign Language corpus. In *International Journal of Corpus Linguistics*, 20(1), 102–120.

Piotrowski T. 1994. *Problems in Bilingual Lexicography*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Prokopidis, P., Georgandopoulos, B. & Papageorgiou, Ch. (2011) A suite of NLP tools for Greek. In *The 10th International Conference of Greek Linguistics*. Komotini, Greece.

Schembri, A. (2008). British Sign Language corpus project: open access archives and the observer's paradox. *In the 6th International Conference on Language Resources and Evaluation, LREC, May 26-June 1, 2008*, Marrakech.

Snell-Hornby, M. (1986) The bilingual dictionary – Victim of its own tradition? In R.R.K. Harmann (ed.) *The History of Lexicography*. John Benjamins, Amsterdam.

Tognini-Bonelli, E. (2001) The corpus-driven approach. In W. Teubert, R. Krishnamurthy (eds.) *Corpus Linguistics at Work*. John Benjamins, Amsterdam/Philadelphia.

Vacalopoulou, A. & Efthimiou, E. (2015). Multilingual lexicography for adult immigrant groups: bringing strange bedfellows together. In I. Kosem, M. Jakubíček, J. Kallas, S. Krek (eds.) *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference, Herstmonceux Castle, 11-13 August 2015*. Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., Ljubljana/Brighton.

Vettori, C. & Felice, M. (2008) e-LIS: Electronic Bilingual Dictionary Italian Sign Language-Italian. In E. Bernal, J. DeCesaris (eds.) *Proceedings of the XIII EURALEX International Congress*, pp. 791-796. Universitat Pompeu Fabra, Barcelona.

Zgusta, L. (1971) *Manual of Lexicography*. Janua Linguarum. Series Maior 39. Mouton, The Hague/Paris.