



EURALEX XIX
Congress of the
European Association
for Lexicography

Lexicography for inclusion

7-11 September 2021
Ramada Plaza Thraki
Alexandroupolis, Greece

www.euralex2020.gr

**Proceedings Book
Volume 1**

Edited by Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

EURALEX Proceedings

ISSN 2521-7100

ISBN 978-618-85138-1-5

Edited by: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

2020 Edition

Sign Language Corpora and Dictionaries: a Multidimensional Challenge

Vacalopoulou A.

ILSP-Institute for Language and Speech Processing / Athena RC

Abstract

This paper is an analysis of the main challenges in developing sign language resources such as corpora and dictionaries. Although difficulties in data collection and processing are common with those in similar projects for vocal languages, there are extra complications that seem to be unique to the creation of resources for sign languages. These, more language specific, problems could be categorised under three general headings: (a) linguistic obstacles, (b) financial obstacles, and (c) social obstacles. Most of the challenges in studying and describing any sign language spring from the nature of these languages themselves, which is why this nature is briefly described. Instead of dealing with the typical two-dimensional, linear representation of the linguistic message, researchers have to cope with a more complex and dynamic medium involving elements including hand position and movement, eye gaze, facial expression as well as head and body movement. All these, among others, make the acquisition and processing of signed material more expensive and time-consuming. Finally, the activity of building and exploiting sign language resources can also be held back by social factors, including choice of informants, communication barriers and prejudice.

Keywords: sign language lexicography; multimodal lexicography; sign language corpora; sign language resources; Greek Sign Language.

1 Introduction

As anyone who has ever contributed to the making of any dictionary knows, general and more detailed typological issues dictate both the content and the form of lexicographic products. As dictionary typology is among the fundamentals that guide a lexicographer's work, standard works on lexicography never fail to dedicate short or more extensive descriptions of it and how it affects dictionary writing (Zgusta 1971: 198-221; Béjoint 2000: 32-41; Hartmann 2001: 57-74; Atkins & Rundell 2008: 24-43). Although most lexicographers tend to specialise in a particular type of dictionary, they sometimes find themselves involved in very diverse projects. These can be commercial or academic, print or electronic, offline or online, purely linguistic or more encyclopaedic, monolingual or bilingual and multilingual, diachronic or synchronic, general or specialised, intended for decoding or encoding, short glossaries or multi-volume works, written for native speakers or targeting learners of the language. Conscientious lexicographers who are faced with a new type of project tend to research different aspects of the anticipated product and its end users to adjust their craft accordingly. Nothing, however, can fully prepare a lexicographer for the challenges of compiling, for the first time, a reference work involving sign language.

Based on their academic background and/or experience, most dictionary compilers would tend to assume that this is yet another bilingual project and would try to approach it in such a way. In many ways, any bilingual reference work can be more perplexing than a monolingual one simply because of the need to study more than one language at the same time (Lew 2013: 289). As a result, bilingual lexicography involves not only the extra element of comparison but also the collaboration between at least two native speakers of different languages. Nevertheless, one discovers that awareness of classic pitfalls of bilingual dictionaries in theory and practice is not enough to provide solutions to the problems that occur in sign language lexicography.

To a great extent, this is due to historical reasons as sign language lexicography is a relatively new discipline worldwide (Schermer 2006: 321; McKee & Vale 2017: 6-7), which leaves several aspects yet unstudied. However, as others have shown (Zwitserslood 2010: 444-445) a great part of the challenge lies in the nature of sign languages themselves. The aim of this paper is to list and categorise different challenges involved in the design and creation of sign language resources based on twenty years of professional involvement in the field as well as on testimonies by researchers with similar experience in the hope that researchers who are about to embark on similar ventures gain some perspective on the subject. In the following section, a brief description of the characteristics of sign languages is provided in relation to those of spoken languages so that the reader can have an overview of the nature of sign languages. Next, an account of the different challenges involved in the development of sign language resources is given; these challenges are classified here under three general headings: linguistic, financial, and social issues. The paper closes with a recapitulation of the points mentioned.

2 Spoken and Sign Languages: Similarities and Differences

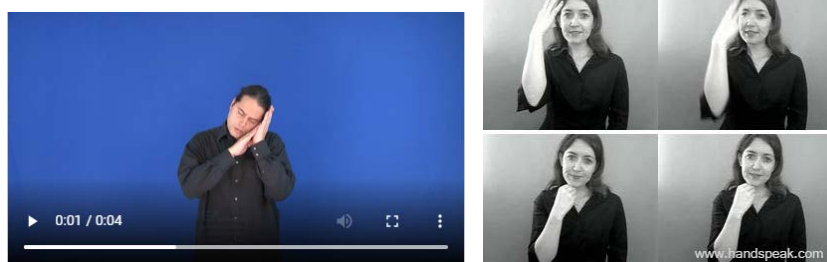
There have been quite a few misconceptions regarding the nature of sign languages. Based on stereotyped notions, people often tend to expect any language system to be similar to the one (or the ones) they themselves recognise and use. As a result, the mere fact that utterances in sign language are not formed through speech but through signing has led some to think that sign language is less of a language (Armstrong & Karchmer 2002; Zwitserslood 2010: 444; Wilcox & Occhino 2016: 1) more similar to fabricated language (Zwitserslood 2010: 444; Wilcox & Occhino 2016: 2; Vale 2017: 14), a visual

interpretation of spoken language (Wilcox & Occhino 2016: 2; Vale 2017: 14, Yule 2020: 230), or simply mime (Zwitserslood 2010: 444; Vale 2017: 14, 18). Another popular myth that still seems quite prevalent is that only one sign language exists and is shared by deaf people universally (Crystal 1992; Wilcox & Occhino 2016: 3; Vale 2017: 14). Although all these ideas of sign language are still common fallacies among the general public, sign language linguistics has been describing and researching very real and distinct sign languages since the mid-1960s. At the same time, sign language lexicography emerged, the first notable example of which was the celebrated *Dictionary of American Sign Language on Linguistic Principles* (Stokoe, Casterline & Croneberg, 1965). Although there had already been attempts to create glossaries and dictionaries to serve the needs of specific communities, the ground-breaking element of this reference work was that, for the first time, it arranged a sign language through a phonological system instead of alphabetised glosses (McKee & Vale 2017: 6). Eventually, research has established the linguistic status of different sign languages offering new insights to the ways in which they develop and operate in use. Such findings help lexicographers – among other sign language researchers – to decipher the true nature of sign languages and challenge stereotypical views of sign languages by understanding what makes spoken and signed languages similar and what makes them different. The fundamental similarity among all signed languages is that they are natural, that is, they are spontaneously created by members of a community to serve their communicative needs as opposed to “the artificially constructed systems used to expound a conceptual area (e.g. ‘formal’, ‘logical’, ‘computer’ languages) or to facilitate communication (e.g. Esperanto)” (Crystal 2008: 265). Consequently, every sign language is the creation of a specific deaf community, the members of which also share a common culture and are native signers of that language; as any other similar system – it can also be learnt by non-members of that community. Each of these languages is distinct in terms of lexical, morphological, syntactic and semantic aspects (Mayberry & Squires 2006: 291; Wilcox & Occhino 2016: 3), which do not directly correspond to with those of other spoken or signed languages.



Figure 1: The sign for *umbrella* in GSL (NOEMA+).

On the other hand, there is a vital difference between spoken and signed languages, which relates to their modality, i.e. the fact that they are visual-gestural as opposed to oral-aural (Zwitserslood 2010: 457; McKee & Vale 2017: 2). In other words, instead of being linear, the structure of the language is multidimensional as it is produced, perceived and understood in space. Signs (which give these types of languages their name) are the building blocks of communication in the sense that they usually convey the intended meaning. Signs, however, cannot be taken to have a one-to-one correspondence to words as, instead of sounds and syllables, they consist of different elements, which often carry some meaning themselves (Johnston & Schembri 1999: 117-118): (a) handshapes, i.e. the specific shape formed by one or both hands, (b) hand position, i.e. where hands are located, e.g. in front of the body or next to it, (c) hand movement, i.e. the way in which the hands move, (d) hand orientation, i.e. the direction in which the hands are placed, e.g. fingers facing the body, (e) non-manual elements in the face and other body parts apart from the hands, e.g. facial expressions or head tilting. Figure 1 shows a video still of the sign for *ομπρέλα umbrella* in Greek Sign Language (GSL), exemplifying some of these components. The sign starts by both hands shaped in fists (handshape) facing inwards (orientation) and placed in front of the body (position), followed by the top hand moving upwards (movement) as if opening an umbrella. An example of a non-manual element can be seen in Figure 2 (left), where the head tilts towards the hands to represent *sleep*



in GSL.

Figure 2: The sign for *sleep* in GSL (left, NOEMA+) and in ASL (HandSpeak).

A direct consequence of the visual-gestural modality is that the articulation of the abovementioned elements happens not only sequentially but sometimes also simultaneously, marking another difference between spoken and signed languages (Sandler 2006: 336). A second obvious element springing from this modality is the fact that a lot of signs seem to be

characterised by iconicity in demonstrating meaning, as shown in the GSL example of *umbrella* in Figure 1. In fact, as Taub (2001) has argued, it is this iconicity that has misled some into thinking that sign languages are universal in nature rather than separate arbitrary systems (as cited in Sandler 2006: 336). In fact, there is a lot of arbitrariness in sign languages not only concerning signs that represent things that cannot be demonstrated in such a way (such as abstract concepts) but in those corresponding to concrete things as well. This can be demonstrated by the fact that the same concrete thing can be represented by very different signs across sign languages. An example would be the representation of *sleep* in two different languages, GSL and American Sign Language (ASL) shown in Figure 2. In the video still, *sleep* is represented in GSL by putting one hand on top of the other, holding them both next to one side of the head and then closing the eyes and bending the head onto the hands as if they were a pillow. On the other hand, in the printable version in ASL *sleep* is signed by opening one hand in front of the face with the palm facing the face, then moving the hand towards the chin while joining the fingers together and touching them with the thumb.

3 Issues for Sign Language Lexicography

After this general description of the nature of sign languages, which springs from their modality, an attempt is made to analyse and categorise a series of the challenges involved in sign language lexicography partially drawing on a research group's experience in various Greek sign language lexicography projects. These projects include the development of relevant resources, such as the capturing of Greek Sign Language material in video, the annotation of the respective corpora as well as the design and development of various GSL dictionaries (Efthimiou et al. 2004; Efthimiou et al. 2017; Efthimiou et al. 2018; Vacalopoulou, Efthimiou & Vasilaki 2018; Vacalopoulou et al. 2018). The analysis will only refer to general challenges deriving from the multimodal nature of sign languages rather than more technical difficulties relating to the lexicographic treatment of specific signs of parts of signs, how detailed the information for each sign should be, how lemmas are selected or organised, etc.¹

3.1 Linguistic Issues

In order to represent any sign language and process it computationally, several transcription systems have been developed depending on project-specific use and needs. Among the most popular conventions internationally is the representation of signs using word glosses; in cases when one word is not enough to describe the respective sign, more words are included, typically joined together by hyphens. Glossing, which is a very helpful technique for alphabetising sign lists, has been widely used in the description of sign languages in terms of their morphology, syntax, and discourse (Miller 2006: 353) and it remains a standard way of sign transcription to this day.

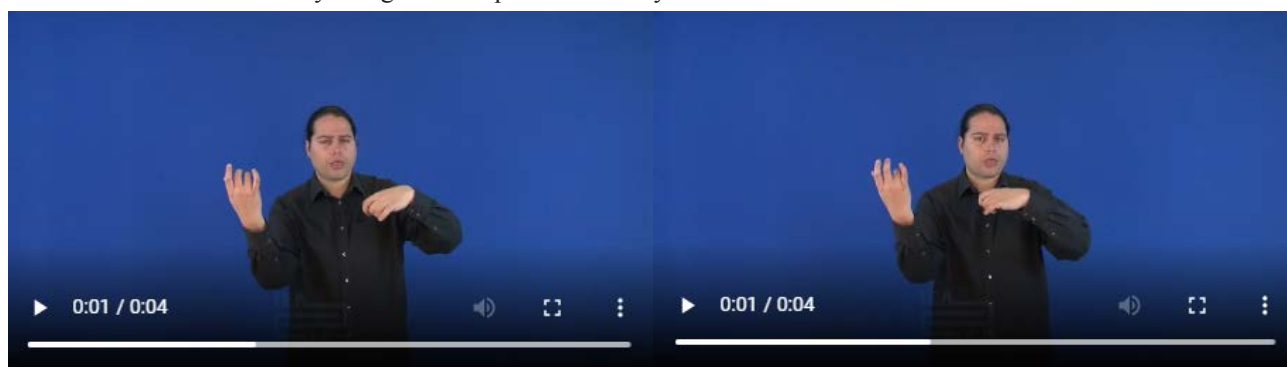


Figure 3: Video stills of the sign for *persuade* in GSL (NOEMA+).

This type of representation, however, poses several challenges, among which the fact that not all glosses include the full range of meanings carried by a sign. For instance, relevant research in Australian Sign Language has shown that the glosses OPEN-WINDOW and SHUT-WINDOW can be misleading as the same signs are generically used to represent the opening and shutting of any flat-surfaced thing (Johnston 2001: 251). Another problem results from cases of ambiguity as to what exact glossing should be used when the meaning of the sign is not straightforward; in other words, when the referent is not easily translatable to the respective oral language (Mesch & Wallin 2008: 135). In such occasions, the equivalent of a paraphrase is used in the gloss, which comprises more than one words. An example would be one of the signs for the verb *to persuade* in the specialised sense of “charming or humouring someone in order to convince them about something”. The GSL sign for this (Figure 3) is formed by playing an invisible fiddle while pushing out one's slightly open lips hinting to the charm element behind this action, which is depicted in the gloss “CONVINCE-DIPLOMACY”. In this case of polysemy, the same sign is, therefore, used to denote both *persuade* and *diplomacy*.

In addition, there are several spatial verbs – including verbs of movement and location – that tend to be represented in more than one translational equivalent in sign languages due to the fact that the type of movement linked to each of them is visually different depending on the context. As a result, there will be different signs in a language for the verb *to close* depending on what closes or what is being closed: “close the window”, “close one's eyes”, “close the curtains”, “close the

¹ Attempting to further expand on these or on the description of specific projects would exceed the scope of this study. For an overview of most of these issues, as well as an account of current sign language lexicographic practices, see Zwitserlood 2010.

Figure 6: Classifier annotation by semantic function in *iLex*.

Notation systems are typically used in combination with specialised software to facilitate the computational recording of the phonological aspects of sign languages. An example is shown in Figure 6, where a sentence in GSL is broken down in parts for annotation in the *iLex* software for transcribing and linking signed corpora and dictionaries (Hanke & Storz 2008). Nevertheless, as extensive and detailed as most of these systems are, they are not always successful in fully grasping signs, as sign boundaries are seldom clear-cut in real utterances when there is considerable overlapping (Herrmann 2008: 69). A central issue here is the time parameter, that is, the fact that, apart from sequential articulation, there is also simultaneous articulation. The problem of segmentation, of course, is not unique to sign language processing, as boundaries between words are also often difficult to set in continuous speech (Creer & Thompson 2004; Himmelmann 2006). Indeed, signed and spoken language share a lot of similarities, such as being largely spontaneous and informal. This is why reliance on signed corpora for the description of a sign language could be compared in terms of difficulty to only having spoken corpora to describe a spoken language. Unit segmentation, as many more of the challenges in signed language linguistics, has been linked to the relatively new emergence of this scientific domain (Álvarez Sánchez, Báez Montero & Fernández 2008: 10), a perspective that brings hope for future solutions through relevant research.

3.2 Financial Issues

All sign languages are minority languages. A key obstacle in researching minority languages as well as collecting and creating resources is that they are not always financially supported from official organizations as the respective audience is not expected to be large enough (Ivanova 2010: 125; Vale 2017: 3). As a result, the number of available resources for most sign languages is still limited; this is particularly true for signed corpora that involve a considerable investment in terms of both money and time in order to be suitably collected and annotated (Crasborn & Zwitserlood 2008: 49; Naert et al. 2018: 139; Bragg et al. 2019: 19, 25). Such an investment mainly includes two types of expenses: those relating to the acquisition of signed material and those involving its processing.

Costs relevant to material acquisition are connected with the process of obtaining the appropriate equipment for the recording and processing of sign language utterances. This is usually done in video recordings and typically includes various types of cameras and other sensors. In order to fully grasp the multidimensional nature of sign messages, researchers tend to use high-resolution, multiple and/or depth cameras, a choice that increases the overall cost. When other types of sensors are used, body suits that capture motion will also need to be bought (Jedlička, Krňoul & Železný 2006: 102; Kanis & Krňoul 2008: 88; Bragg et al. 2019: 24). In addition, the recording of authentic videos is very often facilitated by sign language interpreters, another costly addition to the overall budget. The high cost of collecting sign language material has led some researchers to more inexpensive options such as using existing material from online sources or involving outside users in crowdsourcing platforms. However cheap, these possibilities do not come without drawbacks including quality control issues and the lack of appropriate annotation (Bragg et al. 2019: 24). Apart from acquisition, there are financial issues to consider in terms of training annotators on the use of relevant technologies, a process which usually is also time-consuming (op. cit: 21). Part of these costs could be decreased, however, if a standard system for annotating sign languages was to be adopted (Bragg et al. 2019: 25); indeed, the necessity for having standardised any collection of data intended for lexicographic use has been acknowledged by professional lexicographers (Atkins & Rundell 2008: 84). Reduction of costs is also one of the reasons why research has been aiming at the direction of automating the entire annotation process as much as possible (Meurant 2016).

3.3 Social Issues

It is only in recent years that sign languages, though not (yet) all of them, have been granted language status (Kristoffersen & Troelsgård 2012: 294). GSL, for that matter, was officially recognised as a language in 2000 (Timmermans 2005: 104) but had not been granted equal status with Modern Greek until 2017 (HFD 2017). Consequently, and given the various types of prejudice mentioned earlier, it comes as no surprise that research in the field of sign languages is a newcomer in linguistics. The social parameter has, therefore, been the main constraint for the shortage of signed resources. Indeed, this marginalisation has made deaf communities of the world more or less sceptical towards endeavours initiated by hearing people or organizations. Jones (2002: 56), for instance, reports examples of prejudice against professionals in the wider field of deafness who are not deaf themselves as mentioned by Lane back in the early 1990s. Given this tendency, reluctance to participate in such projects is not a rare phenomenon in sign language research.

As already mentioned, sign language lexicographic projects are more complex than any ordinary bilingual project. First, although not always the case, it is considered best practice among lexicographers to involve native speakers of both languages in the compilation of bilingual dictionaries (Atkins & Rundell 2008: 102; Stamper 2012). The activity of building and exploiting sign language resources can sometimes be held back by social factors as, in the context of GSL, linguists who are also native signers are scarce and there are hardly any lexicographers around sharing the same background. This is a reality for most sign languages, the users of which are linguistic minority groups within much broader communities. As a result, sign language lexicography, has been following, along with minority language lexicography, an inevitable tradition of resources compiled and processed by non-native users (Chelliah & de Reuse 2011: 56; Cristinoi & Nemo 2013). Thus, sign language lexicographers unavoidably rely on native informants to ensure that their attempt to describe the language is accurate and up to date, as no pre-existing sign language corpora are readily available. Selecting the right informants for each project can be a complex procedure involving a series of different criteria (Langer et al 2018: 492), some of which are listed in Álvarez Sánchez, Báez Montero & Fernández (2008: 10):

Social background: “place and date of birth, age of deafness occurrence, deafness degree, deaf/hearing family, job of closest family members”; Education: “degree and type of studies, special/ordinary school, use/absence of SL in school”; “Linguistic skills”: in the research sign language, oral language, lip-reading, written language. In addition, any balanced selection of informants would include both men and women from various age groups. Given the fact that the population in question belong to a linguistic and cultural minority, it is evident that a well-adjusted selection of informants is a very demanding task. If this is seen in combination with the bias against hearing professionals, it is evident that the task borders on the impossible.

Furthermore, it is considered good practice that every item intended for inclusion in either a corpus (when this is not a spontaneous one) or a dictionary be reviewed by more than one informants so as to ensure that the actual meaning of the utterance is established. This process, however, is not without challenges, as no native signer consensus is established for a large number of issues in most sign languages (Johnston 2008: 82; Chen Pichler et al. 2016: 31). In fact, diversity among native signers is significant and relates to various factors such as “ethnicity, geographic region, age, gender, education, language proficiency, hearing status, etc.” (Bragg et al. 2019: 18). The fact that there are added parameters (such as hearing status) influencing diversity in sign languages combined with the scarcity of relevant research make the prospect of reaching consensus even more distant.

As if forming a balanced set of informants and trying to reach consensus among them is not enough of a challenge already, it has been noted that lack of formal teaching of sign languages to native signers may result to lack of linguistic conscience among the group (Álvarez Sánchez, Báez Montero & Fernández 2008: 11). Whatever the reason, practice has shown that it is often complicated to present informants with a set of glosses and ask them to represent them in sign. Indeed, in several occasions, we have found that abstracting the actual meaning or use of specific GSL lexical items can be quite difficult for informants who tend to concentrate on the glosses or words presented to them instead. This misleading one-to-one correspondence has often led informants to claim that several items “do not exist” in GSL, only to discover – along with researchers – that they very much exist, when informants are prompted to use them in context in actual GSL conversation. This one of the (several) reasons why there is an increasing tendency for more authentic signed resources as well as for the inclusion of authentic examples in sign language dictionaries (Langer et al 2018; Mesch & Schönström 2018: 121).

4 Recapitulation

Much like the nature of sign languages itself, the challenge of creating signed resources is a multidimensional one. For the adventurous linguists and lexicographers who get involved in related projects, this means that three types of issues will occasionally get in the way: linguistic, financial, and social. This paper attempted to describe and classify most of them in a way that is meaningful to researchers (about to be) involved in the design and creation of sign language resources. In the near future, technical issues such as standardising annotation systems and further automating the transcription process are expected to significantly lower the now high cost for acquiring and processing signed data. The availability of more authentic signed material will hopefully result not only in more accurate representations of these languages but in some that are more generally embraced by the deaf communities. As this field of research grows, most challenges relating to its recent emergence will no doubt start becoming milder and easier to meet.

5 References

- Álvarez Sánchez, P., Báez Montero, I.C., & Fernández Soneira, A. (2008). Linguistic, sociological and technical difficulties in the development of a Spanish Sign Language (LSE) corpus. In LREC 2008 Workshop Proceedings: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. Marrakech, 1 June 2008, pp. 9-12.
- Armstrong, D., Karchmer, M. (2002). William C. Stokoe and the study of signed languages. In D. Armstrong, M. Karchmer, & J. Van Cleve (eds.), *The Study of Signed Languages: Essays in honor of William C. Stokoe*, Washington, D.C., Gallaudet University Press, pp. xi–xix.
- Atkins, S.B.T., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Béjoint, H. (2000). *Modern Lexicography: An Introduction*. Oxford: Oxford University Press.
- Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., Morris, M. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. The 21st International ACM SIGACCESS Conference on Computers and Accessibility, Oct 2019, Pittsburgh, pp. 16-31.
- Chelliah, S.L., de Reuse, W.J. (2011). *Handbook of descriptive linguistic fieldwork*. Dordrecht: Springer.
- Crasborn, O., Zwitserlood, I. (2008). The Corpus NGT: an online corpus for professionals and laymen. In LREC 2008 Workshop Proceedings: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. Marrakech, 1 June 2008, pp. 44-49.
- Creer, S., Thompson, P. (2004). Processing spoken language data: the BASE experience. In LREC 2004 Workshop Proceedings on Compiling and Processing Spoken Language Corpora. Lisboa, 24 May 2004, pp. 20-27.
- Cristinoi, A. Nemo., F. (2013). Challenges in endangered language lexicography. In *Lexicography and Dictionaries in the Information Age: Selected Papers from the 8th ASIALEX International Conference*, Bali, 20-22 August 2013. Denpasar: Airlangga University Press, pp. 126-132.
- Crystal, D. (1992). Sign Language. In T. McArthur (ed.), *The Oxford Companion to the English Language*, Oxford: Oxford University Press, pp. 935.
- Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*. Malden: Blackwell.

- 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 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 HCI 2017 Proceedings: 11th International Conference on Human-Computer Interaction. Vancouver, 9-14 July 2017, pp. 14-26.
- Efthimiou, E., Fotinea, Vacalopoulou, A., Goulas, T., Vasilaki, K., Dimou, A.L. (2018). Sign language technologies in view of future internet accessibility services. In PETRA '18 Proceedings: 11th Pervasive Technologies Related to Assistive Environments Conference. Corfu, 26-29 June 2018, pp. 495-501.
- Efthimiou, E., Vacalopoulou, A., Fotinea, S.E., Stainhaouer, G. (2004). Multipurpose design and creation of GSL dictionaries. In LREC 2004 Workshop Proceedings: 3rd Workshop on the Representation and Processing of Sign Languages: From Sign Writing to Image Processing. Lisbon, 30 May 2008, pp. 51-58.
- HandSpeak: American Sign Language Online*. Accessed at: //www.handspeak.com/ [27/04/2020].
- Hanke, T., Storz, J. (2008). "iLex – A database tool for integrating sign language corpus linguistics and sign language lexicography". In LREC 2008 Workshop Proceedings: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. Marrakech, 1 June 2008, pp. 64-67.
- Hartmann, R.R.K. (2001). *Teaching and Researching Lexicography*. Harlow: Longman-Pearson.
- Herrmann, A. (2008). Sign language corpora and the problems with ELAN and the ECHO annotation conventions. In LREC 2008 Workshop Proceedings: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. Marrakech, 1 June 2008, pp. 68-73.
- HFD (Hellenic Federation of the Deaf). (2017, September 7). *Recognition of the Greek Sign Language as Equal to the Greek Language* [Press release]. Accessed at <https://bda.org.uk/wp-content/uploads/2017/09/906-2017-PRESS-RELEASE-Recognition-of-the-Greek-Sign-Language-as-Equal-to-the-Greek-Language.pdf> [30/04/2020].
- Himmelman, N.P. (2006). The challenges of segmenting spoken language. In J. Gippert, N.P. Himmelmann, & U. Mosel (eds.) *Essentials of Language Documentation*. Berlin: De Gruyter Mouton, pp. 253-274.
- Ivanova, N. (2010). The Icelandic sign language dictionary project: some theoretical issues. In LREC 2010 Workshop Proceedings: 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. Valletta, 1 May 2010, pp. 125-128.
- Jedlička, P., Krňoul, Z. & Železný, M. Methods for recognizing interesting events within sign language motion capture data. In 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining. Portorož, 28 May 2016, pp. 101-104.
- Johnston, T. (2001). Nouns and verbs in Australian sign language: an open and shut case? *Journal of Deaf Studies and Deaf Education*, 6(4), pp. 235-257.
- Johnston, T. (2008). Corpus linguistics and signed languages: no lemmata, no corpus. In LREC 2008 Workshop Proceedings: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. Marrakech, 1 June 2008, pp. 82-87.
- Johnston, T., Schembri, A. (1999). On defining Lexeme in a signed language. *Sign Language and Linguistics*, 2(2), 115-185.
- Jones, M.A. (2002). Deafness as culture: a psychosocial perspective. In *Disability Studies Quarterly* 22(2), pp. 51-60.
- Kanis, J. Krňoul, Z. (2008). Interactive HamNoSys notation editor for signed speech annotation. In LREC 2008 Workshop Proceedings: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. Marrakech, 1 June 2008, pp. 88-93.
- Kristoffersen, J.H., Troelsgård, T. (2012). The electronic lexicographical treatment of sign languages: The Danish Sign Language Dictionary. In S. Granger, M. Paquot (eds.) *Electronic Lexicography*. Oxford: Oxford University Press, pp. 293-318.
- Langer, G., Müller, A., Wähl, S., & Bleicken, J. Authentic examples in a corpus-based sign language dictionary - why and how. In XVIII EURALEX International Congress: Lexicography in Global Contexts. Ljubljana, 17-21 July 2018, pp. 483-497.
- Lew, R. (2013). Identifying, ordering and defining senses. In J. Howard (ed.). *The Bloomsbury Companion to Lexicography*. London: Bloomsbury, pp. 284-302.
- Mayberry, R.I., Squires, B. (2006). Sign Language: Acquisition. In K. Brown (ed.) *Encyclopedia of Language and Linguistics* (2nd ed.), 291-296. Amsterdam: Elsevier. Accessed at: <https://www.sciencedirect.com/science/article/pii/B0080448542008543> [27/04/2020].
- McKee, R., Vale, M. (2017). Sign Language Lexicography. In P. Hanks, G.-M. de Schryver (eds.), *International Handbook of Modern Lexis and Lexicography* Accessed at: <https://www.semanticscholar.org/paper/Sign-language-lexicography-McKee-Vale/e3d1d3a158ec1fd1652fa6462129616fd1baf86a> [27/04/2020].
- Mesch, J., Wallin, L. (2008). Use of sign language materials in teaching. In LREC 2008 Workshop Proceedings: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. Marrakech, 1 June 2008, pp. 134-137.
- Mesch, J., Schönström, K. From design and collection to annotation of a learner corpus of sign language. In LREC 2018 Workshop Proceedings: 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. Miyazaki, 12 May 2018, pp. 121-126.
- Meurant, L., Cleve, A., & Crasborn, O. (2016). Using sign language corpora as bilingual corpora for data mining: contrastive linguistics and computer-assisted annotation. In 7th Workshop on the Representation and Processing of

- Sign Languages: Corpus Mining. Portorož, 28 May 2016, pp. 159-166.
- Miller, C. (2006). Sign Language: Transcription, Notation, and Writing. In K. Brown (ed.) *Encyclopedia of Language and Linguistics* (2nd ed.), 328-338. Amsterdam: Elsevier. Accessed at: <https://www.sciencedirect.com/science/article/pii/B008044854200242X> [27/04/2020].
- Morgan, G., Woll, B. (2007). Understanding sign language classifiers through a polycomponential approach. *Lingua*, 117: 1159-1168.
- Naert, L., Reverdy, C., Larboulette, C., & Gibet, S. Per channel automatic annotation of sign language motion capture data. In LREC 2018 Workshop Proceedings: 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. Miyazaki, 12 May 2018, pp. 139-146.
- NOEMA+. Online Greek Sign Language Dictionary. Accessed at: <http://sign.ilsp.gr/signilsp-site/index.php/el/noima/> [27/04/2020].
- Chen Pichler, D., Hochgesang, J., Simons, D., & Lillo-Martin, D. Community input on re-consenting for data sharing. In 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining. Portorož, 28 May 2016, pp. 29-34.
- Prillwitz, S., Leven, R., Zienert, H., Hanke, T., & Henning, J. (1989). *HamNoSys. Version 2.0. Hamburg Notation System for Sign Language: An Introductory Guide*. Hamburg: Signum Verlag.
- Sandler, W. (2006). Sign Language: Overview. In K. Brown (ed.) *Encyclopedia of Language and Linguistics* (2nd ed.), 328-338. Amsterdam: Elsevier. Accessed at: <https://www.sciencedirect.com/science/article/pii/B008044854200239X> [27/04/2020].
- Stamper, K. (2012, May 9). A Letter to a Prospective Lexicographer. *harm-less drudg-ery*. Accessed at: <https://korystamper.wordpress.com/> [27/04/2020].
- Schermer, T. (2006). Sign Language: Lexicography. In K. Brown (ed.) *Encyclopedia of Language and Linguistics* (2nd ed.), 321-324. Amsterdam: Elsevier. Accessed at: <https://www.sciencedirect.com/science/article/pii/B0080448542002315> [27/04/2020].
- Timmermans, N. (2005). *The status of sign languages in Europe*. Report drawn in co-operation with the Committee on the Rehabilitation and Integration of People with disabilities. Strasbourg: Council of Europe.
- Vacalopoulou, A., Efthimiou, E., Fotinea, S.E., Goulas, T., & Dimou, A.N. (2018). Making online educational content accessible in Greek sign language. In EDULEARN2018 Proceedings: 10th International Conference on Education and New Learning Technologies. Palma, 2-4 July 2018, pp. 7305-7310.
- Vacalopoulou, A., Efthimiou, E., & Vasilaki, K. (2018). Multimodal corpus lexicography: compiling a corpus-based bilingual Modern Greek—Greek Sign Language dictionary. In XVIII EURALEX International Congress: Lexicography in Global Contexts. Ljubljana, 17-21 July 2018, pp. 17-21.
- Vale, M. (2017). Folk definitions as a model for sign language dictionary definitions: A user-focused case study of the Online Dictionary of New Zealand Sign Language. PhD thesis. Victoria University of Wellington, Wellington, New Zealand.
- Yule, G. (2020). *The Study of Language* (7th ed.). Cambridge: Cambridge University Press.
- Wilcox, S., Occhino, C. Historical Change in Signed Languages. (2016). In *Oxford Handbooks Online*. Accessed at: <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935345.001.0001/oxfordhb-9780199935345-e-24> [27/04/2020].
- Zgusta, L. (1971). *Manual of Lexicography*. Prague: Academia.
- Zwitsersloot, I. (2010). Sign language lexicography in the early 21st century and a recently published dictionary of Sign Language of the Netherlands. In *International Journal of Lexicography*, 23(4), pp. 443-476.

Acknowledgements

I acknowledge support of this work by the project “Computational Sciences and Technologies for Data, Content and Interaction” (MIS 5002437) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund). I am also indebted to my colleague, Kiki Vasilaki, for her helpful insights in GSL.