Proceedings of the 10th Workshop on the Representation and Processing of Sign Languages (sign-lang@LREC 2022), pages 118–123
Language Resources and Evaluation Conference (LREC 2022), Marseille, 20-25 June 2022

Open Repository of the Polish Sign Language Corpus: Publication Project of the Polish Sign Language Corpus

Anna Kuder, Joanna Wójcicka, Piotr Mostowski, Paweł Rutkowski
Section for Sign Linguistics, Unit for General Linguistics and Sign Linguistics,
Faculty of Polish Studies, University of Warsaw
Krakowskie Przedmieście 26/28, 00-927 Warsaw, Poland
{anna.kuder, j.filipczak, piotr.mostowski, p.rutkowski}@uw.edu.pl

Abstract

Between 2010 and 2020, the research team of the Section for Sign Linguistics collected, annotated, and translated a large corpus of Polish Sign Language (*polski język migowy*, PJM). After this task was finished, a substantial part of the gathered materials was published online as the Open Repository of the Polish Sign Language Corpus. The current paper gives an overview of the process of converting the material from the Corpus into the Repository. If presents and explains the decisions made along the way and describes the process of data preparation and publication.

There are two levels of access to the Repository, which are meant to fulfil the needs of a wide range of public users, from members of the Deaf community, through hearing students of PJM, sign language teachers and interpreters, to users with academic background. We describe how corpus material available in open access was prepared to be searchable by text type and elicitation tasks, by sociolinguistic metadata, and by translation into written Polish. We go on to explain how access for research purposes differs from open access. We present possible ways in which data gathered in the Repository may be used by members of the signing community in Poland and abroad.

Keywords: sign language corpus, sign language data annotation, repository, corpus data publication, open access, Polish Sign Language (PJM)

1. Introduction

The Polish Sign Language Corpus project was launched in 2010 at the University of Warsaw by the Section for Sign Linguistics (SSL) of the Faculty of Polish Studies. The endeavour was undertaken for two main reasons: firstly, to document Polish Sign Language (*polski język migowy*, PJM) – the endangered language which is the main vessel of the Deaf culture in Poland. Secondly, to serve as a solid empirical foundation for detailed grammatical and lexical studies of the language (which at that time had only been analysed on the basis of small data samples or researchers' and users' intuitions). The PJM Corpus project was designed to match other modern big-scale sign language corpora that were being developed around that time in different parts of the world, e.g., in Germany[1] (Prillwitz et al., 2008), Australia[2] (Johnston, 2009), Great Britain[3] (Schembri et al., 2013) and the Netherlands[4] (Crasborn and Zwitserlood, 2008).

2. Polish Sign Language Corpus

In the 10 years of the project's development (2010-2020), we obtained elicited recordings from 150 Deaf PJM signers from all over the country. The group of informants was balanced with respect to their age, place of origin and gender. Taken together, all the frontal-view recordings of individual signers are approx. 565 hours in length, as each recording session involved two participants and lasted approx. 3-5 hours. The elicitation scenario contained approx. 24 tasks that the informants were asked to complete in a face-to-face conversation context. The elicitation materials used in the PJM Corpus project were designed in close collaboration with the German Sign Language corpus team operating at the University of Hamburg (Nishio et al., 2010).

After each recording session, the collected data was backed up, stored on the University of Warsaw servers and annotated in the iLex software (Hanke and Storz, 2008). The methodological choices made during the PJM Corpus creation are described in detail in: Rutkowski et al., 2013; Rutkowski and Łozińska, 2014; Rutkowski et al., 2017.

The annotation team consisted of approx. 20 trained Deaf signers and hearing linguists. The annotation protocol used in the PJM Corpus (overview in: Filipczak, 2014 and in section 3.2.2.2 of this text) was based on annotation schemas used in other sign language corpus projects, including the ECHO project (Nonhebel et al., 2003), Australian Sign Language Corpus (Johnston, 2010, 2019), German Sign Language Corpus (Prillwitz et al., 2008) and others. During the annotation process the team distinguished 687,971 sign tokens (which were grouped into 15,384 sign types) and inserted 1,340,536 additional tags. The latter were used to represent the following levels of linguistic analysis:

- segmentation into clause-like units (developed based on the CLU tagging methodology proposed by Johnston (2019)),
- parts of speech,
- negation (overview in: Kuder, 2021),
- transcription into HamNoSys notation (Prillwitz et al, 1989).

On the basis of these primary annotations, 67,698 PJM sentences were translated into written Polish.

While the PJM Corpus was still being annotated it served as the empirical basis for different research and educational projects. The most important ones have included: the creation of *The Corpus-based Dictionary of Polish Sign Language*[5] (Łacheta et al., 2016); detailed linguistic analyses of PJM verbs (Łozińska, 2014) and of negation (Kuder, 2021); the creation of *Sign with us 1* and *Sign with*

[1] http://www.sign-lang.uni-hamburg.de/dgs-korpus
[2] http://www.auslan.org.au/about/corpus
[3] http://www.bslcorpusproject.org
[4] http://www.ru.nl/corpusngt
[5] https://www.slownikpjm.uw.edu.pl/en

us 2[6] – two parts of an extensive multi-media PJM course for children corresponding to A1-B2 CEFR proficiency levels. A number of smaller research endeavours have included studies on different linguistic, socio-linguistic, and psycho-linguistic features of PJM and Deaf Culture that have been presented through the years at various conferences and workshops in Poland and internationally, and published in numerous papers and articles.[7]

3. Open Repository of the Polish Sign Language Corpus

The PJM Corpus project team decided to make a significant portion of the collected data publicly available, following good practices set by other corpus projects of this type (e.g., Schembri, 2008 for British Sign Language; Bono et al., 2014 for Japanese Sign Language; Jahn et al., 2018 for German Sign Language). Therefore, when the project was completed, the project team moved onto designing the most sustainable way of on-line data publication. The platform used for that purpose was entitled the *Open Repository of the Polish Sign Language Corpus* (see section 3.1 below, henceforth the Repository).

Firstly, we decided to publish only a limited number of tasks performed by each of our informants. This was related to the need to protect sensitive information shared during the recording sessions (see Crasborn, 2008 and section 3.2 for details).

Our second decision concerned publishing different types of signed texts, so as to represent different genres. Therefore, the Repository features mono- and dialogical texts (see section 3.2 for details), including retellings and narratives that were produced in response to materials eliciting the use of different language structures and expressions (e.g., depicting signs, constructed actions, expressions involving the use of sign space, negation, different types of discourse structures).

Our third decision concerned the sociolinguistic data. The portion of the PJM Corpus data published in the Repository corresponds to the entire Corpus, balanced with respect to the informants' age, place of origin and gender (as the Repository is meant to reflect regional variation and differences in signing between younger and older signers).

Lastly, we distinguished four potential Repository user groups:

- casual visitors (deaf and hearing members of the signing community and all persons interested in sign language who wish to use PJM materials created for non-professional purposes, e.g., to see the diversity of signing styles, regional variation, etc.; this group includes persons concerned with preserving PJM as an endangered language documenting Deaf Culture);
- sign language teachers and students (who are, on the one hand, interested in obtaining materials that can be used in teaching sign language – e.g., for lesson practice or homework – and, on the other hand, in studying sign language and expanding their language competences by observing signing styles of deaf people from various environments and backgrounds; Crasborn, 2008);
- sign language interpreters (who can benefit from obtaining signed videos in which the same topic is covered by informants with different signing styles, which can serve as an excellent source of training materials; Leeson, 2008);
- sign language researchers (who are especially interested in obtaining materials for conducting detailed analyses of the PJM linguistic structure and other aspects of signed communication, including for comparative studies).

Our choices informing data publication were made to match the needs of those groups (see section 3.2.1 and 3.2.2 for an overview).

3.1 Name and Visual Representation

Among one of the first choices we made before publication, was the choice concerning the name and the visual representation of the published materials. As not all of the PJM Corpus materials are publicly available, we decided its shared part should go under a different name – not to cause any confusion as to its character. A decision was thus made to call the corpus website the *Open Repository of the Polish Sign Language Corpus* and make it publicly accessible at: www.korpuspjm.uw.edu.pl/en (Wójcicka et al., 2020). The website can be accessed by scanning the QR code below. We decided to publish the Repository on the University of Warsaw servers, to ensure good quality of hosting and to minimize the financial cost of maintaining the site.

Figure 1: QR code linking to the Repository website.

Then, we developed the visual identification of the Repository, consistent with the visual identification assumptions of other projects carried out by the SSL. A separate logotype (Fig. 2) is meant to indicate that the Repository should not be confused with the PJM Corpus.

Figure 2: The Open Repository of the Polish Sign Language Corpus logotype.

The Repository webpage contains all the video files that were qualified for publication. Additionally, the page includes information on PJM, the PJM Corpus project, the SSL team and a thorough instruction on how to use the resource. All Polish written texts are translated into PJM, written English, and International Sign.

[6] https://www.gov.pl/web/edukacja-i-nauka/kurs-polskiego-jezyka-migowego-pjm

[7] https://www.plm.uw.edu.pl/publikacje/

3.2 Data Publication

After each recording session, the collected video recordings were first segmented into 24 separate clips, in line with the elicitation task schema. This segmentation into individual tasks was not limited to specific task-related productions, but rather included instructions from the moderator, follow-up questions, clarifications and nonrelevant signing and, therefore, the pre-segmented recordings contained sensitive information that should not be published. For that reason, the recorded material to be included in the Repository was carefully examined and extracted from the extensive recording sets to minimize the possibility of publicly sharing any personal data unrelated to the elicitation tasks. Additionally, in some cases our informants did not fully grasp the instruction to the given task and their signing did not fully correspond with the elicitation materials. All of these clips were excluded from publication.

What is more, we also came across a number of videos in which our informants shared some information that could potentially be demeaning for them personally and for people mentioned by them, and also uncomfortable for the users of the Repository. We share the concerns of other sign language researchers who point out the issue of the anonymization of publicly available corpus data (Isard, 2020). Since we did not want to modify our data by means of concealment methods (as it would result in having to blur out too many clips), we decided, for ethical reasons, not to publish those parts of the PJM Corpus videos that contained sensitive or uncomfortable and disturbing data (Crasborn, 2008).

Having prepared the video material, we focused on designing the way in which the data will be presented on the website. In this regard we followed the choices made by the teams publishing the British (Schembri, 2008) and Japanese (Bono et al., 2014) Sign Languge Corpora. We drew upon the idea of publishing the collected data in different ways in order to better serve the needs of different potential Repository users (see section 3). After careful considerations, we decided to create two ways of accessing the Repository:

- fully open access for all interested parties;
- access requiring previous registration, for researchers, interpreters, and teachers.

3.2.1 Open Access

To make the PJM Repository openly accessible from different angles and to fulfill the needs of various users and various research purposes, different browsing options have been provided. As shown in Figure 3, the open access mode offers three ways to search the set of video clips constituting the Repository:

- by the birthplace, age, and gender of the signer (sociolinguistic data),
- by the type and topic of discussion,
- by words present in written Polish translations.

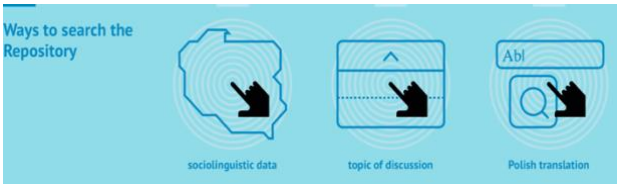

Figure 3: Three different ways to search the Repository.

After clicking the chosen icon corresponding to the particular search method, the user is presented with further options to specify the search. When browsing by sociolinguistic data, the user is presented with a map of Poland (Fig. 4), from which they can choose the part of the country they are interested in. In the next step the user is taken to the page with clips which allows a more fine-grained selection, the user can choose the interesting age range (possibilities: 18-30; 31-45; 46-60; 60+) and gender(s) of signers (as shown in Fig. 5 below).

Figure 4: Browsing through sociolinguistic data.

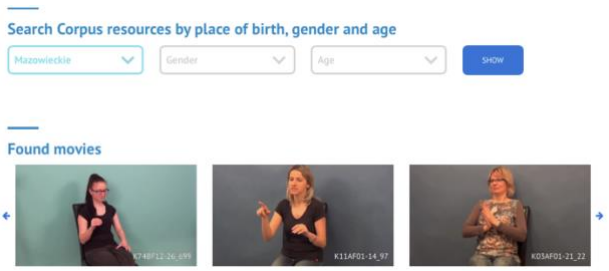

Figure 5: Selection page after browsing by sociolinguistic data.

When browsing by discussion topic, the user is presented with a list of drop-down menus to specify their selection. They can choose different text types, stimulus types, elicitation material types and, finally, specific elicitation tasks.

These search procedures reflect the extent of variation of the PJM Corpus elicitation materials. Some of the materials were designed to elicit monologue responses, other clips result from tasks that elicited conversational interaction.

In the Repository, all monologue tasks are divided into groups depending on the type of stimulus used for elicitation – either a movie clip or a drawing. Movie clips used in the elicitation showed either films/animations (therefore, eliciting narrative retelling) or a signing person (therefore, eliciting retelling of a signed story).

In the last browsing step, the user can choose the name of a specific task. To get to know which elicitation materials have been used for which task, the user should consult the detailed description in the 'how to use' section on the Repository website.
When it comes to dialogues included in the Repository, the informants either:

- discuss the interpretation of the pictures they are shown, or
- are asked to set the date for the meeting based on the calendar pages they are shown, or
- describe short comic strips to one another.

In the PJM Repository, clips with conversations show both signers performing the task simultaneously (to reflect the interactional nature of the recorded discourse). In the case of monologues, we decided to publish videos of the single informant (as shown in Fig. 6).

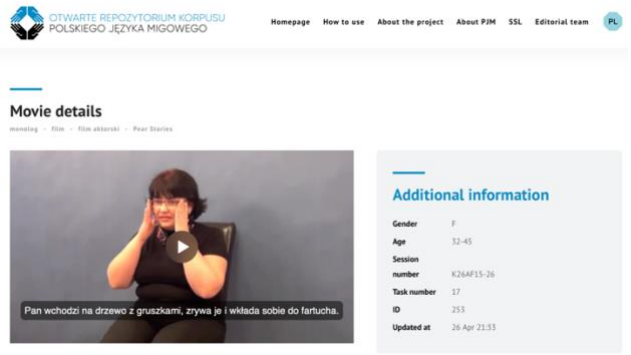

Figure 6: Individual video page (https://www.korpuspjm.uw.edu.pl/en/videos/178_).

When searching by Polish translations, the user is asked to type a Polish word into the search box. The search is performed in the form of a full-text search — the form of a word in the text is matched without the possibility of searching by lemmas or grammatical categories. The search results provide a list of videos that contain the given form, without specifying the precise timecode in which the written word occurs. More fine-grained search on the translation tier is possible after obtaining the "access for research purposes" – see section 3.2.2 – and using .eaf files.
After specifying their selection, the user is taken to a page listing all video clips in the Repository that meet the chosen criteria. Clicking on a particular video clip takes the user to the video individual page (Fig. 6), giving the possibility to watch the recording with Polish subtitles (based on the free translation extracted from the PJM Corpus) and providing detailed information about the clip, including the following:

- gender and age of the signer,
- the number of the recording session during which this particular clip was recorded,
- number of the task from the PJM Corpus that the clip corresponds to,
- ID of the clip,
- date of the last update,
- title of the task,
- text type
- the type of the stimulus used to elicit material.

3.2.2 Access for Research Purposes

The videos and data available in the open access mode are meant to serve the needs of general users, such as casual visitors. However, that form of access has its shortcomings from the point of view of professional users. For this reason, another type of Repository access is offered for those who wish to register. It is aimed at users with research interests who will be able to justify their need to have broader access to the Repository (for example, teachers, interpreters, or linguists). The access is granted individually by the Repository editorial team upon the receipt of an on-line request.
After obtaining the editorial team's approval and creating a personal account, the user can still use the Repository in the standard way available in the open access. Additionally, they can download the .eaf file for each of the video clips contained in the Repository, to be opened in the ELAN software (Crasborn and Sloetjes, 2008).

3.2.2.1 Annotation Files

Each .eaf file corresponds to a video with a single informant (regardless of the type of activity: monologue or dialogue). The .eaf files accessible from the Repository contain four tiers of annotation extracted directly from the PJM Corpus:

- gloss for dominant hand,
- gloss for non-dominant hand,
- HamNoSys transcription,
- free translation into written Polish.

Tiers with annotation follow the conventions applied in the PJM Corpus (see section 3.2.2.2. and Filipczak, 2014).
The dominant/non-dominant hand tier distinction is not meant to reflect the structure of particular signs, as two-handed signs are annotated on the dominant hand tier. The non-dominant hand tier is used for independent articulations (such as buoys, simultaneous signing using two hands, changing hand dominance, anticipations, perseverations, etc.). The layer with free translation was prepared for public access purposes to convey the general meaning of PJM utterances in Polish.
The .eaf files are extracted from the PJM Corpus with a custom software created to facilitate and accelerate export of iLex files according to the needs of the Repository publication team. The custom-built exporter allows to adjust annotation files with specific filters (e.g., time stamps, particular informants, chosen tasks or tiers) and additional configuration for extracting selected indexed parts of annotations or fixing Unicode encoding issues.

3.2.2.2 Annotation Conventions in the PJM Corpus and Repository

The Repository annotation conventions follow the annotation schema used in the creation of the PJM Corpus, where the videos are glossed using conventions that differentiate between lexicalized ('frozen') signs and non-lexicalized signs.
The latter group contains non-frozen signs and gestures divided in groups such as productive signs (including partly-lexicalized classifiers, glossed with the prefix *$:KL:*), phatic gestures (glossed with symbols & and @), palm-ups (glossed as %) and other gestures (glossed with the prefix *G:*). Fingerspelling glosses (annotated as: *A.B.C.D.*) are not treated as lexicalized.

Lexicalized forms have a hierarchical structure differentiating between types and subtypes. Types are used as labels for main glosses (Johnston, 2008) and a Polish equivalent is chosen to represent the sign's approximate meaning. Types (labeled 'Signs' in the iLex software) are written in capital letters and are always presented in the form of an uninflected Polish word. Lexical variants are distinguished by numbers: e.g., *BUS1*, *BUS2*. In the PJM Corpus, type labels are accompanied by conventionalized notations of hand configuration for the right and left hand for the most frequent form (mainly used for facilitating annotation process in identifying lexical variants): e.g., *BUS1 P:O;L:O*, *BUS2 P:CC;L:CC*, where *P:* stands for the right hand (*prawa* in Polish), *L:* stands for the left hand (*lewa* in Polish), and the letters following the colon specify a given handshape.

For searching purposes some lexical signs are grouped by using specific prefixes: e.g., *NUM:* for numerals, *IDENTYF:* for sign names.

Division into subtypes serves the purpose of differentiating between articulatory variants of the main type, including variants with modified handshapes (e.g., *CAT 1.1 P:O;L:O*, *CAT 1.2 P:Z;L:Z*), weak-hand drop (*CAT 1.3 P:O;L:Ø*, where *Ø* indicates that the hand is not in use), orientation, localization, or modification of movement. This way of formulating type glosses provides additional options when it comes to analyzing phonological and phonetical variation in signs. In the Repository files, data are represented by the main gloss accompanied with numbers for lexical variants and basic hand articulation notation.

This way of representing signed data in annotation is not identical but consistent enough with annotation methods used in other large-scale corpus projects worldwide, so it allows for comparative studies of sign language phenomena between different sign languages.

4. Conclusion

Publishing materials extracted from an existing sign language corpus is a complex, multi-dimensional process. Decisions must be made regarding the choice of materials suitable for publication, their anonymization, form of presentation and features offered to the website users.

In this paper, we have presented a detailed description of the Open Repository of the Polish Sign Language Corpus. We have focused mainly on the user perspective, but we have also explained the decisions made in the publication process. We have presented various methods of browsing and searching the published data (including the ability to download annotations in the form of .eaf files). Publication in two modes (open access without registration and registered access) is meant to serve the needs of different users.

The Repository can be used by linguists, researchers in the field of deaf studies, sign language teachers, translators/interpreters, L2 learners, and everyone interested in sign language. The Repository also serves as one of the largest available datasets documenting and archiving the language and culture of the Deaf community in Poland.

The Repository is a closed publication, but its form remains open. It is possible for it to be expanded with additional materials created in subsequent research projects.

5. Acknowledgements

The first phase of the PJM Corpus project was supported financially by Poland's National Science Center (*Narodowe Centrum Nauki*) under the project *Iconicity in the grammar and lexicon of Polish Sign Language (PJM)* (grant number: 2011/01/M/HS2/03661) and by the Foundation for Polish Science (*Fundacja na rzecz Nauki Polskiej*) under the project *Grammatical categorization through space and movement in Polish Sign Language* (grant number: 1/2009). The second phase was financed by the Polish Ministry of Science and Higher Education (*Ministerstwo Nauki i Szkolnictwa Wyższego*) under the National Program for the Development of Humanities (*Narodowy Program Rozwoju Humanistyki* – project title: *Multi-layered linguistic annotation of the corpus of Polish Sign Language (PJM)*; grant number: 0111/NPRH3/H12/82/2014).

6. Bibliographical References

Bono, M., Kikuchi, K., Cibulka, P. and Osugi, Y. (2014). A Colloquial Corpus of Japanese Sign Language: Linguistic Resources for Observing Sign Language Conversations. In *Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel. International Conference on Language Resources and Evaluation* (*LREC 2014*), pages 23–31, Reykjavik, Iceland, May. European Language Resource Association (ELRA).

Crasborn, O. (2008). Open Access to Sign Language Corpora. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation* (*LREC 2008*), pages 33–38, Marrakech, Morocco, May. European Language Resource Association (ELRA).

Crasborn, O. and Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation* (*LREC 2008*), pages 39–43, Marrakech, Morocco, May. European Language Resource Association (ELRA).

Crasborn, O. and Zwitserlood, I. (2008). The Corpus NGT: an online corpus for professionals and laymen. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation* (*LREC 2008*), pages 44–49, Marrakech, Morocco, May. European Language Resource Association (ELRA).

Filipczak, J. (2014). Anotacja korpusu PJM [The PJM Corpus Annotation Process]. In P. Rutkowski & S. Łozińska (eds.), *Lingwistyka przestrzeni i ruchu. Komunikacja migowa a metody korpusowe* [*Linguistics of space and movement. Sign language communication and corpus methods*]. Warsaw: Faculty of Polish Studies, University of Warsaw, pp. 91–105.

Hanke, T and Storz, J. (2008). iLex – A database tool for integrating sign language corpus linguistics and sign

language lexicography. In *Proceedings of the 3rd Workshop on the Representation and Processing of Signed Languages: Construction and exploitation of sign language corpora. International Conference on Language Resources and Evaluation* (*LREC 2008*), pages 64–67, Marrakech, Morocco, May. European Language Resource Association (ELRA).

Isard, A. (2020). Approaches to the Anonymisation of Sign Language Corpora. In *Proceedings of the 9th Workshop on the Representation and Processing of Signed Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. International Conference on Language Resources and Evaluation* (*LREC 2020*), pages 95–100, Marseille, France, May. European Language Resource Association (ELRA).

Jahn, E., Konrad, R., Langer, G., Wagner, S. and Hanke, T. (2018). Publishing DGS Corpus Data: Different Formats for Different Needs. In *Proceedings of the 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. Language Resources and Evaluation Conference* (*LREC 2018*), pages 83–90, Miyazaki, Japan, May. European Language Resource Association (ELRA).

Johnston, T. (2008). Corpus linguistics and signed languages: no lemmata, no corpus. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages*: *Construction and Exploitation of Sign Language Corpora. Language Resources and Evaluation Conference* (*LREC 2008*), pages 82–87, Marrakech, Morocco, May. European Language Resource Association (ELRA).

Johnston, T. (2009). Creating a corpus of Auslan within an Australian national corpus. In M. Haugh, K. Burridge, J. Mulder and P. Peters (eds.), *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages*. Somerville: Cascadilla Proceedings Project, pp. 87–95.

Johnston, T. (2010). From archive to corpus: transcription and annotation in the creation of signed language corpora. *International journal of Corpus Linguistics*, 15(1):106–131. https://doi.org/10.1075/ijcl.15.1.05joh

Johnston, T. (2019). Auslan Corpus Annotation Guidelines. August 2019 Version, manuscript, Macquarie University, Sydney, Australia, on-line at: https://www.academia.edu/40088269/Auslan_Corpus_Annotation_Guidelines_August_2019_version

Kuder, A. (2021). Negation markers in Polish Sign Language (PJM), *Sign Language and Linguistics*, 24(1):118–131. https://doi.org/10.1075/sll.00055.kud

Leeson, L. (2008). Quantum leap. Leveraging the signs of Ireland digital corpus in Irish Sign Language/English interpreter training. *The Sign Language Translator and Interpreter,* 2:231–258.

Łozińska, S. (2014), Czasownik w polskim języku migowym. Studium semantyczno-gramatyczne [Verb in PJM. Semantic and grammatic analyses]. Unpublished PhD thesis. Warsaw: Faculty of Polish Studies, University of Warsaw.

Nishio, R., Hong, S.-E., König, S., Konrad, R., Langer, G., Hanke T., and Rathmann, C. (2010). Elicitation methods in the DGS (German Sign Language) Corpus Project. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. Language Resources and Evaluation Conference* (*LREC 2010*), pages 178–185, Valletta, Malta, May. European Language Resource Association (ELRA).

Nonhebel, A., Crasborn, O. and van der Kooij, E. (2003). Sign language transcription conventions for the ECHO project. University of Nijmegen, Tech. Rep. 6, on-line at: https://repository.ubn.ru.nl/bitstream/handle/2066/57889/57889.pdf

Prillwitz, S., Leven, R., Zienert, H., Hanke, T. and Henning, J. (1989). HamNoSys: Version 2.0. Hamburg Notation System for sign languages. An introductory guide, Hamburg: Signum.

Prillwitz, S., Hanke, T., König, S., Konrad, R., Langer, G. and Schwarz, A. (2008). DGS Corpus Project – Development of a Corpus Based Electronic Dictionary German Sign Language / German. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation* (*LREC 2008*), pages 159–164, Marrakech, Morocco, May. European Language Resource Association (ELRA).

Rutkowski, P., Łozińska, S., Filipczak, J., Łacheta, J. and Mostowski, P. (2013). Jak powstaje korpus polskiego języka migowego (PJM)? [How is the PJM Corpus created?]. *Polonica,* 33:297–308.

Rutkowski, P. and Łozińska, S., eds. (2014). Lingwistyka przestrzeni i ruchu. Komunikacja migowa a metody korpusowe [Linguistics of space and movement. Sign language communication and corpus methods]. Warsaw: Faculty of Polish Studies, University of Warsaw.

Rutkowski, P., Kuder, A., Filipczak, J., Mostowski, P., Łacheta, J. and Łozińska, S. (2017). The design and compilation of the Polish Sign Language (PJM) Corpus, In P. Rutkowski (ed.), *Different faces of sign language research*. Warsaw: University of Warsaw, Faculty of Polish Studies, pp. 125–151.

Schembri, A. (2008). British Sign Language Corpus Project: Open Access Archives and the Observer's Paradox. In *Proceedings of the 3rd Workshop on the Representation and Processing of Signed Languages: Construction and exploitation of sign language corpora. International Conference on Language Resources and Evaluation* (*LREC 2008*), pages 165–169, Marrakech, Morocco, May. European Language Resource Association (ELRA).

Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S. and Cormier, K. (2013). Building the British Sign Language Corpus. *Language Documentation & Conservation*, 7:136–154

7. Language Resources References

Łacheta, J., Czajkowska-Kisil, M., Linde-Usiekniewicz, J. and Rutkowski, P. eds., (2016). The Corpus-based Dictionary of Polish Sign Language, Warsaw: Faculty of Polish Studies, University of Warsaw. http://www.slownikpjm.uw.edu.pl.

Wójcicka, J., Kuder, A., Mostowski, P. and Rutkowski, P. (2020). Open Repository of the Polish Sign Language Corpus. Warsaw: Faculty of Polish Studies, University of Warsaw. https://www.korpuspjm.uw.edu.pl/en