Using Components of Corpus Linguistics and Annotation Tools in Sign Language Teaching

Katarzyna Barczewska
AGH University of Science and Technology, Department of Automatics and Bioengineering, Department of Electronics,
Krakow, Poland
E-mail: kbarczew@agh.edu.pl

Jakub Gałka, Dorota Szulc, Mariusz Mąsior, Rafał Samborski
AGH University of Science and Technology, Department of Electronics, Krakow, Poland

Tomasz J. Wilczyński
AGH University of Science and Technology, Department of Mechanics and Vibroacoustics, Department of Electronics,
Krakow, Poland

Abstract—As an interdisciplinary group of people working on automatic sign language recognition, authors of the article developed concept how to facilitate the process of understanding Sign Language (SL) utterances by hearing learners. Concept is based on Polish SL, and was created in response to signals from the students who indicate understanding of SL messages as the hardest part of learning process, but can be easily adapted to other SLs. In comparison to speech corpora, SL databases constitute only a small fraction. From the other side thanks to the Internet there are available video recordings with native-signers, as well as tools which enable their analysis. Developed concept is based on using one of the available annotation tools used by professionals all over the world to describe SL corpora in SL classes. Making annotations helps students in understanding foreign language, corpora analysis helps to find objective rules governing SL, computer assisted language learning can be attractive way of study and Internet – great source of materials.

Index Terms—Sign language, annotation tools, computer-assisted language learning, sign language corpora.

I. INTRODUCTION

In every society, there is a group of people who communicate in Sign Language (SL). Despite the fact that they are minority, their needs should not be forgotten. Their different way of communicating has an impact on many situations: access to information, services, education and employment. Matters of everyday life, which for hearing people are ordinary routine, for deaf can be real problems which are often insurmountable without help of the others. These people, using different language, are like foreigners in their own country. Communication barrier creates a feeling of isolation and makes it very difficult to tell the world about their unique experience, and thus causes a conviction of impossibility of being an equal partner to the other and problems with equal access in the information society. There are some works, projects which are carried out on the various facilities that could help deaf people overcoming these barriers [1], [2]. They can be divided into two groups: 1) solutions that consists of automatic translation of the phonic language utterances into a form understandable by deaf people; 2) solutions that perform translations but in opposite direction. However there are more and more research projects which aim is to create such systems, as well as algorithms and methods developed for automatic speech recognition and translation systems which are adapted to SL recognition, there are not many commercially available solutions. The main reasons are: lack of formal description of SL structure and grammar (for majority of SLs, despite American SL which is considered to be the most widely described SL), insufficient amount of data (especially in comparison to data needed by phonic language translators) that could be used to prepare such description required by natural language processing methods which have to be applied to model the language.

Learning SL is a difficult task for hearing people. It requires a lot of hard work and effort. Difficulties that may occur in the process of learning SL as a foreign language are primarily due to a colossal difference between sign and spoken languages: the first is constructed on the base of visual modality, and the second – is based on the auditory modality. A big challenge for hearing people is to change their way of perception from auditory to visual. This skill helps in understanding how to function in the deaf community [4]. Although it is hard, it is possible for hearing people to reach the level of communicative competence in SL. If it is hard, but possible, there is the question: how can we help hearing students in facilitating the process of learning SL and educate specialists that would be able to create missing corpora?

As an interdisciplinary group of people working on a

project of Automatic Sign Language Recognition (ASLR), the authors of that article, taking into account mentioned problems, proposed solution that could help in enlarging databases of well annotated material crucial for ASLR systems. Proposed solution is long-term solution and is to educate hearing students who learn SL in special way that they could build and describe SL databases and corpora. The concept is based on Annotation Tools (ATs) that could be used as a teaching aid during SLs classes. ATs are used by researchers to prepare database which contains recordings with SL gestures or utterances for further analysis. Annotated material is used to develop signs, gestures models as well as to build SL models which are necessary for ASLR. According to the literature review it was found that using annotations in the process of language learning/teaching can have many advantages. The effectiveness of using multimedia annotations on foreign language improve reading comprehension and attitudes was previously described in [5]. The usage requirements of ATs was widely discussed in [3]. The example of employing the AT for large amounts of files with SL recordings is shown in [6]. Work presented by Dina and Ciornei [7] gives some more general remarks on the advantages and disadvantages of employing computers in teaching foreign languages.

The article describes elements of corpus linguistics used in technical science (Chapter II), annotation of SL (Chapter III), examples of ATs (Chapter IV) and at the end the concept of their use as a teaching aid during SL classes (Chapter V).

II. Elements of Corpus Linguistics in Technical Science

The research on SL is a quite new direction of investigation. It covers the area of linguistics and communication as well as technical science.

The entire task of any language study is performed by analysis of samples, text or recordings, and is expressed in a concept of corpus linguistics. A rich set methods of tracing path from data to theory was created during development of the corpus linguistics researches. The main idea consists of three additional, descriptive layers for corpus data [8]: Annotation (structural markups, part-of-speech tagging, parsing, etc.), Abstraction (terms translation into a theoretically motivated model or dataset) and Analysis (statistical evaluations optimizations and generalizations of the dataset). Thus well described and annotated corpus of SL is a mine of knowledge for linguists and their language studies as well as for technicians and development of automatic SL synthesis or recognition systems.

The last problem, ASLR, is the most interesting for technical science and this entire task could be referred to the task of building the Automatic Speech Recognition (ASR) system. Both of these possibilities of communication use different methods of transferring information (image versus sound), but many phenomena for both issues are similar. Creation of advanced signal processing software, such as the ASR systems, demands a large amount of speech recordings for building phonetic and language models. The most successful ASR systems in existence are based on hundreds and thousands of hours of speech [9]. Creating an extensive set of recordings is expensive and time-consuming and making annotations of all the recordings makes it an even more costly process. It is assumed that SL research requires data of quality similar to that of speech technology, but it turns out that it requires a much bigger effort. The first reason is greater complexity and diversity of specific gesture signs, which depends on coarticulation conditioning and signing person specificity. Moreover, all of this concerns not only whole sign gestures, but also subunits (which are a constituent element of SL) [10], [11]. This means that language corpora will require much more work on each language unit, as compared to speech.

The aforementioned issues cause the need of gathering a great quantity of recordings. The necessity of implementing entire tasks results in a very large increase in the cost of the creation and annotation of SL recordings. The visual aspect of SL could be collected in many different ways, for example by: 2D cameras, time-of-flight cameras, ultrasonic sensors and any motion-sensing devices and motion-sensing video game controllers.

Not many SLs have sufficiently extensive corpora of recordings for the purpose of automatic recognition development. The example of SL, on which this article is based on is Polish SL (PSL, PJM). The work on such corpora for PJM is in progress in various research centers, but they encounter difficulties at the stage of annotation. The UW corpora project [12] (the one of the most mature) consists of 500 hours of pre-recorded videos (15TB of data), but only about 10% of them, are annotated. There are corpora for different SLs: corpus NGT (http://www.ru.nl/corpusngt/) which consists of 10 sec - 12 min recordings of more than two thousands conversations in Dutch SL with annotations, RWTH-Boston corpora (http://www-i6.informatik.rwth-aachen.de/aslr/) which contain from 110 to 843 sentences in American SL. Majority of other databases contain only isolated signs or sequences of isolated signs, what is not relevant to modeling language.

Teaching students how to create, use corpus and ATs would be a step to educate specialists who could annotate existing video material as well as create databases for different, not sufficiently investigated SLs or dialects and for technical applications.

III. Annotation of Sign Language

The tradition of describing natural SLs is relatively short as it dates back to the 1960s, when the first scientific paper on the American SL (ASL) was published. It was not until 1980, however, that linguists involved in the study of the language and visual-spatial communication discovered its non-manual components, recognized them as very important and selected as a subject of their study. Although various initiatives in the area of signed language annotation have been undertaken

over the years, no clear standards have been developed. An important contribution to the researches has been Johnston's [13] emphasis on the use on 'ID-glosses'. A gloss is a type of annotation. It is a brief one or two word translation in one language to a word or morpheme in another language. Importantly, different glosses for the same sign are usually used in different contexts to reflect the meaning of the sign in those specific contexts. Consequently, it is often very difficult to know with certainty which sign form is being referred to by a particular gloss because it does not contain any information about the sign form. To address this problem – as proposed by Johnston – signs are identified uniquely and consistently by using an ID-gloss. The process of assigning an ID-gloss to lexical signs in a corpus is essentially lemmatization, which reduces inflected forms of words to their basic forms (lexemes or lemmas) just as ID-glossing ignores idiosyncratic variants or systematic modifications in the form of signs (provided they are not lexicalized) in favor of the underlying citation form (lemma) [13].

The glossed text is enriched by symbols of lexical or morphological items. The Hamburg Notation System (HamNoSys) is an example of such a system of symbols/ technical pictograms. It has its roots in the Stokoe notation system. The HamNoSys notation for a single sign consists of a description of the initial posture plus the action changing this posture in a sequence. It consists of: hand shapes, their orientation and location, actions and non-manual components which are symbolically presented in Fig. 1.
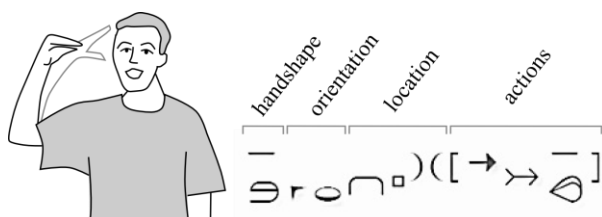

Fig.1. The HamNoSys Structure. DGS Example, Adopted from [14].

Individual research teams, taking into account their specific needs and goals, can make their own marks of linguistics elements. A part of SL annotated text, whose authors are students and lecturers of the Section for Sign Linguistic of the Warsaw University (Poland) is presented in Table 1. Analyzing examples of symbols used in this annotation, one can observe that annotated version contain information about order of signs in the utterance, which is different than in phonic language, description of language structure elements (e.g. classifiers, locators), about simultaneity of performed actions, information about manual, non-manual features, facial expressions, about dynamic of hands movement and their relative positions. There can be added more symbols which describe handshapes, type of movement and other SL components mentioned before.

The basic factor that impedes the linearization of SL texts is their visual and spatial qualities. While phonic language phenomena are linear and sequential in character, their SL counterparts take place in the signing space, and they may occur simultaneously and on many levels. The mechanism of its usage boils down to locating the signs with regard to both the speaker and the recipient: Depending on the message conveyed sings may vary in terms of the distance to both the speaker and the recipient and it term of height.

The simultaneity of SL poses another challenge to those who wish to annotate it. [16] identifies manual and non-manual simultaneity. Manual simultaneity occurs when more than one sign is articulated simultaneously in the manual message. Non-manual simultaneity occurs when other parts of the body rather than just hands are simultaneously involved in the process of conveying a message. These include gestures (eyebrows, lips, eyes, facial expressions), movements of the head, movements of the torso (tilts and turns) and of the entire body. The interaction between manual and non-manual signs proves to be the essence of visual languages. Non-manual components may form an integral and obligatory part of the sign, impart a quality to it, and at the same time, intensify this quality; express its definiteness or non-definiteness, enhance signs which denote activities, create the person category, identify various signs and change their meaning. They may also occur on their own and carry a separate meaning; they may be used to create affirmative utterances, questions and negations as well as exclamations or signals of submission. They are a very important part of the SL, just as important as signs themselves: Without them, the SL would not be adequate enough. Thus, the meaning of one sign comes as a result of combining several manners of articulation and the influence they have on one another. The process of conveying a message in the SL involves the entire body and is very expressive [16].

Table 1. Fragment of SL text writing. The source: [15] pp. 34-39. Symbols meaning: x – hands' folding; @ - pause; /k:…/ - classifier; oo – eye contact with the viewer; & - changing of the direction of signing; Q – changing of mimicry; oo/ruch – spatial, dynamic locator; X* - repetition.

Original Polish Version		
No.	**Symbolic annotation**	**Meaning**
1.	[DZIEŃ DOBRY oo] x@	'Dzień dobry'
2.	[JA [MYŚLEĆ@ {zastanawianie się} WYGLĄDAĆ2 OPOWIADAĆ BAJKA]] %]	'Zastanawiam się, jaką bajkę opowiedzieć.'
3.	& Q [MAMA $ZDEJMOWAĆ_/k:słoik/$*n oo/ruch]	'Mama zdejmuje słoiki z półki, jeden po drugim.'
English Translation		
No.	**Symbolic annotation**	**Meaning**
1.	[MORNING GOOD oo] x@	'Good morning.'
2.	[I [TO THINK@ { to ponder}TO LOOK2 TO TALK STORY]] %]	'I am thinking which story to tell you.' (…)
3.	& Q [MOTHER $ TO TAKE OFF_/k:jar/$*n oo/motion]	'Mum was taking off jars, one after the other.'

Table 2. Comparison of Different Annotation Tools.

Feature of annotation tool	iLex	SignStream	ELAN
It is free	Yes (after contacting the authors)	Yes	Yes
Multiple layers of annotation (tiers)	Yes	Yes	Yes
Multi-tier regular expression search and easy navigation	No	No	Yes
Number of video files that can be associated with an annotation document	?	4	4
Operating system	Mac OS, Windows	Mac OS	Windows, Linux, Mac OS
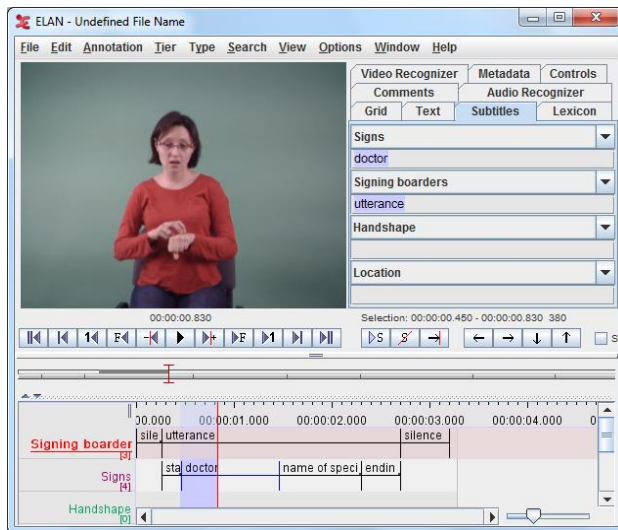
Fig.2. ELAN Main Window with Notes Added In Two Layers (Signing Boarder and Signs).

Yet another challenge is to set the end of utterance. There are no real gaps between signs when signing. Signers do not (and cannot) crisply articulate one sign after another, returning to a neutral position between each sign, nor can a sign sequence be articulated without any transitional movements between each signs [17].

To summarize, a number of components must be taken into account in the process of annotating video recordings of SL texts, i.e. both manual and non-manual components that have no equivalents in the phonic language. Such a representation by its very nature simplifies things, and numerous facts may be lost while other may be preserved in a seriously altered form. However, only then is it possible to search the corpus, create lists of terminology and compare structures. Using software in the process offers one serious advantage: It is possible to freely supplement descriptions of individual fragments or create them at later stages of the study. This means that it is possible to quickly produce a corpus that can be later processed digitally.

IV. ANNOTATION TOOLS

In much research studies on speech or, more generally, linguistics, one of the first steps is the creation of an orthographic or phonetic transcription of some recorded event or experiment [18]. More and more analytic layers of annotation are then added to this transcription by the researchers. These additional layers usually contain some higher level information, such as words, syntax, co-reference or even general comments. To do so, researchers employ ATs.

A great advantage of ATs is the fact that a large part of them is distributed with licenses which allow their free, non-commercial use. This is particularly important when the wide range of features offered by this software is taken into consideration in comparison with tools dedicated to teaching SL. The most popular ATs within the community of researchers working on SL are iLex, SignStream and ELAN. iLex (short for ``integrated lexicon'') was designed and developed by the researchers at the Institute of German SL and Communication of the Deaf in the University of Hamburg in 2002 [19]. SignStream was created at Boston University in 2002 to process American SL data. However, the tool may be applied to any kind of language data captured on video [20]. ELAN was created by The Language Archive, a unit of the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands [6], [21].

All of these tools allow adding text comments, annotations or glosses to video recordings. Simultaneously different features can be described using multiple tiers which can be added to video recording. Each tier can correspond to different articulatory element. They can be used as well to add different levels of annotations or translation. The main window of ELAN is shown in the Fig. 2, the comparison between some features of different ATs is presented in Table 2.

V. ANNOTATION TOOLS AS A TEACHING AID

Teaching the SL to hearing persons primarily aims to develop the skill to communicate through visual and spatial channels and to receive non-verbal messages [22]. This is because when using the SL, hearing persons have to switch from the phonic mode to the visual one and to activate a manner of perceiving and interpreting the world that is new and different from those inherent in phonic languages. For this reason, [23] and [22] recommend that phonic speech should not be used in class, and instead teachers should utilize a selection of dialogues, narratives and utterances in the Polish SL or in special transcription. They suggest that video recordings of Deaf Persons using the SL should be utilized as one of the main teaching aids. But what if hearing students wants to make some notes, as usually when they learn foreign language? And how to make notes on video recordings?

Generally in the foreign language classes annotations are used as a teaching aid or can be treated as students notes. Making marks on text is important in traditional learning activities [24]. There are two types of annotations used in foreign language teaching: learner generated and instructor provided. The first type is when students make annotations by themselves, trying to comprehend the text, and the second – when they obtain to analyze text with annotations prepared previously by the teacher. Each type of annotations brings benefits for students: annotating by themselves students keep in mind content or significant points; analyzing material prepared by instructor – they can work with comprehensive and correct material [25]. Glosses can have different functions. Written in the learner's first language can act as a prompting, highlighting or clarifying. For more advanced students they can have linguistic character and lexical or syntactical function [25]. It is easy to create annotations writing on the sheet of paper with a text while learning foreign spoken languages, but it is more complicated in the process of learning SLs while using video material. In this case multimedia annotations can be good option. What is important, multimedia annotations made using special software tools have positive influence on reading comprehension [24], [5]. As [5] and [7] claim, there are many advantages of Computer Assisted Language Learning (CALL) and teaching which has become an important part of the language learning process.

Young generation can be called generation of digital technologies era [7] thus the use of technology, including computers and Internet resources and capabilities can enhance process of foreign language learning, making lessons more interesting for students. From SL classes it is especially important, because students can find materials with native signers on popular video-sharing websites, that they know very well. Regardless of the foreign language that the student learns, there are many other advantages of CALL [7]:

- Facilitating completing repetitive language drills and practice by providing access to teaching materials at any time, without deadlines, what makes students the opportunity to work at their own pace and is beneficial for owing foreign language;
- Using computer as stimulus for students conversations, critical thinking and creativity or as a tool which helps in understanding of the language usage;
- Providing easy access to multimedia resources which are available on the Internet, access to authentic learning environment, possibility to share students' own materials, contact with native speakers;

Incorporation of technology in language classes is a challenge for teachers, who have to find and integrate new methods, tools and approaches in the learning process. But well thought-out use of computers for SL learning seems to be particularly important, because of the visual nature of language and video materials. One possible source of such materials are SL corpora.

Work [26] emphasizes the role of the SLs corpora in the process of teaching SL as a second language (L2). There is noted the problem of lack of standardized materials for SL teaching by what teachers rely on their intuition about the language what very often causes that language is overgeneralized or oversimplified and not based on naturalistic corpora. They describe few examples of differences between the principles taught by the teachers in SL classroom, and those observed after analysis of natural Australian SL corpus. Corpus analysis helps to provide more objective view of the SL. Work [26] lists four important steps that should be taken to incorporate corpus linguistic elements into SL classes: the first, foundational – refers to teachers knowledge, not intuition, about language; the second – refers to awareness of differences that may occur between intuition and knowledge; the third – refers to implementation of corpus based knowledge into classroom practice; the fourth – integration of corpus-based and teaching materials, including video clips with natural examples of discussed signs, vocabulary, grammatical structures.

There is large number of publications on the pedagogical benefits of the use of corpora for language teaching. According to [27] data-driven learning and the use of innovative corpus based tools cause learning gains in the transfer of vocabulary knowledge of learners. Incorporating these tools in language laboratory can bring benefits in improving students' awareness of words meanings by introducing new words in different contexts. While studying phonic foreign languages, using corpora linguistics elements help students not only in acquiring new academic vocabulary but also learning and use of verb-noun collocations, passive and active forms, linking adverbials. It also enhance students' performance in language activities. Work [27] notices that however there is vast range of research on the application of corpus approaches for language teaching, there is lack of implementation of corpus research findings in didactic materials. Author of [27] believes that this state will begin to change dramatically.

Hearing learners find it difficult to understand messages in SL. In the short survey carried out by authors of this article at the Institute of Pedagogy, Jagiellonian University in Krakow, students were asked about SL learning element of higher difficulty. The survey was conducted in February 2014 among 49 students from the 1st to the 4th year. Among them 40 were beginners in SL, just after 30 hours of SL course, the rest was more advanced, after 90 hours of lessons.

Students were asked to indicate the element in the process of SL learning which causes them more problems (choice was possible between understanding and creating statements in SL) and preferred method of SL learning (with or without video recordings). Nearly 84 percent of students assessed understanding of Polish SL statements as higher difficulty element of SL learning. What is very important, and was mentioned before, making annotations helps in understanding utterances in foreign

language. Respondents mentioned the following advantages of exercises based on detailed video analysis: a) contact with Deaf Persons' natural mode of communication, b) the opportunity to learn more about different Polish SL styles, including its regional variants, c) a better understanding of the role non-manual signals play in the process, together with the use of the SL space, temporal axis (temporal sequence of events), classifiers and links between various utterance parts, d) an attractive method to learn new signs. They also pointed out the advantages that stem directly from the use of technology, e.g. the ability to e) stop the recording or rewind it to watch its previous fragments and perform detailed analysis (step-by-step or movement-after-movement analysis), g) play the recording at a suitable pace, g) learn at home.

Taking into account recommendations formed by [23] and [22] and literature review as well as the students' response, authors developed the interdisciplinary concept of the incorporation of the professional AT, like the ELAN, to SL classes as a teaching aid which can support the process of understanding SL utterances as well as enhance the process of active learning [28].

In Table 3 there are presented advantages of the use of AT (on the example of ELAN) from the point of view of SL education process. The first important thing is that ATs described in the article can be used for free, students can install them on their own computers. Different features can be used during the classes depending on the students level of SL proficiency. Students can use this tool to:

- Add simple notes to video recordings with SL utterances, marking them in the right place on the timeline;
- Describe signs using simple notes, what can simplify the process of repeating new vocabulary;
- Indicate different grammatical elements in utterances;
- Indicate articulatory elements which are used to create different types of sentences: affirmative, questions, negations, exclamation;
- Indicate the ends of sentences in longer utterances;
- Find and mark differences between the same utterances or signs but in different dialects;
- Describe simultaneous articulatory elements using multiple tiers;
- Indicate interactions between manual and non-manual components;
- Make professional annotations using one of the annotation systems;
- Analyze annotated videos;

The ability to play the recording at a suitable pace as well as to stop it and return to the task after some time seems to be of particular importance to those who are at the early stages of SL acquisition. They can use ELAN to the simplest applications, adding simple notes at appropriate time points, or more sophisticated annotations (like these described in chapter 3), analyze in details simultaneous (tiers) as well as sequential character (temporal axis) of SL utterances. Exporting annotations into text files allows to compare annotations created by different students, and verify them by the teacher.

Table 3. The Advantages of the use of ELAN in SL Classes Associated with Different Features of the Tool [21]. Note: Items Marked with an Asterisk * Were Mentioned by SL Students in the Survey as the Advantages of the use of Technology in SL Classes.

ELAN feature	Advantage of its usage during SL classes
It is free	Possibility of installing on personal computers; learning at home, at one's own pace.*
Uses multimedia files	Adding text glosses to video recordings with natural dialogue in SL;*
Different types of annotations can be used	Making notes, adding comments, glosses to any visual feature observed in the recording.
Multiple layers of annotation (tiers)	Annotations of simultaneous features of SL;.
Hierarchically interconnected tiers	Hierarchy of annotations, describing both the general and specific features of signs or utterances.*
Synchronization of annotations	Help in understanding the temporal sequence of events (temporal axis).*
Multi-tier regular expression search and easy navigation	Easy search through glosses in the annotated materials.
Multiple speed settings of navigation	Possibility of detailed analysis thanks to stopping or rewinding the recording *
Up to 4 video files can be associated with an annotation document	Comparing of different SL styles, dialects simultaneously.*
Annotation results can be exported as text files	Possibility of printing, consulting text, make statistical analysis of annotations;
Used as AT in corpus linguistics worldwide	As an Inspiration, indicating the new directions of interest and research (linguistics, corpus linguistics, natural language processing).

IV. Conclusions

The process of SL acquisition requires considerable intellectual effort, willingness and commitment especially from hearing students who begin learning this language. While it requires great focus and powers of observation, it also exercises one's imagination, visual memory and the ability to associate ideas. Hearing students have to change their mode of perceiving language from sound to image and as they mentioned in our short survey, they

find the comprehension of SL utterances as the most difficult part of learning that language. Thus, it is necessary to arrange classes, select and apply teaching methods in such a way to make the understanding of signing easier for them.

As the group of people working on a project of ASLR system authors use ATs to describe recordings with SL utterances during preparation of SL corpus. This is very important to prepare proper annotations and prepare material for further analysis, it is very laborious and time consuming process as well, but necessary to broaden the

knowledge about this language. There are not many well annotated corpora of SLs, especially there is not available Polish SL corpus, thus authors think that it is needed to teach students how to create and describe such databases. There was developed an idea of incorporation professional ATs during SL classes to achieve this goal, as well as to facilitate students understanding SL utterances, not only in Polish, but in any other SL.

The concept is based on ELAN AT, which integrate video materials with different levels of annotation. Depending on the advancement of the student, it may be used as a tool for adding simple notes or for more advanced, annotations. The advantage of using AT as a teaching aid is that students are getting used to computer technology commonly used in corpus linguistics all over the world. Moreover, working on video recordings with native-signers provides a better opportunity for a direct contact with deaf people and their natural signing. It can be a pretext to discuss or consult annotations prepared by students during classes with native-signers as well as a good tool for raising awareness of the deaf people's way of thinking. Use of latest technology can be also an attractive method of learning new language signs.

Important advantage of the method is that it raises awareness of the Culture of Deaf and Hard of Hearing Persons, their values and tradition. When used in class, analyzed in detail and understood, video materials that represent the natural mode in which Deaf and Hard of Hearing Persons communicate with each other help students to appreciate the importance of human dignity and friendship and allow them to locate the Culture of Deaf and Hard of Hearing Persons.

Introducing corpus linguistics during SL classes can give the students new directions of interest and research. It can be also a chance to overcome barriers between hearing and deaf communities, giving both sides better access in information society.

REFERENCES

[1] S. C. W. Ong and S. Ranganath, "Automatic SL Analysis-a survey and the future beyond lexical meaning" *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 27(6) pp. 873—891, 2005.

[2] P. K. Pisharady and M. Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: A review," *Comput Vis Image Und*, vol. 141, pp. 152-165, 2015.

[3] H. Brugman and A. Russel, "Annotating Multimedia/ Multi-modal resources with ELAN". In: *Proceedings of LREC'04.* European Language Resources Association, Lisbon: 2004, pp. 2065–2068.

[4] P. Tomaszewski *Jak osoby słyszące mogą osiągnąć podstawy kompetencji komunikacyjnej w PJM? (How hearing can achieve basic PJM communication skills?)* Paper presented during the training for teachers at the Deaf Institute, Warszawa, 12 Dec. 2001.

[5] C. J. Chen and P. L. Liu "Comparisons of learner-generated vs instructor-provided multimedia annotations," *The Turkish Online J. of Educational Technology,* vol. 11(4), pp. 72—83, 2012.

[6] O. Crasborn and H. Sloetjes "Enhanced ELAN functionality for SL corpora" in: *Construction and Exploitation of SL Corpora. 3rd Workshop on the Representation and Processing of SLs*, Crasborn et al., Eds. Paris: 2008, pp 39—43.

[7] A. T. Dina and S. I. Ciornei "The Advantages and Disadvantages of Computer Assisted Language Learning and Teaching for Foreign Languages" *Procedia Soc. Behav. Sci*., vol. 76, pp. 248—252, 2013.

[8] S. Wallis "Annotation, Retrieval and Experimentation", in: *Annotating Variation and Change*, Meurman-Solin, Nurmi, Eds. Helsinki: Varieng, 2007.

[9] T. Hughes, K. Nakajima, L. Ha, A. Vasu, P. Moreno, M. LeBeau "Building transcribed speech corpora quickly and cheaply for many languages" in *Proceedings of the 11th INTERSPEECH 2010*, pp 1914–1917, 2010.

[10] V. Pitsikalis, S. Theodorakis, C. Vogler, P. Maragos "Advances in phonetics-based sub-unit modelling for transcription alignment and SL recognition", *Gesture* 11, pp. 1–6, 2011.

[11] C. Vogler and D. N. Metaxas "Toward Scalability in ASL Recognition: Breaking Down Signs into Phonemes", in: *Gesture Workshop*, Braffortet et al. Eds. vol. 1739, 1999, pp. 211–224.

[12] P. Rutkowski, S. Łozińska, J. Filipczak, J. Łacheta, P. Mostowski, "Korpus polskiego języka migowego (PJM): założenia – procedury – metodologia" (*PJM Corpus: assumptions – procedures - methodology*), in Sak (ed) *Stan badań nad głuchotą w Polsce*, PSG, Łódź. 2013.

[13] T. Johnston "Creating a Corpus of Auslan within an Australian National Corpus", in *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus. Cascadilla Proceedings Project*, Haugh et al. Eds. Macquaire University, 2008, pp 87–95.

[14] T. Hanke "HamNoSys – The Hamburg Notation System for SLs". http://www.aign-lang.uni-hamburg.de/dgs-korpus/tl-files/inhalt_pdf/HamNoSys_06en.pdf. Accessed 30 April 2014.

[15] M. Świdziński and D. Mikulska "Reprezentacja linearna teksu Polskiego Języka Migowego" (*Linear representation of the Polish Sign Language text*), in: Świdziński, Gałkowski (eds) *Studia nad kompetencją językową i komunikacyjną niesłyszących (Studies of lingustic and communication skills of the Deaf)*. Zakład Graficzny UW, Warszawa, pp 31–59, 2003.

[16] K. Lausz, "O linearności i symultaniczności w Polskim Języku Migowym" (*On the linearity and simultaneity in PJM*), in: Świdziński, Gałkowski (eds) *Studia nad kompetencją językową i komunikacyjną niesłyszących (Studies of lingustic and communication skills of the Deaf)*. Zakład Graficzny UW, Warszawa, pp 79–108, 2003.

[17] T. Johnston, "Auslan Corpus Annotation Guidelines. Centre for Language Sciences, Departament of Linguistics", Macquarie University, Sydney Australia. http://media.auslan.org.au/attachments/AuslanCorpusAnnotation Guidelines_Johnston.pdf. Accessed 22 April 2015.

[18] A. Partington, *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. John Benjamins Publishing Company, Amsterdam, 1998.

[19] T. Hanke "iLex – A tool for SL Lexicography and Corpus Analysis", in *Proceedings of the third International Conference on Language Resources and Evaluation*, Gonzalez et al. Eds. Vol. III. ELRA, Paris, 2002, pp 923–926.

[20] C. Neidle, "SignStream Annotation: Conventions used for the American SL Linguistic Research Project" http://www.bu.edu/asllrp/asllrpr11.pdf. Accessed 22 April 2015

[21] ELAN system Homepage at Max Planck Institute for the creation of complex annotations on video and audio resources. Retrieved April 22, 2015, from http://tla.mpi.nl/tools/tla-tools/elan/

[22] M. Garncarek, "Polski język migowy w nauczaniu słyszących: wskazówki metodyczne" (*Polish Sign Language in teaching of the Hearing: methodological clues*), in: *Język w komunikacji 2* (*Language in communication 2*), Majkowski et al. Eds. ADJ, Częstochowa, 2013.

[23] P. Tomaszewski, M. Bielak, P. Rosik, *Podstawy Polskiego Języka Migowego, Podręcznik do nauki Polskiego Języka Migowego dla kursantów (edycja robocza)* (*Basics of Polish Sign Language, Handbook for students*) Zarząd Główny Polskiego Związku Głuchych, Warszawa, 2009.

[24] W. Y. Hwang, C. Y. Wang, M. Sharples, "A study of multimedia annotation of Web-based materials", *Computers & Education,* vol. 48, pp. 680–699, 2007.

[25] W. B. Roby "What's in a gloss?", *Language Learning & Technology,* vol. 2(2), pp. 94–101, 1999.

[26] D. Cresdee and T. Johnston, "Using Corpus-Based Research to Infrom the Teaching af Auslan (Australian SL) as a Second Language", in: *Teaching and Learning Signed Languages: International Perspectives and Practices*, McKee et al. Eds. Palgrave Macmillan, 2014, pp 85–110.

[27] D. Biber, R. Reppen, E. Friginal "Research in Corpus Linguistics", in: *The Oxford Handbook of Applied Linguistics*, Kaplan Ed. Oxford: Oxford University Press, 2012, pp 548–567.

[28] M. Yahyazade, M. M. Moghaddam, A. Attaran, "Towards an Interactive EFL Class: Using Active Learning Strategies", *IJMECS*, vol. 6(5), pp. 8–14, 2014.

Authors' Profiles

Katarzyna Barczewska received her MSc degree in Biomedical Engineering from AGH University of Science and Technology in Kraków, Poland. Now she is a PhD candidate in the Department of Automatics and Biomedical Engineering at AGH. Her research interests include, machine learning, gesture recognition and human-computer interaction.

Jakub Gałka received his MSc and PhD degrees in telecommunications and electronic engineering from the AGH University of Science and Technology in Kraków, Poland in 2003 and 2008, respectively. Since then he has been with the Department of Electronics at AGH. His research focus lies in speech and language processing and recognition, speaker recognition, multimedia signal processing and data analysis.

Dorota Szulc, MA, completed postgraduate studies in Speech and Language Therapy. Currently she is a PhD student at the Institute of Education at the Jagiellonian University. Her interests focus on deaf higher education and sign language. Since 2007 she has committed to actions for the unhearing people. She works as a Polish Sign Language lecturer and interpreter.

Mariusz Mąsior received his MSc and Engineering degrees in telecommunications and electronic engineering from the AGH University of Science and Technology in Krakow, Poland in 2010. Since then, he has been an assistant professor-lecturer and a member of the Signal Processing Group at AGH University. He specializes in signal processing, speech technology, embedded systems, and systems engineering.

Rafał Samborski received his MSc degree in electronics and telecommunications from AGH University of Science and Technology in Krakow, Poland in 2009. Since then he is a PhD student at AGH. He co-authored several scientific publications on digital signal processing and data analysis as well as a European patent related to biometrics. Rafał is a co-founder and CEO of Consonace, a firm implementing R&D strategies for biomedical industry companies.

Tomasz Wilczyński, MSc eng., graduated from AGH University of Science and Technology in Krakow, Poland, and Cranfield University, UK. Nowadays, PhD student at AGH University Science and Technology in Krakow, Poland. He is focused on music acoustics, signal and image processing, and multimedia realization, also from commercial point of view.