Towards Czech on-line sign language dictionary – technological overview and data collection

Pavel Campr[1], Marek Hrúz[1], Jiří Langer[2], Jakub Kanis[1], Miloš Železný[1], Luděk Müller[1]

[1] Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Pilsen, Czech Republic
[2] Institute of Special Education Studies, Faculty of Education, Palacký University in Olomouc, Czech Republic
campr@kky.zcu.cz, mhruz@kky.zcu.cz, jiri.langer@upol.cz,
jkanis@kky.zcu.cz, zelezny@kky.zcu.cz, muller@kky.zcu.cz

Abstract

In this article we present the current state of our work on an on-line sign language dictionary. The aim is to create both an explanatory and a translation dictionary. It is primarily targeted (but not limited) to the Czech and Czech sign language. At first we describe technological aspects of the dictionary and then our data collection practices. The dictionary is an on-line application build with respect to the linguistic needs. We use written text to represent spoken languages and several representations are supported for sign languages: videos, images, HamNoSys, SignWriting and interactive 3D avatar. To decrease time required for data collection and publishing in the dictionary we use computer vision methods for video analysis to detect sign boundaries and analyze the manual component of performed sign for automatic categorization. The content will be created by linguists using both new and already existing data. Then, the dictionary will be opened to the public with possibility to add, modify and comment data. We expect that this possibility of on-line elicitation will increase the number of informants, cover more regions and makes the elicitation cheaper and the evaluation easier. Furthermore we prepare a mobile interface of the dictionary. The mobile interface will use different format of web pages and different video compression methods optimized for slower Internet connection. We also prepare an offline version of the dictionary which can be automatically generated from the online content and downloaded for offline usage.

1. Introduction

As for spoken languages, sign languages can utilize dictionaries for several purposes. Translation dictionaries are used to translate words (or phrases) from one language to another, explanatory dictionaries define the words in the same language instead of translating them. Traditional dictionaries for spoken languages use written text as main form for content creation. This becomes more difficult for sign languages where written form of the language is not so evolved and spread among the community. Examples of the written forms are HamNoSys (Hamburg Sign Language Notation System, developed in 1985) and SignWriting (developed in 1974). Advances in the field of information technologies allow creation of electronic dictionaries with new possibilities such as interactivity, faster searching, video animations, etc. The Internet brought new platform for on-line applications which opened other possibilities for the dictionaries: availability from anywhere, information sharing, interoperability and replaced static content with dynamic.
There are many existing on-line sign language dictionaries, but not all of them offer expected features, quality and content:

- Easy and intuitive usage
- Searching - not only by text, but by another, sign language specific criteria
- Complete data - to cover whole language, not only limited topics
- Being up-to-date
- Usage of the sign language written forms (HamNoSys, SignWriting, etc.)
- Linguistic information
- Version for mobile devices
- Offline version for download

Usually the existing dictionaries are specialized for selected topics or support only limited features. Our goal is to create state-of-the-art sign language dictionary which supports all mentioned features, is both translational and explanatory, and supports unlimited number of languages so that a dictionary entry (word or collocation) can be translated e.g. from the Czech sign language to the Czech language, English and American Sign language. Our dictionary is being developed now and our main content will be the Czech sign and Czech language.

spoken languages	sign languages
written text	HamNoSys
	SignWriting
	video
	image or illustration
	3D avatar

Table 1: Content forms supported by the dictionary

The dictionary supports several forms of content for sign languages as seen in the table 1. Along the common forms an interactive 3D avatar is available. It can perform the sign and the user can change angle of the view, zoom to a

detailed parts and slow down or pause the animation. Another innovation is usage of computer vision methods for video analysis to speed up video data collection and automatically categorize the signs into groups, which can be used as a criterion for searching.
This article is divided into four main parts: *Dictionary structure and content* describes, which data can be stored and viewed in the dictionary. In *Requirements for dictionary usage*, three different platforms (PC with and without the Internet and mobile devices) and their requirements are presented. *Searching* part describes possibilities of searching in the dictionary database. Finally, *Data collection* part introduces our practises for data collection and usage of computer vision methods.

2. Dictionary structure and content

Our on-line dictionary supports unlimited number of languages, the content can by represented by all forms listed in the table 1 and the dictionary entries can contain explanatory part and translations to other languages.
The functionality is based upon a database structure which is shown on the figure 1.

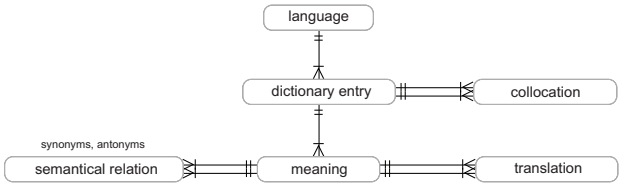

Figure 1: Database structure (simplified view)

Every dictionary entry is connected to one language. The entry can have one or more meanings. An explanation, use cases and linguistic information (e.g. part of speech) are attached to the meaning.
As collocations are widely used in the Czech sign language the database allows linkage between a word from a collocation to the dictionary entry, which corresponds to the standalone word. For example, collocation *personal baggage* is linked to *personal* and *baggage* entries.
Each of the meanings can be attached to another one (of a different dictionary entry) and express synonymic or antonymic relation.
The translation functionality is allowed by the linkage between two meanings which are linked to entries in different languages. This means that each of the meanings can be translated separately.
This database structure fulfils all linguistic needs for creation of explanatory and translation dictionary.
Every form specified in the table 1 is equal to another one and are optional (at least one of them is mandatory). Here we discuss some specific features:

video The dictionary entry can be represented by one or several video clips. They can be recorded separately (different speaker, place) or can be recoded simultaneously from multiple views (e.g. front, side and face view). The video data can be stored on the dictionary server or can be stored on external server anywhere on the Internet, including video sharing websites (Youtube, Vimeo, Dailymotion etc.). The video data will be available in more compression qualities and sizes, mainly for the usage on mobile devices.

image Multiple illustrations, photos or any other images with representation of a sign can be used.

HamNoSys (Hamburg Sign Language Notation System) The dictionary includes a special editor for HamNoSys strings which allows the users to create new or modify existing HamNoSys strings.

SignWriting Similarly to HamNoSys, special editor for SignWriting is required. This editor is being developed now and will be added lastly.

3D avatar Synthesis of the sign language creates a computer animation of the signing avatar (see fig. 2). For this purpose, we have specially created 3D animation model of the human figure.
For the web environment we had to convert the animation model to Collada format[1]. This format allows us to save 3D data, define the skeleton avatar animation and import the control trajectories.
An important part of the synthesis system is a conversion algorithm which converts a symbolic entry into the control instructions that are transmitted to the animation model. The entry of the algorithm is one or more signs noted in HamNoSys. The conversion algorithm was originally designed for the manual component of the sign language (Krňoul et al., 2008) and the version HamNoSys 3. We can convert not only isolated signs but the phrase, or continuous speech. The initial perceptual study shows good clarity of the animation of the manual component. The non-manual component was initially expressed by the visual speech, i.e. the articulation of words spoken language (mouthing). New extension of the conversion algorithm, however, allows transfer of the non-manual signals (NMS). For this purpose, a methodology for notation of NMS is designed (Krňoul, 2010). The notation of NMS is now a part of signs and the user is allowed to edit movements of torso, head and facial expressions.
An animation from the symbols has the benefits from the possibility of easy editing signs. The user can change the notation and determine the best form of the sign. One sign may be used for creation another sign with a similar form. The synthesis system provides two types of interactivity for the dictionary purpose. The first type is a preview of the figure. The animation model is rendered in the window and user can turn it in three axes or zoom the facial details, etc. Unlike video, which is always defined in one direction, the user can adjust it for best view. The animation is not in principle blurred or noisy. The second type of interactivity is phasing of the animation. User can suspend animation, re-run or step frame by frame. In particular, stepping allows the user to find ”an articulatory target”. The articulatory target is shape and position of the hand, body posture or facial expression that establishes meaning of the sign. From the educational point of view it allows the users quickly understand and learn new sign.

[1] OpenCOLLADA Framework, www.collada.org

Figure 2: 3D avatar. Left: front view. Right: face expression.

3. Requirements for dictionary usage

Primary platform for the dictionary usage is a PC connected to the Internet, with any modern internet browser (Internet Explorer version 6 and above, Firefox, Opera, Chrome, etc.) with installed Adobe Flash plugin. To enable hardware accelerated, high quality 3D avatar animation, a special plugin (Google O3D) is required. Without this a lower quality 3D avatar is used, without hardware acceleration. In the future, WebGL (new specification for writing web applications utilizing hardware accelerated 3D graphics) can be used.

Secondary platform is a mobile device (PDA, smartphone, etc.) with installed internet browser. For this platform the dictionary will be formatted with respect to the device capabilities and the video clips will be resized and compressed for the needs of those devices.

Another secondary platform is a PC in the same configuration as above but without internet connection. The dictionary will be able to automatically create offline version, which will be automatically created every day and available for download. This offline version will be limited in functionality in comparison to online version, mainly in searching capabilities.

4. Searching

Key feature of the dictionary is searching. The goal is to create searching functionality which will provide relevant results for user query. For spoken languages the user provides searched term, language and optionally a topic and grammatical information (e.g. part of speech). Result is a list of dictionary entries which satisfy the given search criteria. For the Czech language a lemmatization engine is used to enable searching among different inflected forms of same words. Furthermore, the searching is not limited only for dictionary entry title but provides fulltext search in all text items (meanings, explanations, use cases etc.).

For sign languages the searching feature is more complicated since the sign language words and sentences aren't represented in text form and thus we cannot use tools used for text searching. Our goal is to examine possibility of HamNoSys and SignWriting usage as search criteria and find a way how to find related dictionary entries for the given criteria. Because we expect that the resulting list for this way of searching will contain many items, other criteria can be used to limit the search as for text search (topic, grammatical information).

5. Data collection

The content of the dictionary will be continuously extended and modified. For this purpose a special administration section is available where the users can (depending on their permissions) create, update or delete dictionary entries. Special workflow management is prepared for administration users with limited permissions, where all modifications must be confirmed by administrators with full permissions. Thus the quality of the content is preserved with the possibility for many users to edit the content. The workflow can be easily changed after we get some experience after the dictionary is be released.

The decision if a new or updated dictionary entry is valid will be supported by a discussion under each dictionary entry, where the community can decide, whether the provided information is correct.

Most of the dictionary entries will be provided by professional linguists. The process of video recording is quite time consuming, to reduce the required time we use several tools, such as automatic detection of sign boundaries in a recorded session.

5.1. Utilization of computer vision methods

We use computer vision techniques to automatically detect boundaries of signs in a recorded session. There are certain conditions that need to be met in order to successfully obtain the boundaries. There should be a neutral pose of the signing person. This pose defines the beginning and the end of the sign. Also, the stage where the person is signing should have laboratory-like conditions so that the hands of the person are clearly visible and easily distinguishable from the background. Since the intention of the recordings is to use them in a SL dictionary these conditions are rational.

5.2. Sign boundaries detection

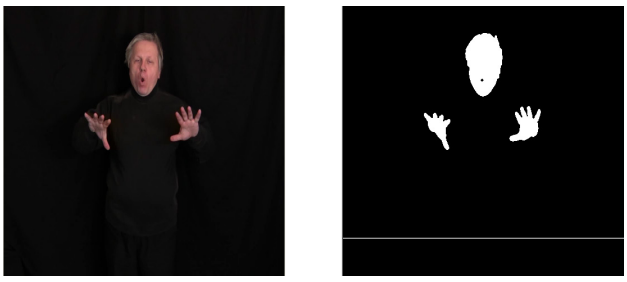

Figure 3: Video file processing and segmentation.
Left: original frame from a video. Right: segmented image, white parts correspond to head and hands.

We detect two features: motion and position. First, the image is pre-processed and segmented so that we obtain parts of human body. In some cases a simple thresholding can be used (e.g. the signer wears dark clothes), see fig. 3. In

more complex situations when the brightness level of pixels is not enough to distinguish between parts of human body and the rest of the scene, we use skin color segmentation. Next, we use object detection in the segmented image. We compare the position of objects (hands and head) with the trained initial position. If the distance is below a threshold we assume the signer is in the initial pose (fig. 4). In some cases we do not need to compute the distances but rather examine the position of the object and check whether it is in some predefined region. This is just an alternative approach with the same mathematical foundation.

In the next step we describe the movement as the sum of pixels in the difference image. This does not give us a detailed description of the movement but rather an estimate of total movement in the image. This value is normalized by the resolution of the image. A threshold is set and when the relative motion in the image is above this threshold we assume there is a significant movement of hands (fig. 4).

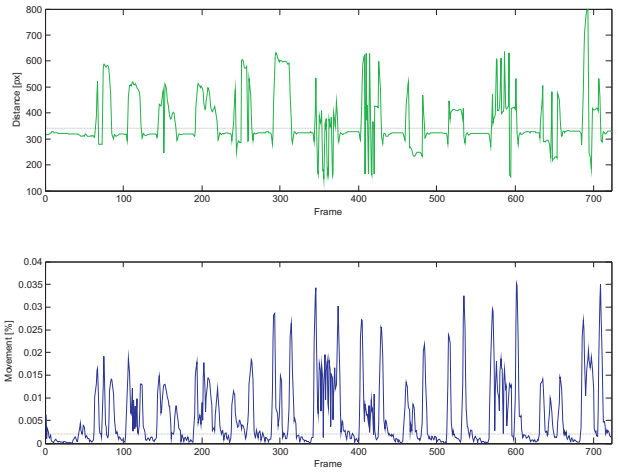

Figure 4: Features used for sign boundaries detection: motion and position

The features of movement and initial pose are measured over the recording. The first frame in which the neutral pose is not detected and movement is detected is considered as the beginning of the sign. Respectively, the first frame in which the neutral pose is detected and no movement is detected is considered to be the end of a sign. We have to shift the boundaries of the detected sign a bit so that the resulting cuts begin and end in a stationary pose. Usually we use the value of +/- 50 ms.

5.3. Automatic processing of signs

According to work described in Hrúz et al. (2008) we are able to track hands and head in recordings designed for sign language dictionary. For now we are able to obtain the trajectories of both hands and the head. On a relatively small dataset (Campr et al., 2007) we achieved good recognition results (Trmal et al., 2008) with features describing the manual component of SL. It is a baseline system and the features can be used for the annotation of a portion of manual component in the desired form. One of our goals is to develop a new system capable of describing the manual component in more detail. Based on that we can automatically group similar signs and utilize this information for searching purposes.

6. Conclusion

In this article we presented our progress on ongoing project *Czech on-line sign language dictionary*. Some parts of the dictionary are nearly finished (database system, administration interface, 3D avatar, video players, fulltext search), other parts are being developed (sign search, SignWriting editor, frontend interface). We expect the first public release (and first feedback from the users) in the second half of the year 2010, but some limited pre-release versions are already available.

Our goals are both to create high-quality on-line sign language dictionary system and to provide high-quality content for the Czech and Czech sign language.

7. Acknowledgement

This research was supported by the Grant Agency of Academy of Sciences of the Czech Republic, project No. 1ET101470416, by the EU and the Ministry of Education of the Czech Republic, project No. CZ.1.07/2.2.00/07.0189, by the Ministry of Education of the Czech Republic, project No. ME08106 and by the Ministry of Education of the Czech Republic, project No. MŠMT LC536. We would like to thank Dr. Petr Peňáz, Masaryk University, Brno, for helpful discussions about linguistic aspects of the dictionary.

8. References

Pavel Campr, Marek Hrúz, and Miloš Železný. 2007. Design and recording of signed czech language corpus for automatic sign language recognition. *Interspeech*, pages 678–681.

Marek Hrúz, Pavel Campr, and Miloš Železný. 2008. Semi-automatic annotation of sign language corpora. *LREC Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 78–81.

Zdeněk Krňoul, Jakub Kanis, Miloš Železný, and Luděk Müller. 2008. Czech text-to-sign speech synthesizer. *Machine Learning for Multimodal Interaction, Series Lecture Notes in Computer Science*, 4892:180–191.

Zdeněk Krňoul. 2010. New features in synthesis of sign language addressing non-manual component. In *4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*. ELRA.

Jan Trmal, Marek Hrúz, Jan Zelinka, Pavel Campr, and Luděk Müller. 2008. Feature space transforms for czech sign-language recognition. *Interspeech*, pages 2036–2039.