Title of the Article: *A Methodology for Automatic Sign Language Dictionary Creation*
The Name of the author(s): *Jakub Kanis*[1], *Petr Peňáz*[2], *Pavel Campr*[1], *Marek Hrúz*[1]
Affiliation (University, City, Country): *(1) Faculty of Applied Sciences, Department of Cybernetics, University of West Bohemia, Pilsen, Czech Republic, (2) Teiresias Centre, Masaryk University, Brno, Czech Republic*
Keywords:
E-mail: *(1) jkanis, campr, mhruz@kky.zcu.cz, (2) penaz@teiresias.muni.cz*

Abstract: In this article we present the the sign language dictionary being developed by a research team of University of West Bohemia, Masaryk University and Palacký University. The aim is to create both an explanatory and a translation dictionary with respect to the linguistic needs. The dictionary is designed as an on-line application with access for all registered users who compose a hierarchical structure to ensure a lexicographic consistency of the dictionary content. The basic structure of the dictionary is proposed to meet linguistic criteria of both spoken and sign languages. We use written text to represent spoken languages and several representations are supported for sign languages: videos, images, HamNoSys, SignWriting and an interactive 3D avatar. To decrease the time required for data collection and publishing in the dictionary we use computer vision methods. We detect the boundaries of the dictionary entry. If the entry is a single sign we analyze the manual component of the performed sign for automatic categorization. The content is being created by linguists using both new and already existing data. Later on the dictionary will be opened to the users 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 make the elicitation cheaper and the evaluation easier.

1 Introduction

A dictionary is an important source of information about given language. It holds especially in case of sign languages, where exists no official written form and dictionaries or lexicons are one of main sources for teaching and linguistic research of sign languages.
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
- Completeness of lexicographical data, to cover most of the language, not only limited topics
- Being up-to-date
- Usage of the sign language written forms (HamNoSys, SignWriting, etc.)
- Version for mobile devices
- Off-line 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.
Because adding of new signs into the dictionary is usually a time consuming and difficult task. We introduce in this article a methodology for automatized dictionary creation to speed up this process. The methodology is based on using of advanced computer vision techniques.

2 Dictionary Structure and Content

2.1 The linguistic principles of the dictionary

The dictionary combines three main principles:

- Full autonomy of the data for each language (the signs are not considered equivalents to some words and primarily, a sign should be explified by means of a sign language).
- Philological exactness and completeness (using the same lexicographic procedure for signs as normally used in thesauri online such as Merriam-Webster, Duden, Garzanti, Tresor de la langue francaise etc.).
- Semantic synsets (groups of synonyms) as conceived by George Miller and Christiane Fellbaum et al. for the well-known WordNet structure[1].

In such a way, each language represents an independent lexicological network connected by synsets and collocations, and these independent networks (lexica) are freely connected to each other just by means of identical semantic features of their lexical items. Each lexical item (entry, lemma) is described by means of two groups of fields:

Formal description

Presentation of the lemma form:

- alphabetic text (standard spelling as primary word presentation in spoken language, alphabetic transcription of a sign if such a system exists)
- audio/video presentation (one or more audio or video formats to present pronunciation and/or mouthing in a spoken language; signing in a sign language)
- analytic description of the acoustic or visual reality (IPA transcription in spoken language, SW or HamNoSys in a sign language etc.)
- verbal description of the acoustic or visual reality (text comments to pronunciation or mouthing, verbal description of signs)

Classification of the lemma form:

- part of speech

[1] http://wordnet.princeton.edu/wordnet/

- grammatic categories
- stylistic categories

Semantic description

Presentation of the meaning:

- verbal description of the meaning (semantic features) with examples for each semantic unit
- synsets: hypernyms/hyponyms, antonyms/synonyms, coordinate terms
- collocation, phrases with examples for each syntactic unit

Classification of the meaning:

- semantic categories

Organisational and legal description:

- performer: who filled in the data
- authorisation: who is the author of or responsible for the data
- history: when which data have been entered or cancelled

2.2 Content Management

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 work-flow 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 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.

3 Methodology for Automatized Dictionary Creation

3.1 Scene Description

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.

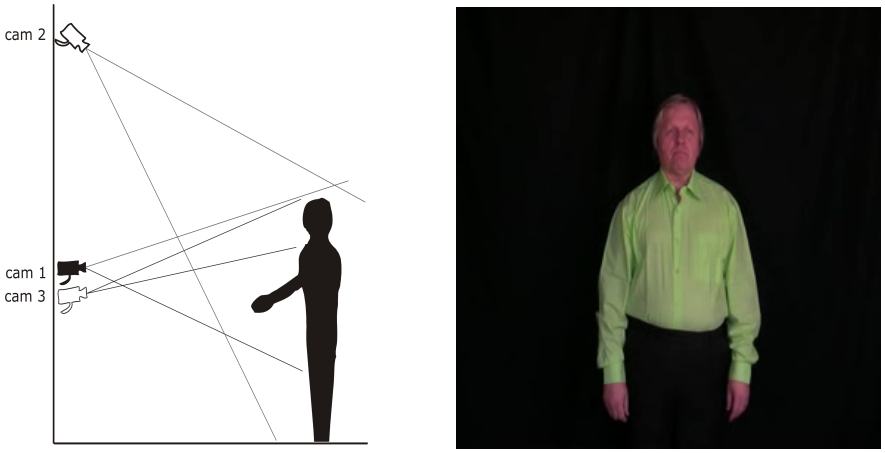

Figure 1: Left: An example of a configuration of cameras. Right: Example of the laboratory-like conditions and an initial pose.

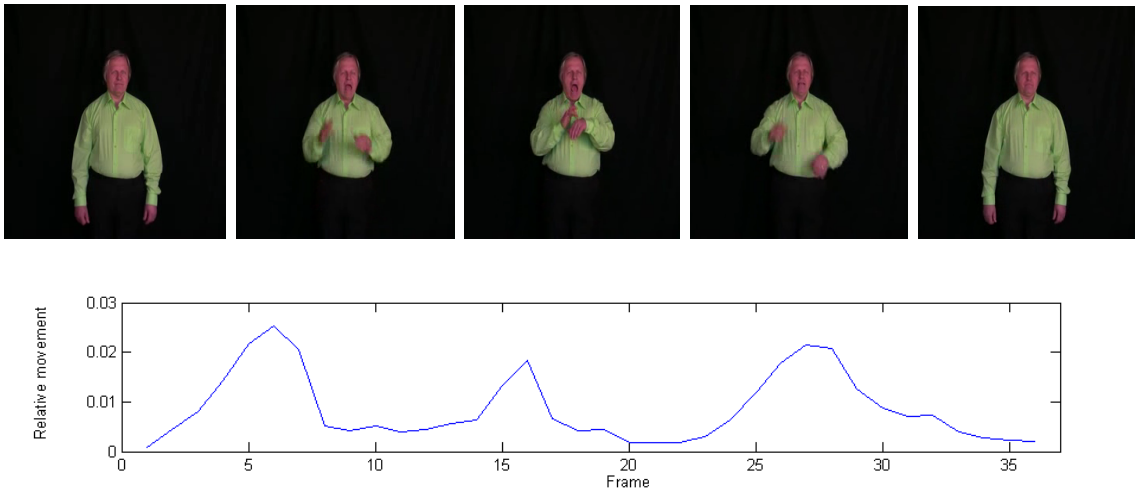

Figure 2: Top row: selected frames from a sign video. Bottom row: relative movement through the sign video.

3.2 Sign Boundary Detection

We detect two features: motion and position. First, the image is preprocessed 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). 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 (Figure 1). 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

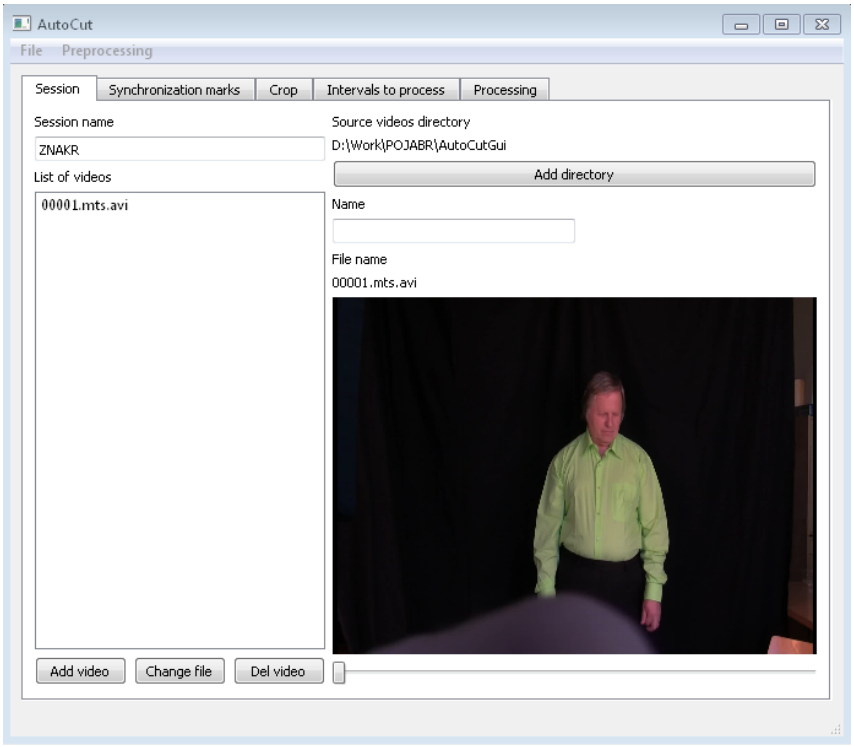

Figure 3: AutoCut main window.

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 (Figure 2).
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.

3.3 AutoCut

To speed up the process of video recordings we made a specialized program. The program was developed with aim to take the user trough the whole process of videos from preprocessing to made the recorded signs ready to publish. The first step of video processing is conversion of recorded videos to the video format appropriate for the next processing. This can be done by user in simple dialog with file managing and running of the preprocessing. The main window (see Figure 3) of the program has five cards in the order of usual processing of chosen videos. On the first card the user can add, change or delete source video files (made in preprocessing step). On the second card a synchronization mark, which is important for processing of video records from different point of views, can be set. The next card is for the crop adjusting and the third for choosing the intervals to process. The last card is for the automatic sign boundary detection, adjusting of the skin color model and final generation of particular signs for publication.

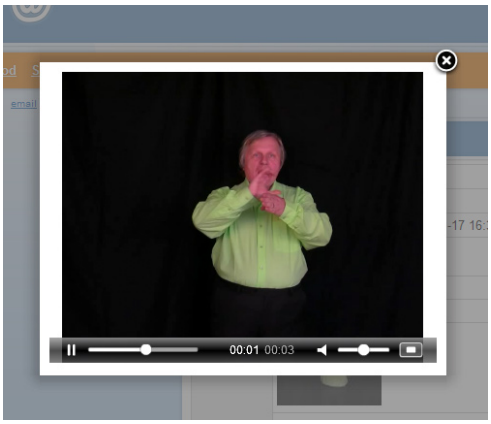

Figure 4: Video player showing a sign.

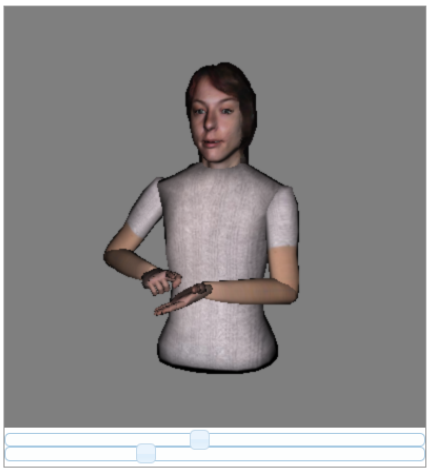

Figure 5: 3D avatar displayed in the web page.

4 Sign Language Dictionary On-line

The dictionary is a joint venture of University of West Bohemia, Masaryk University and Palacký University, accessible at the following URL's: *http://slovnik.zj.teiresias.muni.cz/* and *http://znaky.zcu.cz* . 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.
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 next platform is a PC in the same configuration as above but without internet connection. The dictionary will be able to automatically create off-line version, which will be automatically created every day and available for download. This off-line version will be limited in functionality in comparison to on-line version, mainly in searching capabilities.

4.1 Sign Language Representations

video The dictionary entry can be represented by one or several video clips (Figure 4). 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.

3D avatar Synthesis of the sign language creates a computer animation of the signing avatar, [1] (Figure 5). For this purpose, we have specially created 3D animation model of the human figure.

image Multiple illustrations, photos or any other images with representation of a sign

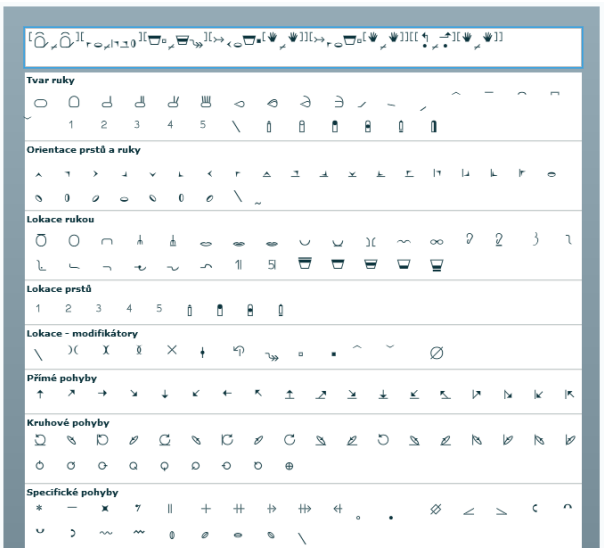

Figure 6: HamNoSys editor.

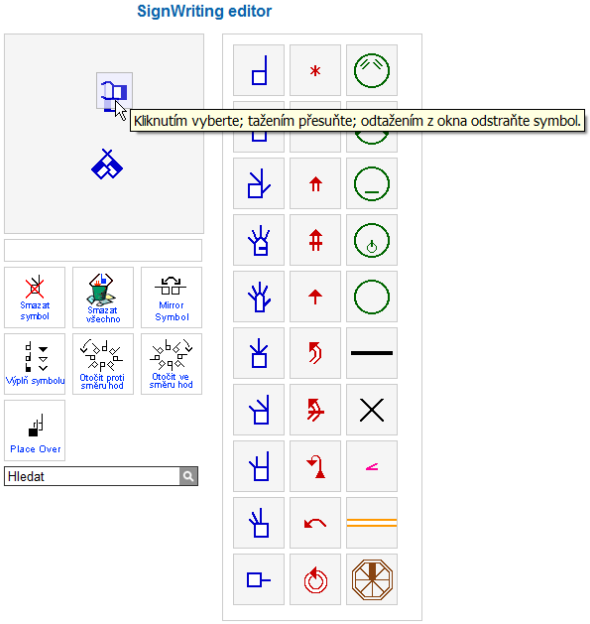

Figure 7: SignWriting editor.

can be used.

HamNoSys The dictionary includes a special editor for HamNoSys (Hamburg Sign Language Notation System, [2]) strings which allows the users to create new or modify existing HamNoSys strings (Figure 6). In this time the Hamnosys notation of each sign is crucial for automatic synthesis of sign language.

SignWriting Similarly to HamNoSys, special editor for SignWriting is included, [3] (Figure 7).

4.2 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 there are basically two ways to enter data for searching: text data via standard spelling (possibly in combination with grammatical and stylistic or semantic information), or acoustic data via voice. A specific combination of both ways is represented by searching based on phonetic transcription (usually intuitive one, rather than an exact system). 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 necessarily limited only for dictionary lemma form but may provide, if needed, full-text search in all text items (meanings, explanations, use cases etc.).

In the case of sign languages both above-mentioned ways are complicated by the lack of standards: there is usually no system of text transcription, there is no standard sign notation, and not enough visual data to use a camera and motion capture as a standard searching technology. 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

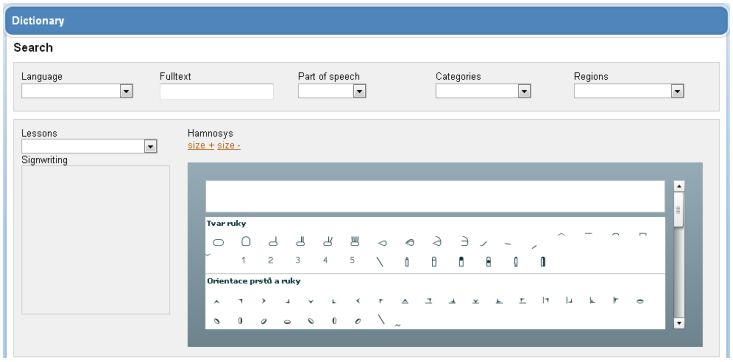

Figure 8: Searching form.

many items, other criteria can be used to limit the search as for text search (semantic and grammatical information) (Figure 8).

4.3 Automatic Processing of Signs

According to work described in [4] 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 (Figure 9). On a relatively small dataset we achieved good recognition results with features describing the manual component of SL. The features can be used for the annotation of a portion of manual component in the desired form. From the signal of trajectories of both hands we are able to determine whether the sign was symmetric or not. We compute Pearson correlation coefficients between the x and y signals of the trajectories. The absolute value of the coefficients determines the probability of the sign being symmetric. The trajectory can be also utilized to determine the location of the sign, the type of movement and the contact between body parts. These are work in progress. The trajectory also gives us information about the position of hands in every frame. Thus, we are able to obtain an image of the hand in every frame. This image can then be used for hand shape recognition. This task is very difficult and up to now is not sufficiently solved. Our approach will take into account only 27 hand shapes thus the problem becomes better conditioned. Based on this information we can automatically group similar signs and utilize this information for searching purposes.

5 Conclusion

The sign language dictionary which offers so many options and meets all the above-mentioned criteria is a complex and prestigious goal that will need co-operation of several teams on the international level. We hope that it can be achieved in spite of the fact that number of native signers is evidently decreasing (due to several reasons) and it is not easy to build up a professional team dedicated to sign linguistics.

6 Acknowledgment

We would like to acknowledge all the colleagues who are developing the on-line dictionary: Zdeněk Krňoul, Miloš Železný, Luděk Müller, Svatoslav Ondra, Jan Fikejs, Vojtvech

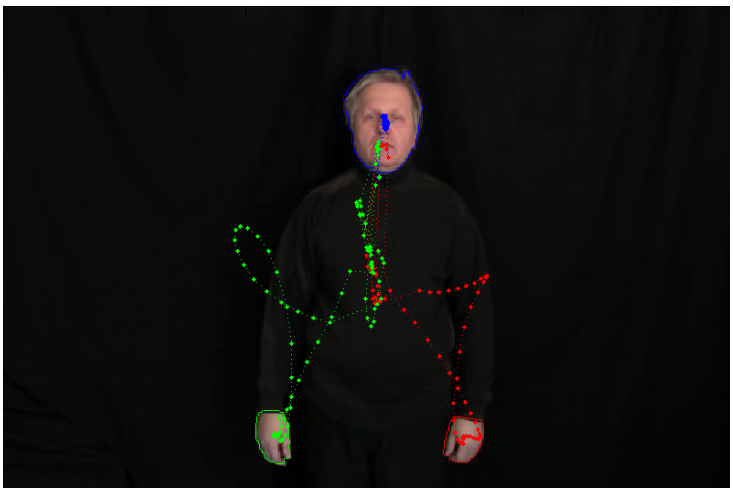

Figure 9: Example of tracking result.

Knyttl and Jiří Langer. 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 Grant Agency of the Czech Republic, project No. GAČR 102/09/P609.

References

[1] Zdeněk Krňoul and Jakub Kanis and Miloš Železný and Luděk Müller. Czech Text-to-Sign Speech Synthesizer. *Machine Learning for Multimodal Interaction, Series Lecture Notes in Computer Science*, 2008, pages 180–191.

[2] Prillwitz et al. HamNoSys. Version 2.0. Hamburg Notation System for Sign Language. An Introductory Guide. *Signum*, 1989.

[3] Amy Rosenberg. Writing Signed Languages In Support of Adopting an ASL Writing System. *University of Virginia*, 1995.

[4] Jan Trmal, Marek Hrúz, Jan Zelinka, Pavel Campr, and Luděk Müller. Feature Space Transforms for Czech Sign-Language Recognition. *Interspeech*, 2008, pages 2036–2039.