Towards Russian Sign Languages Corpora

Alexander Voskresenskiy
NGO "International Academy of Sciences on Information and Information Processes and Technologies"
21, F. Engelsa str., Moscow, Russian Federation
+7 (499) 265-23-93
avosj@yandex.ru

ABSTRACT

In this paper, we describe our approach to building Russian Sign Language corpora.

Categories and Subject Descriptors

D.2.2 [**Design Tools and Techniques**]: *User interfaces*

H.3.1 [**Content Analysis and Indexing**]: Dictionaries, Indexing methods, Linguistic processing

General Terms

Algorithms, Design, Experimentation, Human Factors.

Keywords

Russian Sign Language, Corpora, Tokenization, Lemmatization, Search of Gestures.

1. INTRODUCTION

This paper describes approaches to building Russian Sign Language Corpora. It was assumed that its creation will begin in 2011, but for reasons independent from the developers start is postponed. We hope that this work will begin next year.

Linguists have considerable experience in creating corpora for verbal language. In the case of sign language corpora it is much more difficult task. This paper identifies the problems encountered when trying to create sign language corpora different from the case of verbal language.

2. THE PURPOSE OF THE CORPORA

One of the problems of interaction between deaf and hearing people is the difficulty of communication. This problem leads, for example, to the fact that many Deaf do not go to doctors, in particular to psychologists and psychotherapists. Facilitating communication could improve the quality of life for deaf people, enabling them to better adapt to society of hearing people.

For Deaf people, mastering the skills of visual perception of speech (lip-reading) and pronunciation can not be considered a solution of the problem due to the fact that the establishment of the correct pronunciation in the absence of control by the ear is an extremely difficult task. There is need to recognize that pronunciation of prelingually deaf people, elaborated as a result of training, far from the average quality of pronunciation of hearing people. At the same time forcing deaf people to use only verbal speech (which for some of them for objective reasons is practically impossible) is a violation of their human rights.

These ideas are embodied in the Convention on the Rights of Persons with Disabilities [2], adopted by the UN General Assembly on 13 December 2006 and came into force on 3 May 2008. The Russian Federation acceded to the Convention on the Rights of Persons with Disabilities on September 24, 2009. This means that Russia agrees to continue its policy to adhere to the provisions of the Convention, including, for example, Art. 21, which states that:

> *"States Parties shall take all appropriate measures to ensure that persons with disabilities can exercise the right to freedom of expression and opinion, including the freedom to seek, receive and impart information and ideas on an equal basis with others and through all forms of communication of their choice, as defined in article 2 of the present Convention, including by: ...*
> *b) Accepting and facilitating the use of sign languages, ... in official interactions; ...*
> *e) Recognizing and promoting the use of sign languages."*

In this regard, it becomes an urgent requirement for proficiency in sign language by all those who communicate with a deaf person. For this training, manuals must to be developed on sign language especially for health professionals, law enforcement officers and judicial system, municipal employees.

One of the problems of deaf students is a lack of understanding of textbooks in various subjects[1]. The reason is that the active vocabulary of deaf students includes significantly fewer words than their hearing peers. In addition, understanding of semantic differences between different morphological forms of the tokens of the Russian language is complicated. This is due to the fact that the deaf students' verbal Russian is their second language [11]. This complicates the understanding of speech and text in Russian. At the same time, good knowledge of Russian spoken language for the deaf is the primary means of mastering skills and achieving positive socialization.

Neuropsychological studies suggest that the activity of brain mechanisms for the functioning of both verbal and sign language built on the same principles. It was shown that for deaf children a good command of sign language facilitates the study of verbal language.

[1] This problem is observed not only in our country. For example, in USA the median reading level of deaf high school graduates is fourth grade [3].

Thus, the development of training manuals in Russian Sign Language (RSL) and RSL dictionaries is a prerequisite for achieving the desired quality of life for deaf people. We hope that the RSL corpora will be a basis for the development of such manuals and dictionaries.

3. COLLECTION OF TEXTS

Corpora studies are usually conducted on verbal language, with a significant number of written texts. Currently, the researchers were able to form corpora to its tasks, using the resources on the Internet. Setting the appropriate requests to search engines, such as Google, it can be relatively easily to retrieve texts of some domain and form a corpus of a few million words and more. In this case the search engine automatically removes duplicates of web pages from the output.

For RSL this is not possible. This language has no written form and is not properly described. In fact, about 30 years ago, its existence was practically denied. As a means of communication with deaf persons, Signed Russian (SR) is used in official events, which is not a proper language but a system for encoding the words of spoken Russian by gestures.

Until now, vocabularies of Russian signs are a mixture of RSL and SR gestures, with no labels, to isolate, for example, signs that are used only in SR or only in RSL. The analysis shows that even the dictionaries, which the authors argue that these dictionaries are genuine RSL dictionaries, contain about 30% gestures that belong to RS.

The total number of Russian signs described in dictionaries (the total number of dictionaries produced in the USSR and Russia does not exceed 10), is about 5000, including various versions of the same signs. We compared the contents of the three dictionaries which were published in Moscow, St. Petersburg and Dnepropetrovsk with the dictionary published in Moscow whose authors argued that in this dictionary only RSL signs were included. Matches in the content of dictionaries (same signs for identical concepts) are between 30 and 58%.

But how representative are these dictionaries? Long-term studies of verbal language helped to create frequency dictionaries, illustrating the frequency of use of certain tokens in texts of various domains. Dictionaries of Russian signs were created in most cases by employees of schools for deaf children. Therefore, these dictionaries contain the signs used in the living conditions of the residential schools. In terms of content, they cannot be compared with dictionaries of Russian, containing the most frequent words. For example, many morphological analyzers for Russian are made on the basis of the grammatical dictionary of the Russian language, which was created by the Academician A.A. Zaliznyak [10] and contains more than 100000 frequently used words. However, when creating a multimedia Russian Sign Language dictionary RuSLED [9] (containing about 2500 entries), we needed to create our own morphological analyzer, because the word lists of RuSLED and Zaliznyak's dictionaries did not match. If these discrepancies were related only to terms that have emerged after the creation of the Zaliznyak's dictionary (eg, *fax, video, Internet*), then this would not be worth mentioning. However, these discrepancies include such words frequently used in the child speech and school life as *take, festivities, children, hole, uncle*, etc.

At this time, we cannot now judge the representativeness of the existing dictionaries of the Russian sign language. Accordingly, it is difficult to assess the representativeness of our sign language corpus, at least in the initial stages of its creation.

4. REPRESENTATION OF SIGNS

Obviously, printed images of signs do not always provide the necessary information about the performance of a sign that occurs in space and time. Electronic dictionaries, in which video is used for the sign demonstration also does not always give a complete picture of the performance characteristics of the gesture because of the loss of spatial information in a flat image captured by one camera.

We need to use 3-D animation to demonstrate signs. A three-dimensional model allows us to rotate the image for viewing at any angle. As a result, images of signs can be seen from the side, top, rear, which provides capabilities not achievable by other means of visualization and very useful in the study of signs as well as in sign language training. However, the means of creating such animations are quite time consuming and not always ensure the necessary quality.

For example, in the 90s the company Vcom3D (Orlando, Florida, USA, http://www.Vcom3D.com) worked on a dictionary of American Sign Language (ASL). The company created the gesture editor Gesture Builder, which allows you to edit the position of the hands in space for key frames in the animation of the gesture and record information about them on the timeline (basic configuration, the coordinates in space, as well as the accompanying gesture facial expression). This information, together with data on the execution time of individual movements, is stored in a script file, which can be played in the program Sign Smith Studio. However, despite the existing library of core configurations, editing gestures is difficult. In addition, this model provides the American English articulation only, so that can not be used to create sign language dictionaries for other languages.

In 2003 – 2005, the European Union project eSIGN (Germany, UK, Netherlands, http://www.sign-lang.uni-hamburg.de/esign) developed several avatars. Control of the avatars is achieved using hand written scripts created in the special notation (HamNoSys – Hamburg notation system). This process is time-consuming.

Created by the same procedure at the University of West Bohemia (Plzen, Czech Republic) and modified for the demonstration of Russian speech at the Institute SPIIRAN (St. Petersburg, Russia), an avatar used in the project "Information kiosk" [6] has the same disadvantages. In addition, our tests carried out jointly with the Czech colleagues have shown that the HamNoSys notation is not capable of describing some Russian signs, such as *infection* (*infectious, contagious disease*).

The motion capture [1] method, planned in our project to create 3-dimensional animation to show the sign avoids these shortcomings. Movement of three-dimensional model fully consistent with the movements of a human demonstrator, the phenomenon of "piercing" model by hands, observed in some cases for the movements of the avatar that is managed by a notation HamNoSys, is absent. It should be recognized that

[1] http://en.wikipedia.org/wiki/Motion_capture

recording signs in the form of three-dimensional animated models using motion capture has capabilities superior to other methods.

Experiments show that three-dimensional animated demonstration of gestures, recorded by motion capture techniques, has high quality images of hand movements, body and facial expressions.

The disadvantage of motion capture is its high cost.

5. TOKENIZATION AND LEMMATIZATION

A set of texts in one language collected in accordance with certain principles, which are tagged in accordance with a certain standard and provided with a specialized search engine is called a linguistic corpus. The primary tagging of texts includes the steps required for each case:

- tokenization (splitting into words);
- lemmatization (reduction of word-forms to the dictionary form);
- morphological analysis.

In contrast to existing corpora of verbal language (eg Russian National Corpus, www.ruscorpora.ru), in which lemmatization and morphological analysis are performed on the basis of certain rules of language, we don't know which morphological rules on Russian Sign Language can be used in a sign language corpus. To define these rules, the corpus of Russian Sign Language should include:

- Parallel texts in sign and verbal Russian language, serving to define the semantics and syntax of sign utterances;
- Source dictionary of signs, which serves to determine the modifications of the various signs in the utterances to identify the morphological rules of sign language.

The words in texts of spoken language stand out with spaces and punctuation. But sign language does not contain pauses between individual gestures. Only phrases are separated by pauses. This introduces additional complexity in the implementation of tokenization, resembling those that occur in developing systems of recognition of continuous speech sound. In addition, many gestures are composite, contain a combination of a few gestures and/or signs of finger spelling before the gesture modifying its meaning.

The number of compound signs is large, but statistics give early. This is due to the relatively small volume of available sign language dictionaries, which can lead to a significant shift of the statistical evaluations.

Given the composite nature of the signs, utterance tokenization should be made by choosing from the dictionary the appropriate sign phrases that have the greatest length, and then analyzing the semantics of the resulting expression. If the value does not match the discourse, we can begin to alternately splitting "long" phrases onto "shorter" ones, trying to get a expression the content of which corresponds to the discourse.

Thus, if the verbal text tokenization in most cases is a formal procedure that requires no analysis of the semantics of the utterance, tokenization of sign utterances must be based on semantics. Automated tokenization of sign utterances will be possible when sign dictionaries are large enough (how many entries are needed?) and support semantically driven tokenization.

Lemmatization of verbal text is used to retrieve texts that include your search words, regardless of their grammatical form. In the index files all the words have their original form - lemma. The search query is processed by the morphological analyzer, resulting query words obtain its lemma shape. This allows you to find texts in which search words are in various grammatical forms.

Among the Russian Deaf it is a common belief that signs do not change in the communication process, but retain the same form as that embodied in the dictionary. Thereby the need for stemming in the case of sign language is denied. However, we have observations that show the presence of changes of signs in utterances. In addition, this fact was confirmed by French investigators [1].

Much attention should be given to this, because changes applied to signs will affect the search in the dictionary when implementing automatic translation of sign utterances. A corpus of Sign Language should include the allowable options of a given sign and their possible changes in the utterance. If we can identify patterns of such changes, it will be possible to develop a software tool - in analogy to the morphological analyzer used when working with the verbal text.

6. FEATURES OF THE CAPTURE AND ANALYSIS OF SIGN UTTERANCES

Shooting of gestures by motion capture is carried out in a studio, which affects human behavior. Informants are in a more assembled state, trying to show how signs can be "correct" and "beautiful." This can be useful when creating a vocabulary of signs, but leads to a loss of naturalness. In these circumstances it is unlikely to find out the changes of signs in utterances. Shooting of signs in a natural environment of the informant (perhaps when he/she did not know at what point of time there is a shooting) may give additional information useful to the researcher.

But in this case we can not use motion capture equipment; a usual camcorder can be used only. The question arises as to convert these two-dimensional images to a three-dimensional model for an avatar. This transformation is needed at least two reasons:

(i) The use of avatars to avoid publishing personal data (image of the person who demonstrated the signs);

(ii) To generate sign utterances from the text, avatars can also group into one utterance signs from the dictionary originally shown by different informants without violating the rules of unity of action, known since the time of the ancient theater.

Recognition of moving video images is a very complex operation. It seems that the presence of a fairly complete dictionary in which gestures are presented by three-dimensional models obtained by motion capture, may to some extent alleviate this problem. Recognition can be reduced to the determination of the trajectories of hands and head [7], as well as the body and individual fingers. Recognition results should be used to locate the key nodes and compare its with nodes of models from the dictionary to find the closest gesture. It may be some variant of the methods used in OCR.

For unresolved (or incorrectly recognized) plots the video must to be processed by the researcher. It is timesaving to do the tokenization and tagging of images (for subsequent retrieval of individual gestures) in one procedure. The most widespread in the processing of video gesture images there are a means of annotating video and audio files ELAN (http://www.lat-mpi.eu/tools/elan/), developed by the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands [8].

However, in our work we found it more convenient to tokenization and tagging sign language videos to use auxiliary screen forms of our dictionary RuSLED [9], which show the contents of individual tables. Partitioning was carried out to facilitate the search of individual gestures, using as search terms to sign handshapes and the places of articulation.

The authors of the Mexican sign language dictionary allow two more parameters, namely movement and orientation of the palm [4]. We have restricted ourselves at this stage to two parameters because our current main aim is to further develop the prototype software for a future corpus. Adding additional tags to search for signs is rather straightforward and does not involve changes in the underlying tables of descriptions of sign, because the labels are contained in separate tables and refer to the entries in the main tables without affecting its contents. Indexing does not depend on how the images of gestures are presented - as two-dimensional video images or as three-dimensional animations.

While the authors of the Mexican sign language dictionary had the aim of finding the exact gesture, we set another goal: to give the learner the opportunity to find a sign, which he/she is interested, by approximate description of the sign, including an indication of the positions and configurations of hands regardless of the order of execution, and compliance of coexistence of pair "position - configuration". This corresponds to the case when the user does not know exactly how the gesture is performed. There are opportunities to search the sign, using all parameters of a learner query or only some of them. For retrieved signs the learner has the opportunity not only to see all the features of their performance, but also to get acquainted with comments explaining the semantic meaning of gestures as well as other explanations that facilitate the learning of sign language.

This method of fuzzy search can be used not only for training purposes but also for studies of sign language, for example, grouping and classification of signs, performed in a particular place or space, having a certain configuration of hands and so on. Also this is providing an opportunity to retrieve from a corpus those signs having a specific grammatical meaning.

The proposed search method is based on set-theoretic representation, in which separate indices are considered independently.

The sets L and K are finite sets of possible positions (L) and configurations (K) of hands. The finite set G represents gestures collected in the dictionary of corpus.

Set of positions of hands in gestures is mapped as $f_1: L \rightarrow G$ and the set of configurations of hands in gestures is mapped as $f_2: K \rightarrow G$, maps are given in tabular form.

Accordingly, the positions and configurations of hand gestures are mapped in $G_1 = f_1(\{x: x \in L\}) \subseteq G$ and $G_2 = f_2(\{x: x \in K\}) \subseteq G$. The set of gestures that meet the specified values of the positions and configurations is the intersection of G_1 and G_2: $G_f = G_1 \cap G_2$.

Learner has the opportunity to search the gestures which correspond to all of these parameters or only some of them.

In the first case (called the search type «И» = “AND”)
$G_f = G_1 \cap G_2$, where
$G_1 = f_1(\{x: x_1 \wedge x_2 \wedge \ldots \wedge x_l \in L_u\});$
$G_2 = f_2(\{x: x_1 \wedge x_2 \wedge \ldots \wedge x_k \in K_u\}).$

Here L_u and K_u denote the user-defined set of positions and configurations, l and k denote the cardinality of these sets – the number of elements contained in them. If l or k exceeds the actual number of positions and configurations for any gesture (ie, the gesture does not contain all user-specified position or configuration), this gesture will not be included in the result set G_f.

In the second case (called the search type «ИЛИ» = “OR”) also $G_f = G_1 \cap G_2$, but
$G_1 = f_1(\{x: x_1 \vee x_2 \vee \ldots \vee x_l \in L_u\});$
$G_2 = f_2(\{x: x_1 \vee x_2 \vee \ldots \vee x_k \in K_u\}).$

In this case, the search results will display all the gestures that contain at least one position and configuration selected by the user.

Obviously, if the user does not have a good knowledge of sign language, it will be difficult to him/her to give an exact list of hand positions and configurations inherent to the sought-for gesture. Specification of the parameter values may lead to the situation that the sign will not be retrieved from a dictionary. Where there is doubt about the sign parameters, it is better to use the second way of search, although the results give a fairly comprehensive list of signs that are only partially satisfying the given search criteria. The final choice will be made by the learner by browsing through all signs retrieved on his request.

If the value of the parameter is omitted, the operation of union of the sets $G_f = G_1 \cup G_2$ is used, ie when $G_1 = \varnothing$, $G_f = G_2$; at $G_2 = \varnothing$, $G_f = G_1$. In this case, you can see, for example, all the gestures that contain the specified configuration of hands (regardless of their place of execution) or performed in specified places (regardless of the configuration of hands). In contrast to [4], compound gestures are indexed and displayed in search results also. A general view of the user interface and the results of both types of searches are shown in Fig. 1 and 2.

At the top of the panel are two lists of valid values of hand positions and configurations which are can be selected by user. Fig. 1 shows the result set when the user clicks "AND"; in Fig. 2 - Button "OR". It is seen that both search variants are using the same set of search criteria (position: in front of the chest and in front of the forehead; configuration: the index finger straightened, the rest are pressed to the palm) but in the second case more signs were retrieved. The user sees the number of retrieved signs and the number of words corresponding to these signs in a separate window. Fig. 2 shows that these numbers may be not equal. There is no one-to-one correspondence between words and signs. In our dictionary RuSLED we tie words and signs using semantic meaning, separating polysemantic meanings.

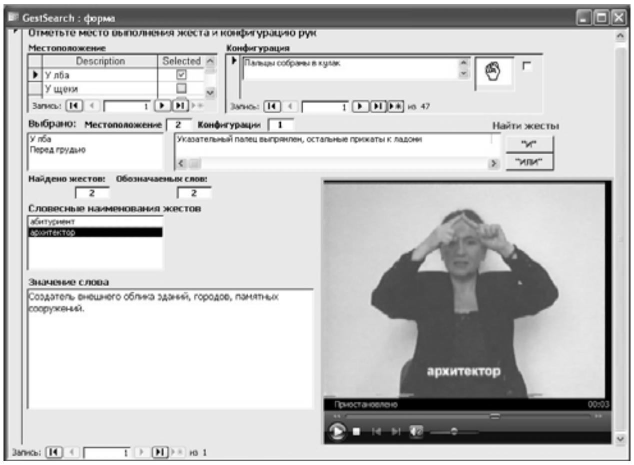

Figure 1. The result of signs retrieve for the command "И" = "AND" (a rigorous search)

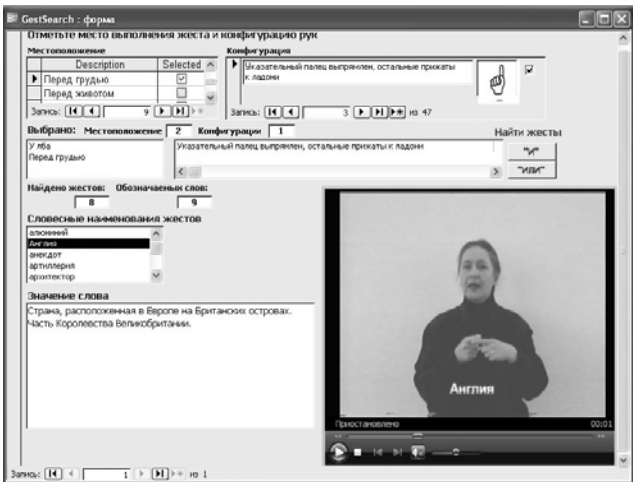

Figure 2. The result of signs retrieve for the command "ИЛИ" = "OR" (a fuzzy search)

In the mark-up process, the corpus annotator also uses the auxiliary lists of valid parameter values similar to those shown in Fig. 1 and 2. They help to choose the right parameter codes, because the number of options is large (for the configurations it is greater than 50). To select the desired value, the annotator uses a mouse, although it is recommended to use the keyboard in order to speed up the work [5].

We hope that these decisions will help create an efficient corpus of Russian sign language and use it for the study of sign language and the development of training manuals and specialized dictionaries.

7. ACKNOWLEDGMENT

The author thanks the reviewers who spent so much energy on parsing virtually unreadable first version of the paper and apologized to them. Special thanks to Thomas Hanke, without whom this paper would not have been corrected.

8. REFERENCES

[1] Braffort A., Bolot L., Chételat-Pelé E., Choisier A., Delorme M., Filhol M., Segouat J., Verrecchia C., Badin F., Devos N. (2010) Sign Language corpora for analysis, processing and evaluation. // *Proc. of the 7th conference on International Language Resources and Evaluation (LREC'10)*. — ELDA, 2010..

[2] *Convention on the Rights of Persons with Disabilities* A/Res/61/106 – http://www.un.org/disabilities/default.asp?id=199

[3] Goldin-Meadows, S. & Mayberry, R. I. (2001). How do profoundly deaf children learn to read? Learning Disabilities Research and Practice (Special issue : Emergent and early literacy : Current status and research directions), 16, 221-228.

[4] Hawayek A., Del Gratta R., Cappelli G. (2010) A bilingual dictionary Mexican Sign Language –Spanish/Spanish – Mexican Sign Language // *Proc. of the 7th conference on International Language Resources and Evaluation (LREC'10)*. — ELDA, 2010.

[5] Hovy E., Lavid J. (2010) Towards a 'Science' of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics // *International Journal of Translation*, Vol. 22, No. 1, Jan-Jun 2010.

[6] Hruz M., Campr P., Karpov A., Santemiz P., Aran O. and Zelezny M. (2009) Input and Output Modalities Used in a Sign-Language-Enabled Information Kiosk // *Proc. of the 13th International Conference "Speech and Computer" SPECOM'2009*. — St. Petersburg: SUAI, 2009.

[7] Lefebvre-Albaret F., Dalle P. (2010) Video retrieval in sign language videos: how to model and compare signs? // *Proc. of the 7th conference on International Language Resources and Evaluation (LREC'10)*. — ELDA, 2010.

[8] Sloetjes, H., Wittenburg, P. (2008) Annotation by category – ELAN and ISO DCR. // *Proc. of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. — ELDA, 2008.

[9] Voskresenskiy A.L., Gulenko I.E., Khakhalin G.K. (2009) From Sounding Speech to Sign Language // *Proc. of the 13th International Conference "Speech and Computer" SPECOM'2009*. — St. Petersburg: SUAI, 2009.

[10] Zaliznyak A.A. (1977) *Grammatical dictionary of Russian language. Inflection*. In Russian: Grammaticheskiy slovar russkogo yazyika. Slovoizmenenie. — M., 1977.

[11] Zaytseva G.L. (1991) *Dactylology. Sign Language*. In Russian: Daktilologiya. Zhestovaya rech. — M.: Prosveschenie, 1991