About Recognition of Sign Language Gestures

Alexander Voskresenskiy, Sergey Ilyin
School for deaf children (1), Academy of Fantasy (2)
Moscow, Russian Federation
E-mail: avosj@yandex.ru, mail@mocaprus.ru

Abstract

A motion capture technique for implementing sign language dictionary is described. Problems of perception and recognition of gestures of Russian sign language in system of the automated sign language translation are discussed. The new approach to morphology of gestures and a method for separate gestures in sign statements are offered. The working definition for "text understanding" is offered.

1. Introduction

The main goal of our work is to create Russia's first explanatory dictionary of Russian sign language, using three-dimensional animated characters created by motion capture techniques and training manuals that contain sign statements-assembled based on this dictionary.

The purpose of the vocabulary and tools is to help deaf people to learn the Russian verbal language, and promoting people who are learning sign language.

A fixed number of examples of verbal and sign statements contained in the manuals are not always able to meet the needs of the student. Therefore, in subsequent stages of work is supposed to create an automated system of sign language interpretation.

Currently there are no word processors, who understand text contents. Available word processors are based on statistical methods. This leads to a significant number of errors, reduction of which using the existing methods is hardly possible.

Our approach to the problem of understanding based on the fact that both verbal and sign language used to describe the same surrounding world. Therefore, we believe that the basic concepts describing the surrounding world for hearing and deaf people are the same.

A comparison of the meanings of words and gestures enabled us to formulate a working definition of the term "understanding of the text[1]".

Basic complexities at translation of text into signs are connected with homonymy resolution, searching of necessary meaning of polysemic word and/or sign, and also with transformation of phrases of Russian language into Sign Language expressions.

Procedure for transfer of sign statements in the text even more complicated, because the gesture utterance does not contain information about the grammatical forms of words from which to generate text, such as noun and verb in many cases are indicated by the same gesture.

The focus of this work is given to the separation of sign utterances into constituent gestures.

[1] The term "text" is used to denote the means of exchange of knowledge between people, including both written text and speech communication (verbal and signed).

2. Current Results

2.1 Short Description of RuSLED Dictionary

Russian Sign Language Explanatory Dictionary RuSLED includes functions of explanatory dictionary as for entered word, so and for gesture representation. On input of dictionary any form of word can be entered, and at the output variants of gesture interpretation of given lexeme are shown.

Dictionary contains 2372 words (with interpretations of their meanings) and 2537 video images of gestures (including variants of the sign) which represent meanings of the words. For 1592 gestures (63% from total number in dictionary) additional explanatory, concerning to manner of execution of gesture or describing semantic nuances are given.

Gestures used in Saint Petersburg and its vicinities are presented in the dictionary. They in part coincide with Moscow gestures but divergence is big enough, what gave occasion to name given dictionary "Petersburg's dialect".

In first version of dictionary digitized fragments of video recording borrowed from video course (IRRC, 2002) are used. Use for viewing of gestures of Windows Media Player ActiveX element allows: to see this gesture repeatedly, at pressing of button ► of player; to suspend performance of gesture in required place, at pressing of button || of player; to see any phase of gesture, moving cursor of player in appropriate position by mouse (fig. 1).

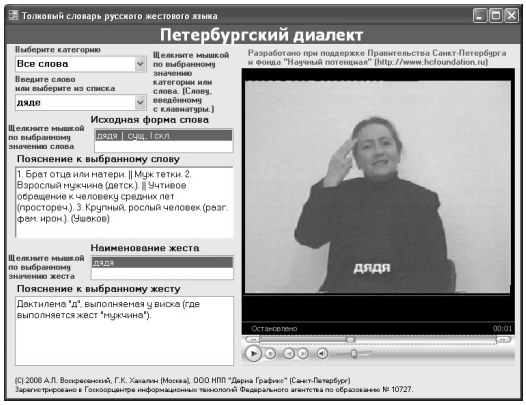

Figure 1: RuSLED dictionary display

Video recording was used for best representation of mimicry accompanying gestures and executing essential role in sign language of deaf persons. So, for example, words «милый», «симпатичный» (darling, nice) are passed with one gesture, but they are differing by movements of lips pronouncing fragments of corresponding words. In new dictionary version video records will be substituted by avatars using motion capture methods.

For some gestures explanatory from (Fradkina, 2001) were used. This dictionary is made on basis of Moscow variant of Russian Sign Language.

For compiling of words explanatory more 30 dictionaries and encyclopedias were used.

On deaf children teacher's recommendations opportunity is provided to filtration of word list of dictionary on grammatical categories (nouns, verbs, adjectives, adverb, pretexts, particles, numerals, pronouns). For viewing all dictionary content it is necessary to choose category "All words".

Separate input of dictionary (separate recording in table of database) is used for everyone semantic value of lexeme (and gesture). This dictionary feature is very convenient for user, and is recommended by lexicographers.

Field «Введите слово» ("Enter Word") allows to enter any word forms or choose lexemes from associated list. In list «Исходная форма слова» ("Initial Word Form") a lexeme corresponding to stem in field "Enter Word" is outputted or several lexemes are outputted if several records are chosen by results of morphological analysis.

When user chooses a lexeme from list "Initial Word Form" as result name of corresponding gesture is outputted in list «Наименование жеста» ("Name of Gesture"). If several gestures correspond to given lexeme then list of names of gestures is outputted. For each word meanings only that gesture is outputted, semantics of which corresponds to meaning of chosen lexeme (Voskresenskij & Khakhalin, 2007).

2.2 Our Approach to Understanding of Text

Word processing is usually divided into successive stages of morphological, syntactic, and, as a final stage, the semantic analysis. However, in some cases, morphological analysis can be performed only on the basis of syntax; in turn unambiguous parsing proposal assumes knowledge of grammatical forms of words in the sentence. Therefore consistent scheme of sentence parsing should be replaced by a scheme of interaction of agents performing different tasks and share the results to refine their work (Majumdar et al., 2008).

Modern systems for semantic text processing for removal of polysemy use ontology and thesauri. As the evaluation of the quality of such systems, the number of errors even in the best samples does not fall below 30% (Loukachevitch, Chuiko, 2007). The main reasons for this are incomplete vocabulary and inadequate procedures for resolving polysemy.

But what is the understanding of the text? The following definition was developed on the basis of comparison and analysis of interpretations of the meanings of words and gestures:

The result of understanding of the text should be the selection and identification of objects described in the text, their spatial positions, as well as registration of changes to their characteristics, actions and conditions in accordance with the change of the text time.

According to the results for each given moment of the passed time of the text we can construct a picture, describing the locations and interactions of the objects are described in the text — the situation. In addition, the interpolation of changes of objects characteristics can provide short-term forecasting of changes of situations.

Supporting examples can be found in the RuSLED dictionary. Some of them are described in (Voskresenskiy et al., 2009).

System of the text understanding should not only store information about semantic relationships of words (often ambiguous), defined by thesauri and ontology, but also must be able to speculate on the possible actions of the subject and the objects described in the text.

Identification of objects includes not only the allocation of group names that describe a particular object, but also recognition that, if the object met earlier in the text; if the objects are the same whether they have the same names (Kazi, Ravin, 2000). For this system, described in (Voskresenskij, 2008), includes not only the basic ontology, storing descriptions of classes and their relations, but also the ontology of the text, including descriptions of specific instances of classes. This ontology will inherit from the basic ontology characteristics of the classes and their relationships, adding to them the characteristics of specific instance (including its position in space).

For example, if the text describes the room in which there are several tables, then to understand what is at some of the tables, not enough to know a general description of the semantic class "table", each instance must be identified. But some of the hallmarks of an instance of a class can be meaningful only within a particular text, so they should not be included in the basic ontology. If they are repeated in different texts and for different instances, these features are important not only within a particular text. Then in the process of system self-learning they must be included in the basic ontology, leading to partition the source class into subclasses.

From this it follows that the ontology of a particular text should not be destroyed upon completion of the text processing, but should be kept for some time. It is necessary to compare information from different texts and identify the most plausible, which may be included in the basic ontology of the system.

The proposed approach to the understanding of the text is useful not only for sign language interpretation, but also for machine translation systems for verbal language. For example, in the Ingush language to convey information about the event, which ended recently, and in which the telling the story subject was present or absent, different forms of the verb are used.

3. Tools for mapping, perception and recognition of gestures

Various versions of the notation, for example Hamburg notation system HamNoSys[2], used to record gestures. In our country, notation proposed by L.S. Dimskis (2002) is used.

Dictionary of Russian sign language RuSLED is added by the function of gesture search using his approximate description. The challenge is that we need to find a gesture that people saw, but does not know its meaning. It uses a simplified notation, hided inside the dictionary, user-accessible lists of possible values: text to describe the place of performance gesture, text with a pattern — for the configurations of fingers. Based on user-selected values search query is formed and returns a set of gestures to meet this request, from which the user selects the gesture.

Demonstration of gestures in the new version of the dictionary made by animated character - an avatar, to record of gestures method of motion capture is used. Record is performed by "The Academy of Fantasy» (www.mocaprus.ru). Movements of demonstrator recorded using 12 cameras and a host of reflectors on the suit (Fig. 2), are converted to 3D-model (Fig. 3), and used to form the shape of an avatar that can be placed into any stage.

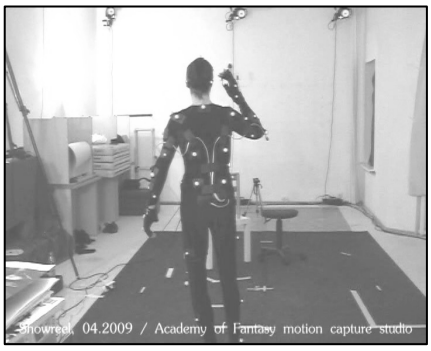

Figure 2: Demonstrator in a suit with reflectors

Figure 3: 3D-model

Movements of the fingers of the demonstrator are recorded using special gloves. To record the facial expressions and articulation the apparent on the face of the reflectors is used (Fig. 4). Their signals are converted into three-dimensional model of facial mimicry (Fig. 5).

[2] http://www.sign-lang.uni-hamburg.de/projects/hamnosys.html.

Figure 4: Demonstrator with glued on the face reflectors

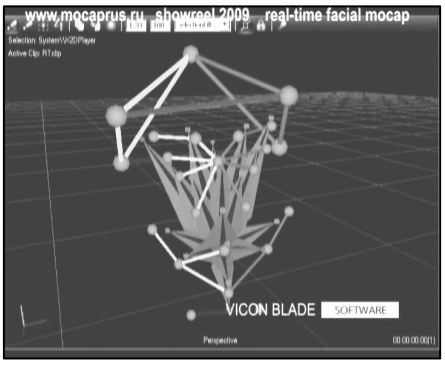

Figure 5: Three-dimensional model of facial mimicry

This will significantly expedite the filling of the dictionary through the use of several sign language interpreters for the demonstration of gestures, while preserving the unity of action expressed by the appearance of a single virtual gesture demonstrator. Formed in such a way dictionary will allow composition of sign statements of gesture collections stored in the dictionary maintaining, as noted above, the unity of action which is important for perception of sign utterances by human.

Studio recording of gestures, allowing you to create original dictionary, obviously, can not be a means of communication with deaf people.

For recognition of gestures there is proposed to develop means of converting raster images of sign language interpreter taken with a camera in a vector images. This transformation includes the recognition of the essential for this task image detail: the head, hands (and the position of each finger), the torso. These details of the image are converted to ellipses and rectangles, the coordinates of which are compared with the parameters of the skeleton of a virtual demonstrator (avatar) of the dictionary.

Transformation perceived image in vector form allows you to significantly reduce the memory requirements of the intellectual system and accelerate the procedure of comparison with etalons.

Methods for converting the image to be used are similar to those used in the pre-processing of images in the systems of character recognition.

Information on the exact position of avatar in space, such as hands, which was absent in two-dimensional scanning

images from a camcorder, it is planned to receive from the knowledge of possible and permissible mutual positions of various parts of the body. To determine the exact pose of avatar the appropriate geometric constructions will be applied, providing the closest match with the original raster image the projections of the avatar on the plane.

4. Methods of processing of sign phrases

In the analysis of sign utterances should take into account that many of the signs are composite, contain a combination of several gestures and pre dactyl signs, modifying the meaning of this sign. When you need to specify, for example, case endings, after a gesture relevant dactyl signs are signed.

Gestural speech does not contain pauses between individual gestures. Only phrases are separated by pauses. This introduces additional complexity in the implementation of automated sign language translation, like those encountered in the development of continuous speech recognition systems.

Given the integral nature of the gestures, the separation of gestural phrases into separate gestures should be maintained by selecting from the vocabulary appropriate gestures, having the greatest length, and analyzing the semantics of the resulting expressions. If its meaning does not match the discourse, we can proceed to successively splitting "long" sign on the constituent elements, trying to get a statement, the content of which corresponds to the discourse. Considering also that the gesture might pass the words of different grammatical forms, construction of syntax tree of a text sentence offers a complex combinatorial problem whose solution is a simple brute force attack is impossible, since it leads to the "exponential explosion".

The solution is to use the method of sequential analysis and retention options without the incremental construction of solutions (Mikhalevich, Volokovich, 1982), which reduces the number of options under consideration. This criterion for excluding unpromising options is contradictory semantics of the resulting text.

5. Conclusion

In the case of sufficiently reliable recognition of gestures using a camcorder (preferably a qualitative recognition using standard web cameras) and establishing a system of sign language interpretation will be possible to ensure prompt communication of the deaf with administration officials and the public, that is a function of "electronic government".

Many details of the process of understanding and explanation, expressed in words, hidden from direct observation in the subconscious, which hampers the development of word processors understanding the text. Based on a comparison of different models of thinking, presented verbally and in sign language, developed a model for understanding the text. Accordingly, there is an idea of the architecture of a system that could perform the required functions.

6. Acknowledgements

The authors express sincere gratitude to N.A. Chaushian. Her very useful comments on sign language allowed drawing attention to the features not always visible to researcher who is not a native signer, and correcting some errors in dictionary RuSLED.

7. References

Dimskis, L.S. (2002). *Start Learning Sign Language*. Moscow: Academy. (In Russian: Димскис Л.С. Изучаем жестовый язык. М.: Издательский центр «Академия», 2002. — 128 с).

Fradkina, R.N. (2001). *Hands Which Are Talking*. Moscow: All-Russian Deaf Association. (In Russian: Фрадкина Р.Н. Говорящие руки: Тематический словарь жестового языка глухих России. М.: Изд-во ВОИ. — 598 С).

IRRC, (2002). Specific communication means of deaf : videocourse in 3 parts. Saint-Petersburg – Pavlovsk : Inter-regional Rehabilitation Center. (In Russian : Специфические средства общения глухих: Видеокурс: В 3 частях. СПб – Павловск: МЦР, 2002).

Kazi, Z., and Y. Ravin. (2000) Who's Who? Identifying Concepts and Entities across Multiple Documents. *Proceedings of the 33rd Hawaii International Conference on System Sciences*. (0-7695-0493-0/00).

Loukachevitch, N., and D. Chuiko (2007). Thesaurus-based Word Sense Disambiguation (In Russian: Лукашевич Н.В., Чуйко Д.С. Автоматическое разрешение лексической многозначности на базе тезаурусных знаний. *Интернет-математика-2007*. Екатеринбург: Изд-во Урал. ун-та. — С. 108 – 117).

Majumdar, A., J. Sowa, and J. Stewart. (2008) Pursuing the Goal of Language Understanding. *Proceedings of the 16th ICCS* /P. Eklund and O. Hammerlé, eds. LNAI 5113, Berlin: Springer, pp. 21-42.

Mikhalevich V.S., Volkovich V.L. (1982) *Computational methods of research and design of complex systems*. Moscow: Nauka. (In Russian: Михалевич В.С., Волкович В.Л. Вычислительные методы исследования и проектирования сложных систем. М.: Наука, 1982. — 288 с).

Voskresenskij, A. (2008). Text Disambiguation by Educable AI System. *The First Conference on Artificial General Intelligence* / P. Wang et al. (Eds.). Amsterdam: IOS Press. — P. 350 – 361.

Voskresenskij, A., and Khakhalin, G. (2007*)*. Semantic Search Engine in a Multimedia Russian Sign Language Dictionary. *Proceedings of the XIIth International Conference "Speech and Computer" SPECOM'2007*, Moscow, Russia, pp. 739 – 744.

Voskresenskiy A.L., Gulenko I.E., and Khakhalin G.K. (2009) From Sounding Speech to Sign Language. *Proceedings of the 13-th International Conference "Speech and Computer" SPECOM'2009*. St. Petersburg: SUAI, 2009. P. 539 – 542.