Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 6079–6085
Marseille, 11–16 May 2020

TheRuSLan: Database of Russian Sign Language

Ildar Kagirov, Denis Ivanko, Dmitry Ryumin, Alexander Axyonov, Alexey Karpov
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences
St. Petersburg, Russia
{kagirov, karpov}@iias.spb.su, denis.ivanko11@gmail.com, {a.aksenov95, dl_03.03.1991}@mail.ru

Abstract

In this paper, a new Russian sign language multimedia database TheRuSLan is presented. The database includes lexical units (single words and phrases) from Russian sign language within one subject area, namely, "food products at the supermarket", and was collected using MS Kinect 2.0 device including both FullHD video and the depth map modes, which provides new opportunities for the lexicographical description of the Russian sign language vocabulary and enhances research in the field of automatic gesture recognition. Russian sign language has an official status in Russia, and over 120,000 deaf people in Russia and its neighboring countries use it as their first language. Russian sign language has no writing system, is poorly described and belongs to the low-resource languages.
The authors formulate the basic principles of annotation of sign words, based on the collected data, and reveal the content of the collected database. In the future, the database will be expanded and comprise more lexical units. The database is explicitly made for the task of creating an automatic system for Russian sign language recognition.

Keywords: Russian sign language, low resourced languages, corpora annotation, image recognition, machine learning

1. Introduction

This paper presents the electronic lexical database of Russian sign language (RSL) collected for one subject area. The database is called TheRuSLan (Thesaurus Russian Sign Language (TheRuSLan, 2019)) and is the first of a kind for Russian sign language. As the authors believe, the collected database can be helpful in tasks of machine learning, gesture and sign language automatic recognition, and sign language linguistics.
The primary objective of the investigation was to create a database of RSL, which could be of interest both for linguistic research and machine learning. This was done with the use of two annotation layers: linguistic ("phonological") and "image recognition-oriented", i.e. class labeling.
The paper is structured as follows: after the Introduction, Section 2 provides some information about RSL as low-resourced natural language. Section 3 sketches the state-of-the-art in the field of corpora and databases of RSL. In Section 4, the authors present their view on building RSL corpora, and provide the basic distinctive features used to annotate the collected database. Following this, Sections 5 and 6 contain information concerning collecting data and general features of the database. Eventually, the last section 7, serves as round-up of the paper where we discuss all the obtained results as well as the main conclusions.

2. Russian Sign Language as a Low-Resourced Language

RSL is the language of communication used by the deaf community in Russia and some of its neighboring countries. As the statistics by the Ethnologue international catalog (www.ethnologue.com) indicates, the number of RSL speakers was over 122 thousand people in 2010. In 2012, RSL was officially recognized as the language of communication in the Russian Federation.
Gestures as a form of communication are of great importance in everyday life and constitute different language systems and sub-systems. Most of them share the property of being independent communication system based on gesticulation. According to the data provided by the World Health Organization, today around 466 million people worldwide have disabling hearing loss, and this number is going to grow during the next 30 years up to 900 million. There are no data on how much of those people use sign languages in everyday life, but their number must be considerable.
Sign language is, like any natural language, a structured form of communication involving gestures and motions. SLs make use of different parts of body to convey meanings. Unlike spoken languages, SLs benefit from space to build up strings of gestures and express semantics. That's why gestures of SL are classified, inter alia, according to position and orientation. Among other distinctive features are handshape and trajectory.
In spite of the official status and active use by the Russian deaf community, RSL was not sufficiently described and presented to the scientific community. Linguistic research on RSL is in its infancy yet. Respectively, RSL can be classified into low-resourced languages due to the fact that there are not much RSL data collected. Databases, however, are essential to training, testing and comparison of automatic (sign) language recognition systems.

3. SL Databases: State-of-the-Art

Nowadays there are many hand gesture databases, collected by different teams for different purposes.
In this paper, the most widely used databases for sign language and hand gesture recognition are given in the Table 1. In contrast to the linguistic resources, these corpora were explicitly created for SL recognition, natural language processing and computer vision tasks. Typically, there is a lack of suitable video corpora for the task of sign language recognition and this kind of data differs significantly from the real language encountered outside the research lab. In the presented table, two of the largest publicly available SL video corpora are listed (7 and 8). The others presented corpora are meant for gesture recognition tasks, which is closely related to SL recognition, and usually used to train parts of sign language recognition systems. Description of the datasets and details such as number of classes, subjects, and samples are given in the table. More gesture recognition databases relevant to SL recognition research can be found in (Pisharady and

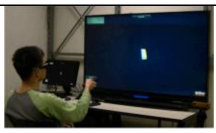
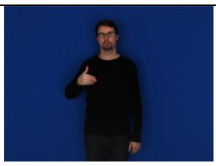
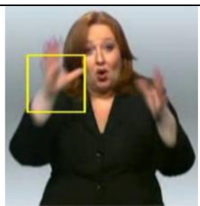
#	Dataset	Year	Task	Sample images	# sub.	# classes	duration	#s per sub.	#samples	Sensor
1	ChaLearn	2011	1) Play game 2) Remote control 3) Learning SL		20	many	-	different	50000	Kinect
					video height	video width	video fps	Data	Free acc.	Link
					320	240	10	RGB-D	Yes	(Malgireddy et al., 2012)
2	MSRC-12	2012	gestures to be recognized by the MS system		#sub.	# classes	duration	#s per sub.	#samples	Sensor
					30	12	6h. 40m.	≈200	6244	Kinect
					video height	video width	video fps	Data	Free acc.	Ref.
					640	480	-	RGB-D	Yes	(Simon et al., 2012)
3	ChaLearn Multi-modal	2013	Multi-modal Italian gestures recognition		#sub.	# classes	duration	#s per sub.	#samples	Sensor
					27	20	-	≈500	13858	Kinect
					video height	video width	video fps	Data	Free acc.	Ref.
					640	480	30	RGB-D + audio	Yes	(Escalera et al., 2013)
4	NUS dataset-II	2012	Hand posture recognition		#sub.	# classes	duration	#s per sub.	#samples	Sensor
					40	10	-	-	2750	Camera
					image height	image width	video fps	Data	Free acc.	Ref.
					320 (160)	240 (120)	N/A	RGB Images	Yes	(Pisharady and Saerbeck, 2013)
5	6D Motion gesture dataset	2011	Gesture recognition		#sub.	# classes	duration	#s per sub.	#samples	Sensor
					28	20	-	200	5600	Camera
					video height	video width	video fps	Data	Free acc.	Ref.
					-	-	-	Row binary	No	(Chen et al., 2012)
6	Sheffield kinect gesture dataset	2013	Learning Discriminative Representations		#sub.	# classes	duration	#s per sub.	#samples	Sensor
					6	10	-	360	2160	Kinect
					video height	video width	video fps	Data	Free acc.	Ref.
					640	480	30	RGB-D	Yes	(Liu and Shao, 2013)
7	SIGNUM database	2008	Sign Language recognition		#sub.	# classes	duration	#s per sub.	#samples	Sensor
					50	455	42.7 h	-	19500	Camera
					video height	video width	video fps	Data	Free acc.	Ref.
					780	580	30	RGB	Yes	(Agris et al., 2008)
8	RWTH-PHOENIX-Weather	2014	Sign Language recognition		#sub.	# classes	duration	#s per sub.	#samples	Sensor
					27	1081	12.54h	-	6841	Camera
					video height	video width	video fps	Data	Free acc.	Ref.
					210	260	25	RGB	No	(Forster et al., 2014)
9	Chalearn Isolated gestures recognition	2016	Isolated gesture recognition		#sub.	# classes	duration	#s per sub.	#samples	Sensor
					21	249	-	-	47933	Kinect
					video height	video width	video fps	Data	Free acc.	Ref.
					640	480	30	RGB-D	No	(Li et al., 2016)

Table 1: Available sign language and hand gesture databases

Saerbeck, 2015).
However, only few databases are available for RSL, and they were created either for educational or linguistic tasks. The only annotated linguistic database is the Russian Sign Language Corpus (RSLC, 2014) by Novosibirsk State Technical University (NSTU). NSTU corpus contains over 230 samples by 43 signers, and was annotated with the ELAN tools, being equipped with metadata search filters. In general, this corpus is an effective linguistic tool for researching RSL. It should be noted, however, that most of the signers use local dialects (primarily Moscow and Novosibirsk), thus the data provided by this corpus are – in some aspects – not of high relevance for other RSL idioms.
In addition to the NSTU corps, there are other databases of RSL. Most of them are video tutorials showing single words, or phrases. The most significant are “Explanatory Dictionary of Russian Sign Language – RuSLED” (2002) and Surdoserver 2.0 project. For a detailed review of RSL databases of this kind see (Kagirov et al., 2020). All of them have basic search functions implemented, but no annotation. Any use of these databases (as well as the NSTU corpus) for automatic gesture recognition requires preliminary work on data processing and annotation.
Moreover, as revealed in (Kharlamenkov, 2017), most of the RSL available databases are of poor quality: some of them are just a mixture of different signs that belong to different speech styles and dialects.
Based on the aforementioned analysis, it was decided to record own database for the task of Russian SL recognition. For this purpose, a Kinect 2.0 based software-hardware complex with a video camera, infrared camera and depth sensor has been developed.

4. Creating a SL Database: Basic Principles

As a rule, any research in the field of machine learning begins either with using of any of the existing datasets, or with collecting a new one. In the latter case, the researchers collect the database based on their own tasks and goals. Quite popular are databases containing isolated words, numbers and letters.
Since 2018, at SPIIRAS an interdisciplinary research to create a multimodal human-machine interaction (HMI) interface that supports recognition of both spoken Russian and RSL has been conducted. We assumed that robotic trolley for supermarkets and grocery stores can be equipped with a similar user interface and this reason has determined the subject area of the database, as well as the content, i.e. isolated words and phrases regarding supermarket products. The reason for choosing this subject area, i.e. supermarket items and interactions, is the practical feasibility of the robotic trolley project. Supermarkets are a natural choice, because are regularly attended by the deaf and hard-of-hearing. The items and locations in supermarkets are regularly changing, thus navigation trough shelves and departments can be a tricky task.
As stated above, the authors aimed to create such a database annotation that would be based on features that could be used not only for linguistic notation, but also for computer analysis and automatic recognition. A feature of the proposed annotation system is that the parameters laid down in the basis of the actual phonological interpretation of the gesture find implementation in those classes that are allocated for constructing probabilistic models.
In (Stokoe, 1960), an innovative approach was introduced, implying decomposition of gestures into three components: 1) handshape and hand orientation; 2) location; 3) trajectory. The set of combination is not very large (Battison, 1978) distinguishes 45 handshapes, 25 localizations, and 12 types of performing a gesture in American SL). Handshapes are determined by "active" fingers and operations with them (fingers can be bent, hooked, flattened, etc.). The principles of gesture description formulated by Stokoe have proven to be good enough to describe isolated gestures. One of the Stokoe principles-based notation systems is the Hamburg Notation System, or HamNoSys (Prillwitz et al., 1989; Hanke, 2004). HamNoSys is well suited for use in electronic databases due to its linear structure, precision and ease of conversion to Unicode format. There have been some research on the integration of HamNoSys into the general system of annotation of sign language corpora (Hanke and Storz, 2008); this notation system was also used to annotate the database of Australian sign language (Johnston, 2001), to create dictionaries of the New Zealand sign language (Kennedy et al., 1997) and German Sign Language (Arbeitsgruppe Fachgebärdenlexika, 1996). This is the reason why the authors of the proposed paper have chosen HamNoSys for TheRuSLan database. For reviews of notation systems for sign languages, see, for example (Karpov and Kagirov, 2011: 130 ff; Frischberg et al., 2012: 1045-1054; Garcia and Sallandre, 2013). To describe the configuration of the hand, 7 main forms of the hand were identified (Table 2):

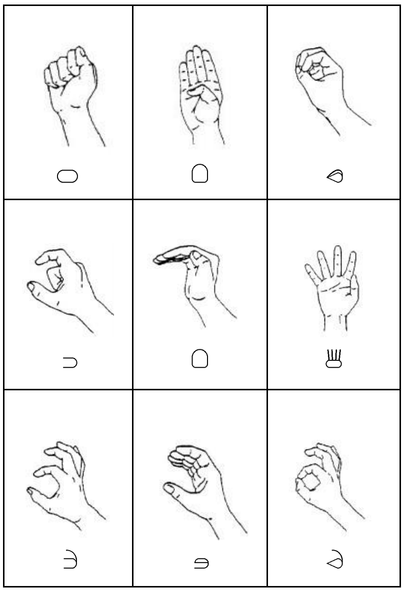


Table 2: Basic handshapes in the database

The rest of forms can be described as their modifications. The table below shows all the modifications in TheRuSLan for one-handed gestures as the result of operations with the selected fingers, see Table 3:

active finger	operation / contact									
none		+	+	+	+	+	+	+	+	+
1	none	+								
	hidden		+							
	hooked	+								
	flattened				+					
	3		+							
	5		+							
	2,3	+								
2	none	+								
	bent	+								
	hooked	+								
3	no		+							
2,3	no	+								
	spread	+								
	hooked	+								
2,5	no	+								

Table 3: Base handshapes modifications

The next parameters to consider are orientation and localization. The hand can be oriented along 18 spatial axes, in addition, the hand can be rotated in 8 ways in the signing space. The standard HamNoSys classification implies the localization of the hand articulators in 30 different ways, but in fact there is a significant variability in signing, which made it possible to enlarge the main localization areas, combining them into 11 ones (Table 4):

level	***explanation***
1	above eye level
2	eyes
3	nose
4	mouth and cheeks
5	chin
6	jaw
7	neck
8	head
9	shoulders
10	torso
11	belly

Table 4: Main locations used in the database

Since the standard gesture recognition implies processing of a sequence of video frames, the trajectory can be determined by comparing localization change frame by frame. Thus, each lexical unit in the database is annotated with the use of HamNoSys in the following way:

MILK

Obviously, such a description, being suitable for linguistic purposes, is of little use for machine learning. Therefore, another level of database annotation is key projections of the hands obtained by combining the data from Table 2 and Table 3, i.e. a hand in a specific configuration and with a specific orientation. A total of 44 projections were obtained that can be used for machine learning:

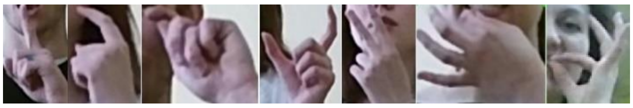


Figure 1: Examples of hand projections.

5. Database Description

At the moment, the database is a record of 164 lexical units and clauses performed by 13 informants with at least 5 iterations (Table 5). The total size is 3.8 TB in the original format, the total video duration is over 8 hours. Iterations were necessary for machine-learning purposes; at the same time, repeating signs is of high importance for analyzing gesture variability and to develop outlines of SLs annotation. Since almost all signers come from different regions of Russia, certain variability in signing was predictable.

№	Signer ID	Gender	Age	Duration (mins)	First language
1	spkr01	female	32	42:00	Russian
2	spkr02	female	40	34:28	Russian
3	spkr03	female	40	38:27	Russian
4	spkr04	female	21	31:33	RSL
5	spkr05	female	20	38:12	RSL
6	spkr06	male	20	30:47	RSL
7	spkr07	female	19	34:41	RSL
8	spkr08	female	20	37:03	RSL
9	spkr09	female	21	36:05	RSL
10	spkr10	male	22	38:43	RSL
11	spkr11	female	20	34:03	RSL
12	spkr12	female	25	37:35	RSL
13	spkr13	female	21	34:48	RSL

Table 5: Signer metadata

The subject area of the database is food products at the supermarket, and the lexical units can be divided into clusters "products", "departments", "moving in space", "looking for products." The vocabulary was mainly built up by exporting text files from the navigation menu of the websites of a range of local supermarkets. Another part of the vocabulary is the list of commands, which includes lexical units related to orientation (forward, backward, right, left), a number of verbs of movement (let's go, let's go) and their modifiers (fast, slower), location requests (where ?, show me) and specific goods (I need ... I want ...). The dictionary of departments includes names of departments in the store. Finally, the same list includes such words as “ticket office”, “exit”, “toilet”, etc. 85% of the records fell into the category “food products and other goods”, the remaining 15% of the records are represented by user commands, department names, questions and answers. Some words are composites, i.e. consist of two (or more) components capable of functioning as independent lexical units, for example: WINE = SWEET.ALCOHOL, HONEY = BEE.SWEET, ETC.

In total, there are 48 different one-handed gestures and 116 two-handed gestures in the database (Table 6).

Jelly	Garlic	Black caviar	Where ...?
Beer	Chicken	Faster!	Water without gas
Pickled cucumbers	Cashier’s office	Waffles	Paste goods
Frozen vegetables	Eggs	Bananas	Yeast
Yogurt	Beet	Red caviar	Where is it located ...?
Wine	Cutlets	Slower!	Sparkling water
Spice	Exit	Chupa Chups	Honey
Vegetable mix	Cake	Grapes	Flour
Kefir	Carrot	Caviar	Where can I buy ...?
Champagne	Sausages	Right!	Kvass
Salt	Toilet	Bagels	Groats
Frozen mushrooms	Pastry	Plums	Baking goods
Kissel	Eggplant	Fish patties	Where is ...?
Cognac	Sausage	Forward!	Lemonade
Sugar	Children’s goods	A loaf of white bread	Buckwheat
Frozen fruits	Gingerbread cookies	Mushrooms	Meat pockets
Milk	Zucchini	Mussels	Goodbye!
Alcohol	Wieners	Back!	Mineral water
Fresh meat	Detergents	Beigels	Rice
Frozen berries	Cookies	Herbs	Manti
Milk porridge	Pumpkin	Shrimps	Unfortunately, there is no way to get there now.
Tomatoes	Ham	Please help me!	Mineral water
Mutton	Toys	Crackers	Oatmeal
Frozen pastry	Cake	Nuts	Chebureks
Cream	Oranges	Live crawfish	Did you found what you were looking for?
Cabbage	Aspic	Show me …	Fruit drink
Pork	Books	Cracker rings	Peas
Chewing gum	Chocolate	Sauerkraut	Frozen cutlets
Butter	Tangerines	Sunflower oil	How else can I help you?
Cucumbers	Meat	Take me to ...	Juice
Beef	Wait!	Bread	Beans
Batteries	Candies	Tea	Pizza
Sour cream	Apples	Vinegar	Where ...?
Potatoes	Live fish	I want …	Water without gas
Hen	Follow me!	White bread	Paste goods
Lighter	Chocolate bar	Coffee	Yeast
Cheese	Pears	Mustard	Where is it located ...?
Onion	Fish	I need …	Sparkling water
Duck meat	Left!	Water	Honey
Cigarettes	Chocolate spread	Pasta	Flour
Quark	Lemons	Marinade	Where can I buy ...?

Table 6: Lexical units of the database

6. Acquiring the Dataset

The standard procedure of acquiring a dataset implies using sensors/video cameras placed in front of the signer. Also, different lighting and background conditions are used, as well as different dress and glasses. Some datasets are acquired with help of specially designed input devices, such as CyberGlove.

In our case, the recording was carried out using MS Kinect 2.0 device, positioned at a distance from 1 to 3 meters to the signer, from different angles and under different lighting and background condition. No additional input devices were used. Software developed by the authors was used to record a 3D video stream from the input, see an example of its interface in the Fig. 2:

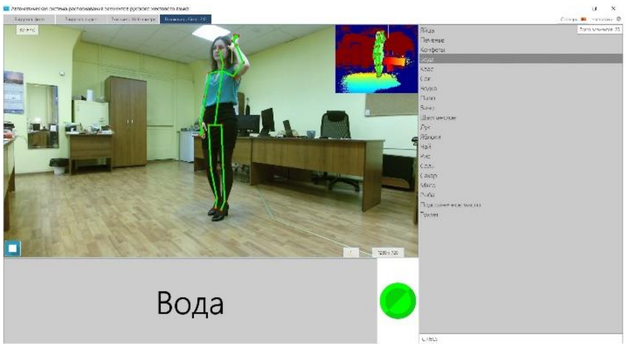


Figure 2: Recording interface

TheRuSLan database is recorded in 3D, which makes it a one-of-a-kind resource of RSL. In fact, even the most extensive databases contain signs as a 2D image. The 3D format was obtained due to the fact that the Kinect 2.0 device not only has ability to record FullHD video, but also the depth map mode. The use of the depth map mode introduces a third dimension, which makes it possible to determine the relative position of articulators with great accuracy. The distance between the active and passive hand, the distance from the body are tools of expressing various semantics in the sign languages of the world.

Moreover, the depth map can be useful in the tasks of automatic gesture recognition. In (Gruber et al., 2018), the authors propose an approach to American Sign Language numerals recognition, using the depth map data obtained via MS Kinect 2.0 sensor.

The authors of this paper would like to highlight, that the main intention was to create a DB of RSL, providing data on RSL in 3D, not to outdo (technically and lexically) the current DBs of the sign languages of the world.

The data obtained as a result of the recording have the following features:

1. .bin files containing a color (RGB) camera recording without compression (optical resolution of 1920x1080 pixels at 30 frames per second, color - 8 bits per pixel);

2. .bin video files containing a depth map data without compression (with an optical resolution of 512x424 pixels at 30 frames per second, color - 16 bits per pixel);

3. text files in XML format with coordinates of the skeletal model of the signer, divided into 25 joints; each joint is the intersection of two axes (x, y) on the coordinate plane and

an additional coordinate value with double precision, indicating the depth of the point, which is measured by the distance from the sensor to the point of the object.

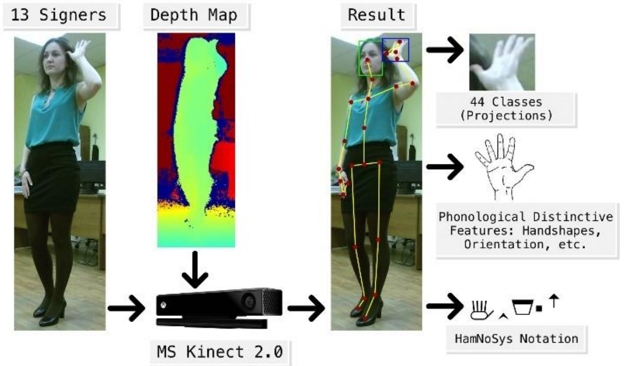


Figure 3: Process of acquiring data

The logical structure of the database can be represented in the form of a directory tree containing information for each gesture shown by a separate speaker: a) video recording of the gesture shown in FullHD format, in the depth map format; b) data on the position of the joints; c) images in jpg format, selected frame-by-frame from videos.

7. Conclusions

This paper presents a database of RSL within the subject area “products at the supermarket”. The value of this database is due to several factors. First of all, RSL is a low-resource language, and any high-quality database enriches our knowledge about this language. Moreover, the presented database was recorded in 3D, which provides significantly more opportunities for both gesture analysis and for using this database for automatic recognition purposes, compared to the current RSL DBs.
The authors focus on – besides of providing the technical features of the database – the principles of annotating the collected data. The phonological annotation is combined with segmentation and labeling, which makes the presented database applicable for a wide range of tasks.
Another feature of TheRuSLan is multimodality. Like any natural language, SLs make use of several streams of information, also referred to as modalities. The fusion of these modalities provides researchers with additional bonuses and has been being actively investigated. In (Deng and Tsui, 2002), parallel HMMs were used to combine modalities for ASL, with accuracy over 90%.
Nowadays, many DNN-based models are used for SL recognition. 3D CNN-based approaches (Ji et al., 2010; Escalante et al., 2016) to gesture recognition tasks showed satisfactory results. Hence, an approach to multimodal (color video stream and depth map) recognition of SLs seems to be quite promising.
In (Kagirov et al., 2019) the dataset based on TheRuSLan was used in an approach to the multimodal recognition of dynamic and static gestures of RSL through 3D convolutional and LSTM neural networks and achieved recognition accuracy of 68.31% for color video sequences, 64.93% for data obtained from the depth map, and 73.25% for the multimodal dataset with two combined modalities. Thus, combining modalities significantly increases recognition results.
Nowadays, TheRuSLan comprises a rather narrow subject area, however, future expansion of the vocabulary can create a larger gesture corpus. In the future, segmentation and analysis of two-handed gestures is planned.

8. Acknowledgements

The research is supported by the Ministry of Science and Higher Education of the Russian Federation, under the agreement № 14.616.21.0095 (ID Number: RFMEFI61618X0095).

9. Bibliographical References

Agris, U., Knorr, M. and Kraiss, K.-F. (2008). The significance of facial features for automatic sign language recognition. In Proceedings of 8th IEEE International Conference on Automatic Face & Gesture Recognition (FG’08), pages 1–6, Amsterdam, The Netherlands, September 17-19.
Arbeitsgruppe Fachgebärdenlexika am Zentrum für Deutsche Gebärdensprache (1996). Fachgebärdenlexikon: Psychologie. Signum Verlag, Hamburg.
Battison, R. (1978). Lexical Borrowing in American Sign Language.: Linstok Press, Silver Spring, MD.
Chen, M., AlRegib, G. and Juang, B.H. (2012). 6dmg: a new 6d motion gesture database. In Proceedings of the Second ACM Multimedia Systems Conference (MMSys’12), pages 83-88, Chapel Hill, NC, USA, February 22-24.
Deng, J. and Tsui, H.T. (2002). A two-step approach based on PaHMM for the recognition of ASL. In Proceedings of Fifth Asian Conference On Computer Vision, pages 126-131, Melbourne, Australia.
Escalante, H., Ponce-López, V., Wan, J., Riegler, M., Chen, B., Clapés, A., Escalera, S., Guyon, I., Baró, X., Halvorsen, P. and Müller, H. (2016). Chalearn joint contest on multimedia challenges beyond visual analysis: An overview. In Proceedings of 23rd International Conference on Pattern Recognition (ICPR’16), pages 67–73, Cancun, Mexico, December 4-8.
Escalera, S., Gonzàlez, J., Baró, X., Reyes, M., Lopes, O., Guyon, I., Athistos, V. and Escalante, H.J. (2013) Multi-modal gesture recognition challenge 2013: dataset and results. In Proceedings of the 15th ACM International Conference on Multimodal Interaction, (ICMI’13), pages 445-452, Sydney, Australia, December 9-13. Available at: http://sunai.uoc.edu/chalearn/
Forster, J., Schmidt, C., Koller, O., Bellgardt, M. and Ney, H. (2014). Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-Weather. In Proceedings of International Conference on Language Resources and Evaluation (LREC’14), pages 1911-1916, Reykjavik, Iceland, May 26-31.
Frischberg, N., Hoiting, N. and Slobin, D.I. (2012). Transcription. In R. Pfau et al. Sign Language. An International Handbook. Berlin – Boston: De Gruyter Mouton, pp. 1045--1075.
Garcia, B. and Sallandre, M.-A. (2013). Transcription systems for sign languages: a sketch of the different

graphical representations of sign language and their characteristics. In C. Müller et al. *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. Vol. 1*. Berlin – Boston: De Gruyter Mouton, pp. 1125--1140.

Gruber, I., Ryumin, D., Hruz, M. and Karpov, A. (2018). Sign Language Numeral Gestures Recognition Using Convolutional Neural Network. In Proceedings. of Third International Conference on Interactive Collaborative Robotics (ICR'18), pages 70-77, Leipzig, Germany, September 18–22.

Hanke, T. (2004). HamNoSys – Representing Sign Language Data in Language. Resources and Language Processing Contexts. Available at: https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/HankeLRECSLP2004_05.pdf.

Hanke T. and Storz J. (2008). Ilex – a database tool for integrating sign language corpus linguistics and sign language lexicography. In Calzolari N. et al., editors, Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC'08), pages 64—67, Marrakech, Morocco, May 28-30.

Ji, S., Xu, W., Yang, M. and Yu, K. (2010). 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:221–231.

Johnston, T. (2001). The lexical database of Auslan (Australian Sign Language). *Sign Language and Linguistics*, 4(1/2):45–169.

Karpov, A. and Kagirov, I. (2011). Formalization of the Lexicon of a Sign Language Computer Synthesis System. *SPIIRAS Proceedings*, 1(16):123–140. (in Russian)

Kagirov, I., Ryumin, D. and Axyonov, A. (2019). Method for Multimodal Recognition of One-Hande d Sign Language Gestures through 3D Convolution and LSTM Neural Networks. In Proceedings of 21st International Conference Speech and Computer (SPECOM'19), pages 191–200, Istanbul, Turkey, August 20–25.

Kagirov, I., Ryumin, D., Axyonov, A. and Karpov, A. (2020). Multimedia Database of Russian Sign Language Gestures in 3D. *Voprosy Jazykoznaniya*, 1:104-123. (in Russian)

Kennedy, G., Arnold, R., Dugdale, P., Fahey, S. and Moskovitz, D. (1997). A Dictionary of New Zealand Sign Language. Auckland University Press, Bridget Willam Books, Auckland.

Kharlamenkov, A. (2017). Sign Language Vocabularies Review. Available at: https://surdocentr.ru/publikatsii/obzory-slovarej-zhestovykh-yazykov

Li, Y., Tian, K., Fan, Y., Xu, X. and Li, R. (2016). Video Gesture Recognition with RGB-D-S Data Based on 3D Convolutional Networks. In Proceedings of International Conference on Pattern Recognition (ICPR'16). Available at: http://chalearnlap.cvc.uab.es/dataset/21/description/#

Liu, L. and Shao, L. (2013). Learning discriminative representations from RGB-D video data. In Proceedings of International Joint Conference on Artificial Intelligence (IJCAI'13), pages 1493-1500, Beijing, China, August 3-9. Available at: http://lshao.staff.shef.ac.uk/data/SheffieldKinectGesture.htm

Malgireddy, M.R., Inwogu I. and Govindaraju V. (2012). A temporal Bayesian model for classifying, detecting and localizing activities in video sequences. In Proceedings of 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'12), pages 43–48, Providence, RI. Available at: http://gesture.chalearn.org/data

Pisharady, P.K. and Saerbeck, M. (2013). Robust gesture detection and recognition using dynamic time warping and multi-class probability estimates. In Proceedings of IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing, (CIMSIVP'13), pages 30-36, Singapore. Available at: http://www.ece.nus.edu.sg/stfpage/elepv/NUS-HandSet/

Pisharady, P.K. and Saerbeck, M. (2015) Recent methods and databases in vision-based hand gesture recognition: A review. *Computer Vision and Image Understanding*, 14:152–165.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T. and Henning, J. (1989). HamNoSys. Version 2.0; Hamburg Notation System for sign languages. An introductory guide. Signum Verlag, Hamburg.

Simon, F., Helena, M.M., Pushmeet, K. and Sebastian, N. (2012). Instructing people for training gestural interactive systems, in Proceedings of International Conference on Human Factors in Computing Systems (CHI'12), pages 1737–1746, Austin, Texas, USA, May 5-10. Available at: http://research.microsoft.com/en-us/um/cambridge/projects/msrc12/

Stokoe, W.C. (1960). Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. Dept. of Anthropology and Linguistics, University of Buffalo, Buffalo.

10. Language Resource References

RSLC 2014 — Russian Sign Language Corpus. Available at http://rsl.nstu.ru/.

TheRuSLan 2019 – Thesaurus Russian Sign Language. Available at http://hci.nw.ru/en/projects/17#ruslan (currently, a very limited number of items).