Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 176 (2020) 3263–3272

Procedia Computer Science

www.elsevier.com/locate/procedia

24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Dealing with Polysemy in the Polish Sign Language Using the OWL Ontology

Wojciech Kozioł[a,*], Krzysztof Pancerz[a], Kazimierz Sikora[b], Kamil Dudek[a]

[a] *College of Natural Sciences, University of Rzeszów, Pigonia Str. 1, 35-310 Rzeszów, Poland*
[b] *Faculty of Polish Studies, Jagiellonian University, Gołębia Str. 16, 31-007 Kraków, Poland*

Abstract

A common problem of natural language processing is synonymy, polysemy, and homonymy. In the paper, we propose to deal with polysemy in the Polish sign language using the knowledge included in the OWL2 ontology created for this purpose. The proposed approach aids the translation process of the Polish sign language into the Polish language by selection from the possible phrases, only those, with the reasonable meaning.

Keywords: Modelling; Polysemy; Polish Sign Language; OWL Ontology

1. Introduction

The problem of polysemy in different sign languages was considered in the literature (cf. [2], [9], [13]). This problem can be considered, among others, from the point of view of cognitive linguistics as well as from the point of view of machine translation. The goal of our research is to deal with polysemy in the Polish sign language in terms of computer translation processes. The translation process of signs requires some external knowledge aiding us to recognize proper meanings of signs. In each language, the problem of polysemy is solved in practice by taking into consideration the contexts of utterances. We take note of this fact in Section 2. The context of a given utterance can be identified by means of looking at individual words (concepts) together with their neighbourhoods, i.e., with their adjacent words (concepts). In each vocabulary, we can determine co-occurrences of words (concepts) in phrases (sentences). To express those co-occurrences of words (concepts) in a formal way, we propose to use ontology. Recently, ontologies have become powerful tools in modern computer systems requiring semantic and well-structured knowledge bases covering different aspects of information that is processed (cf. [3]). Ontologies are also used in the context

* Corresponding author. Tel.: +48-17-8518531.
E-mail address: wkoziol@ur.edu.pl

1877-0509 © 2020 The Authors. Published by Elsevier B.V.

10.1016/j.procs.2020.09.122

of sign languages (cf. [5], [6]). Particularly, the OWL ontologies [8] offer a machine-friendly format for information processing such as XML. A brief description of the created OWL ontology is given in Section 3. This ontology is used in the dedicated application enabling us to recognize proper meanings of signs in utterances for the translation process into the Polish language. The application created by us is briefly described in Section 4. This description is supplemented with simple examples explaining the main idea of operations. Currently, utterances can be created only manually by assembling them from individual signs included in the data base. Because our research is devoted to the Polish sign language, examples of utterances are presented in their original forms (i.e., in the Polish language) and some translation into the English language is placed where appropriate. It is worth noting that the proposed approach can be easily implemented in other sign languages. Such implementation comes down to the development of the proper knowledge base in a form of the OWL ontology.

2. Problem with Polysemy

Problems with determining denotations of such linguistic notions as polysemy (multiple meanings) and homonymy have been repeatedly and comprehensively considered and commented in the scientific literature (cf. [1], [4]). Most linguists are inclined to the thesis that we are dealing with a phenomenon that results from the semantic poly-functionality of language units (conventionalized and approved in the dictionary - the words used have a specific contextual meaning), which boils down to the fact that more than one meaning can be attributed to a given unit, which is in itself something obvious and common in natural languages. Therefore, the proposed approach is insufficient because such semantic differences:

- are regular and can be reduced to a common source (e.g., *wiśnia* means both a cherry tree and a cherry fruit; *wejście* means both input - an action and an entrance) - polysemy,
- are irregular (accidental) and may result either from different origins (etymologies) of two apparently identical words (e.g., in Polish: *bal* means both a ball - from German *ballen* and a log - from Italian *ballare*) or from a long term development of the original meaning of the same word (e.g., in Polish: *pokój* means both a peace and a room) - homonymy.

Therefore, the way in which this linguistic phenomenon is approached depends on the point of view taken:

- if we come from a separate meaning, it leads to homonymy,
- if we come from the identity of the form, it leads to polysemy.

Thus, from a theoretical point of view, it is a less important problem, but it is of primary importance for lexicology and lexicography, which is obliged to present the semantic variance of lexical units in a language.

In practice, the postulate of the dictionary description: *do not multiply beings excessively* boils down to treating the most regular changes of meaning, the so-called universal polysemy (generating the most predictable variants) and specific polysemy (e.g., in Polish *klucz dzikich gęsi* - the key of the wild geese and *klucz* - the key to the test). A difficult question is which examples of semantic derivation (variance) should be included in the dictionary and which ones should be omitted.

The types of this polysemy are as follows:

- *plant* ↔ *plant food*, e.g., *spinach* and *fried spinach*,
- *author* ↔ *his/her work, image*, e.g., *I like Dickens*.
- *animal* ↔ *a meat of this animal*, e.g., *chicken* and *chicken stewed in leeks*,
- *tree* ↔ *wood of this tree*, e.g., *beech* and *cubic meter of beech*,
- *substance* ↔ *portion of this substance*, e.g., *juice* and *two juices please*,
- *material* ↔ *product from it*, e.g., *cotton* and *cotton still in fashion*.

From a semantic and logical point of view, the polysemy can be described as follows [1]. The meanings a_i and a_j of a word A are similar if there are such levels of semantic description on which their definitions (semantic trees) or

Fig. 1. (a) Example 1; (b) Example 2; (c) Example 3.

connotations have a non-trivial (sufficiently explicit) common part and if it fulfils the same function in definitions in relation to other semantic components. Word A is called ambiguous if, for each of its two meanings a_i and a_j, there are such meanings a_1 and $a_2, \ldots, a_k, a_l$, that a_i is similar to a_1, a_1 to a_2, and so on, a_k to a_l, and a_l to a_j.

How does a language in practice solve the problem of polysemy? Such a diagnostic (verification) function is performed by the context, e.g., *I grow spinach - I eat spinach*, *A good confectioner does not fry the brushwood on a gas plate* [1]. However, the problem of the description of the polysemy cannot be underestimated, because the knowledge about the semantic potential of a lexical unit is derived from the dictionaries from which such knowledge is required. Hence, it belongs to the useful knowledge on a language.

The phenomenon of polysemy also occurs in a sign language, probably to a comparable or greater extent as in phonic languages. It is an obvious consequence of the fact that the number of signs is limited, forcing an increase in their semantic poly-functionality. It should be noted that the creation and understanding of sign texts involves a specific sphere of spatial imagination (semantization of space), and requires a cognitive approach in research, allowing (spatial metaphor) to establish possible vectors of association. This is well illustrated in [12].

Estimating the actual scope of polysemy in the Polish sign language is very difficult, it can only be inferred from partial data, mainly dictionary data. We were encouraged to take up this topic by sign language polysemy research, conducted in 2019 by one of the co-authors of this paper - Kamil Dudek (the Polish sign language native speaker). On this basis, a group of lexems was identified, which are characterized by the type of semantic variance described above. It should be noted that the sign language is geographically and socio-linguistically diverse. The materials collected are relevant for its dialect from Rzeszów and its surroundings, used by the young generation. Due to the lack of space we will limit ourselves to a few selected examples.

Example 1: *kwiat* (*flower*) - *maj* (*May*) - *wiosna* (*spring*) - see Figure 1(a). *Kwiat*, as a lexem, has a specific object meaning (*flower*). A side meaning *May* is created by the association *month when there are many flowers*, also known in phonic languages (e.g., in Czech May is *květen*). Its gradational development leads to the meaning of *spring*.

Example 2: *gość* (*guest*) - *klient* (*client*) - *super człowiek* (*superman*) - *bogaty* (*rich man*) - see Figure 1(b). The lexem *Gość* replicates a cognitive structure important in the Polish culture, *gość* is someone important. The shift in meaning is therefore axiological and expressive: a very positive evaluation, admiration, the highest appreciation. Today, it concerns e.g. great sportsmen and social authorities. It can also be e.g. a priest or a friend. The same type of metonymy exists in common Polish.

Example 3: *opowiadać* (*to tell*) - *śpiewać* (*to sing*) - see Figure 1(c). The Polish sign language does not make a lexical difference between these senses. From the point of view of a deaf person, the perceptual difference between these verbal activities is not very expressive, it is based on visual and motoric sensations (including the sense of rhythm externalised in a body language). Let us add that deaf persons also have fun with music, but only if it is properly loud and directed (e.g., speakers on the floor).

3. OWL Ontology

In this section, we give the outline of the OWL ontology created as a central unit of our application described in Section 4. An increasing attention has been focused recently on ontologies since modern computer tools require semantic and well-structured knowledge bases covering different aspects of information that is processed. Ontological engineering refers to the set of activities that concerns the ontology development process, the ontology life cycle, and the methodologies, tools and languages for building ontologies, see [7]. In computer science, ontologies are used to capture knowledge about some domain of interest. In our research, this domain concerns the Polish sign language. There are many definitions and interpretations of the term *ontology* in the literature. One can find a number of them in [7]. In general, an ontology describes:

- concepts in a given domain of interest,
- instances of concepts,
- properties representing:
  - semantic relations expressing various types of associations between instances,
  - various features of instances,
- restrictions on properties.

Semantic relations are very important components in ontology modelling as they describe the relationships that can be established between instances of concepts. A comprehensive review of the literature concerning semantic relations is given in [11]. As the authors noticed, almost every new attempt to analyse semantic relations leads to a new list of relations. A special role is played by three basic semantic relations between concepts:

- *SUBCLASS-OF* (hyponymy), also known as *IS-A*. If c *SUBCLASS-OF* c', it means that c is a kind of c' (c is a more specialized concept than c').
- *PART-OF* (meronymy). If c *PART-OF* c', it means that c is a part of c'.
- *INSTANCE-OF*. If i *INSTANCE-OF* c, it means that i is an instance (example) of c.

More specific semantic relations used in our ontology are described in Subsection 3.3.

One of the key decisions to take in the ontology development process is to select the language in which the ontology will be implemented. Our ontology is built in accordance with the OWL 2 Web Ontology Language (shortly OWL 2) [8]. An OWL ontology consists of three components:

- classes (representations of concepts from the domain of interest, interpreted as sets that contain individuals),
- individuals (instances of classes that represent objects in the domain of interest),
- properties (binary relations on individuals).

There are two main types of properties in OWL 2:

- object properties linking an individual to an individual,
- data properties linking an individual to a data value.

To implement the ontology, we have used Protégé [10] that is a free, open source, platform-independent environment for creating and editing ontologies and knowledge bases (see Figure 2).

3.1. Classes

At the first level of the OWL ontology, we have distinguished classes representing categories of parts of speech as it is shown in Table 1. Moreover, some of the main classes are specialized by the subclasses categorizing parts of speech more precisely in terms of real-world meaning, for example, *Noun* → *living being*.

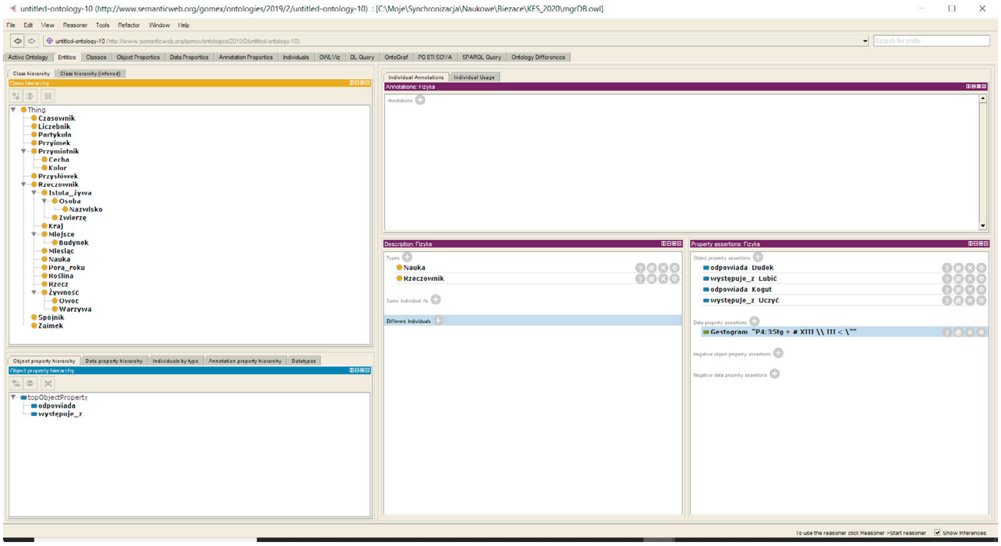

Fig. 2. A part of the OWL ontology in the Protégé environment.

Table 1. Classes representing categories of parts of speech.

Class		Examples of subclasses	
English Term	Polish Term	English Term	Polish Term
Verb	Czasownik		
Numeral	Liczebnik		
Particle	Partykuła		
Preposition	Przyimek	feauture, colour	cecha, kolor
Adjective	Przymiotnik		
Adverb	Przysłówek		
Noun	Rzeczownik	living being, place, plant, season	istota żywa, miejsce, roślina, pora roku
		object, food	rzecz, żywność
Concjunction	Spójnik		
Pronoun	Zaimek		

3.2. Instances

There is a huge number of words used in every natural language. Moreover, in our tool, we are interested in dealing with polysemy in the Polish sign language. Therefore, to add instances (representing words of the Polish language) to our ontology, we have focused primarily on groups of words having the same gestures in the Polish sign language.

3.3. Object Properties

In our OWL ontology, we have distinguished two main object properties, shown in Table 2, expressing the relationships between individuals.

Table 2. Object properties expressing the relationships between individuals.

English Term	Polish Term	Remark
corresponds to	odpowiada	expresses polysemy of gestures (i.e., polysemy at the level of the Polish sign language)
appears with	występuje z	expresses coincidence (co-occurrence) of words (concepts) in the Polish language

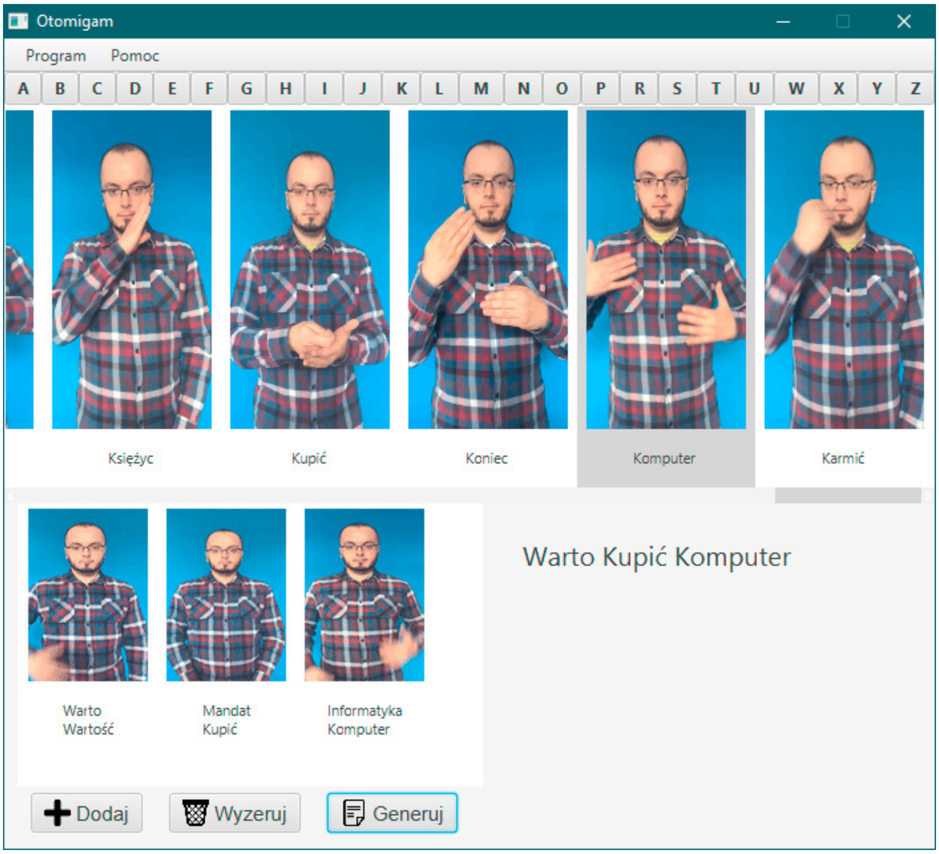

Fig. 3. *OtomigamOntology* - a general view.

3.4. Data Properties

In our OWL ontology, data properties are used to attribute useful entities to the signs. Therefore, data properties assign:

- links to video files with gestures,
- gesture codes.

to individuals. The first attribution is used in the visual presentation of utterances. The second attribution plays an important role if the signs are stored in the textual form. For example the sign representing the term *Fizyka* (*Physics*) has the following gesture code:

```
P4:35tg + # XIII \ III < "
```

4. Computer Tool

To perform our research, a computer application, called *OtomigamOntology*, has been created (see Figure 3). This application uses the ontology, described in Section 3, in the process of translation of simple utterances from the Polish sign language to the Polish language. The interface is created in the Java language with the use of JavaFX technology. Data from the ontology are loaded after running the application. Next, a list (dictionary) of gestures and their meanings is presented in the form of described videos. The dictionary in the form of a horizontal list is in the upper part of the

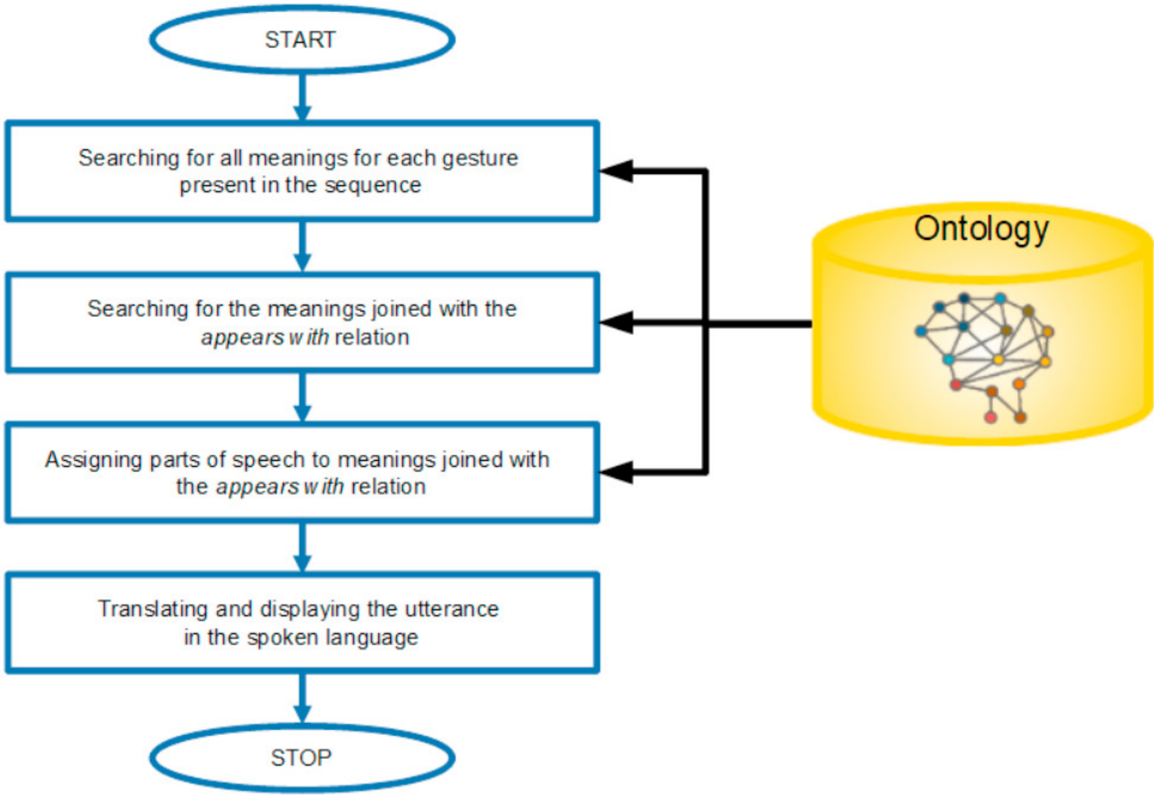

Fig. 4. The procedure.

application window. Individual gestures are displayed in the form of pictures. However, when a mouse is over them, a gesture animation is shown in the form of video. To make a search easier for individual gestures in the dictionary, they are grouped alphabetically. The user can thus browse the list of gestures and select the appropriate one and add it with the *Add* button. In this way, the utterance in the sign language is created as a list of gestures below the dictionary. The pictures in this list are slightly smaller and they contain captions describing the meaning of the selected gestures. One can clear the list using the *Reset* button. In the example, the user has chosen the gestures that make up the sentence *Warto kupić komputer* (*It is worth to buy a computer*). It is worth noting that these gestures also have other meanings that will not be used in this context.

When the utterance is created by the user in the sign language, it can be translated into the Polish language using the *Generuj* button. This triggers the course of actions as it is shown in the diagram in Figure 4.

Example 1. The example illustrates the operation of the program leading to the reduction of polysemy (ambiguity) as a result of three diagnostic contexts. The ambiguous *zostać/pozwolić* (*to become/to allow*) combined with an element of the right-hand side context, allows a semantically acceptable phrase: *X został bogatym dyrektorem* (*X became a rich director*), while eliminating the others: *X został gościem* (*X became a guest*), *X pozwolił gościa* (*X allowed a guest*). Although, *X został gościem* (*X became a guest*) would be a possible phrase, but the system eliminates it by a preferential choice of the longest reasonable sequence of units. There is therefore only one solution for the assumed condition: *X został bogatym dyrektorem* (*X became a rich director*), see Figure 5.

Example 2. This example shows a situation when, after selecting three gestures, the algorithm generates two separate, correct results:

- *Nadchodzi ciepła wiosna* (A warm spring is coming).
- *Nadchodzi ciepły maj* (A warm May is coming).

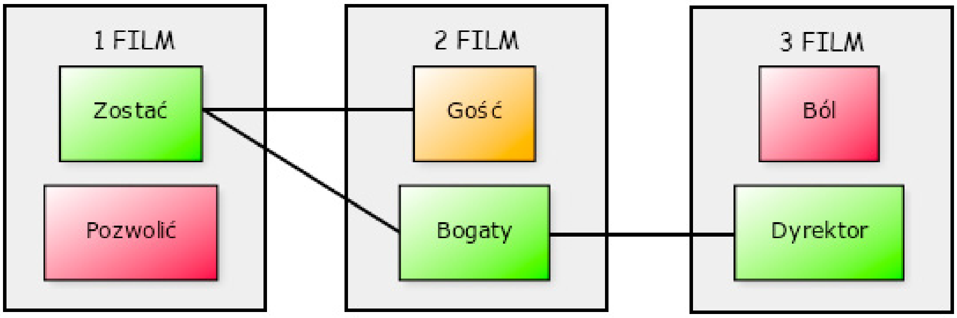

Fig. 5. The example of reduction (1).

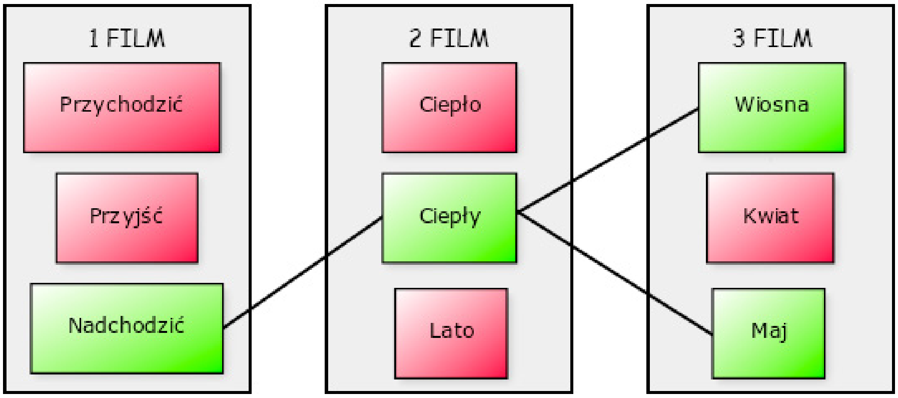

Fig. 6. The example of reduction (2).

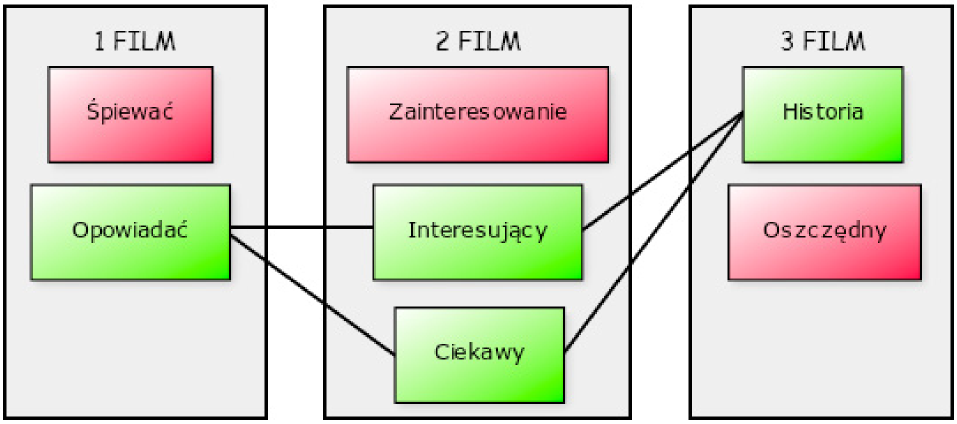

Fig. 7. The example of reduction (3).

Lexems *wiosna* (*spring*) and *maj* (*May*) have a common general semantic characteristic (a temporal term). Therefore, the algorithm found no contradiction at the level of contextual variance and allowed two possibilities (accurate).

Example 3. This example shows that the algorithm proposes two different results:

- *X opowiada interesującą (zajmującą, wciągającą) historię* (*X tells an interesting (engaging) story*).

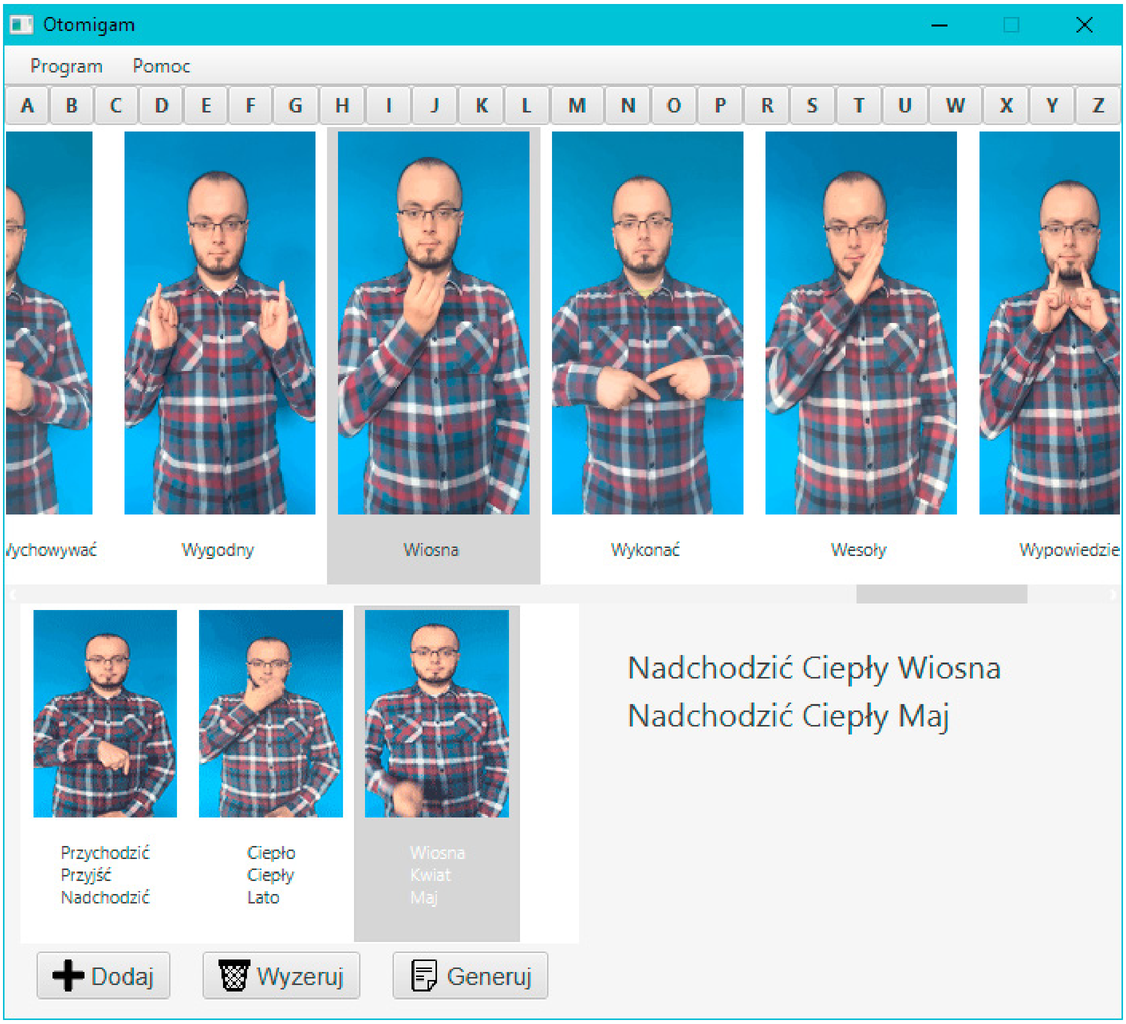

Fig. 8. The screenshot for Example 2.

- X opowiada ciekawą historię (*X tells an interesting (in a sense of the cognitive aspect) story*). The situation described refers to polysemy of the adjective, sometimes difficult to grasp. The algorithm allows two equivalent realizations, equally entitled to the term in a given nominal phrase defined by the connotation requirements of the lexem *historia* (*story*). It rules out such connections to the lexem *oszczędny* (thrifty) and earlier *śpiewać zainteresowanie* (*to sing the interest*).

The application allows the user to view videos and their meanings. In the main program window, all sign language gestures are displayed in the form of videos, which are included currently in the database. The screenshot for Example 2 is shown in Figure 8.

5. Conclusions

We have shown that the ontology is a useful tool to express the knowledge enabling us to deal with the problem of polysemy in the sign language. In our research, we have created the OWL ontology that can be easily processed and searched. At the first step of our research, we considered manual creation of utterances using a dedicated application created by us. The challenge is to deal with polysemy in real-time. This challenge requires developing the procedure for sign recognition during real utterances, especially recognition of sign sequences. This is one of main directions of our research. Another important direction concerns the extension of the ontology developed in our research. Among others, broaden range of vocabulary should be taken into consideration. However, implementation of various semantic relations between concepts can aid proper recognition of meaning of individual signs and the whole utterances.

Acknowledgements

This work was partially supported by the Center for Innovation and Transfer of Natural Sciences and Engineering Knowledge at the University of Rzeszów.

References

[1] Apresjan, J., 2000. Semantyka leksykalna (Lexical Semantics). Zakład Narodowy im. Ossolińskich.
[2] Baker, A., van den Bogaerde, B., Pfau, R., Schermer, T., 2016. The Linguistics of Sign Languages: An introduction. John Benjamins Publishing Company.
[3] Breitman, K.K., Casanova, M.A., Truszkowski, W., 2007. Ontology in computer science, in: Semantic Web: Concepts, Technologies and Applications. Springer, London, pp. 17–34. doi:10.1007/978-1-84628-710-7_2.
[4] Czelakowska, A., 2014. Polisemia regularna - między semantyką a kontekstem (Regular polysemy - between semantics and context). LingVaria IX, 45–58. doi:doi:10.12797/LV.09.2014.17.03.
[5] Elsayed, E.K., Fathy, D.R., 2020. Sign language semantic translation system using ontology and deep learning. International Journal of Advanced Computer Science and Applications 11. doi:10.14569/IJACSA.2020.0110118.
[6] Felice, M., Di Mascio, T., Gennari, R., 2007. A visual ontology-driven interface for a web sign language dictionary, in: Weske, M., Hacid, M.S., Godart, C. (Eds.), Proceedings of the Web Information Systems Engineering Workshops (WISE'2007). Springer, Berlin, Heidelberg. volume 4832 of *Lecture Notes in Computer Science*, pp. 429–440.
[7] Gomez-Perez, A., Fernandez-Lopez, M., Corcho, O., 2004. Ontological Engineering. Springer-Verlag, London.
[8] Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S., . OWL 2 Web Ontology Language Primer. URL: https://www.w3.org/TR/owl2-primer/.
[9] Johnston, T., Schembri, A., 2007. Australian Sign Language: An introduction to sign language linguistics. Cambridge University Press.
[10] Musen, M.A., 2015. The protégé project: A look back and a look forward. AI Matters 1, 4–12. doi:10.1145/2757001.2757003.
[11] Nastase, V., Nakov, P., Séaghdha, D.O., Szpakowicz, S., 2013. Semantic Relations Between Nominals. Morgan & Claypool Publishers.
[12] Przybylska, R., 2002. Polisemia przyimków polskich w świetle semantyki kognitywnej (Polysemy of Polish prepositions in the light of cognitive semantics). Universitas.
[13] Valli, C., Lucas, C., 2000. Linguistics of American Sign Language: An Introduction. Gallaudet University Press.