

The implementation of the example-based machine translation technique for Spanish-to-Polish automatic translation system

MIROSLAW GAJER

High-quality machine translation between human languages has for a long time been an unattainable dream for many computer scientists involved in this fascinating and interdisciplinary field of the application of computers. The developed quite recently example-based machine translation technique seems to be a serious alternative to the existing automatic translation techniques. In the paper the usage of the example based machine translation technique for the development of system, which would be able to translate an unrestricted Spanish text into Polish is proposed. The new approach to the example-based machine translation technique that takes into account the peculiarity of the Polish grammar is developed. The obtained primary results of the development of proposed system seem to be very promising and appear to be a step made in the right direction towards a fully-automatic high quality Spanish-into-Polish machine translation system for unrestricted text.

Key words: natural language processing, computational linguistics, machine translation

1. Introduction

Machine translation is a science that delivers the knowledge how to program the computers, so as they were able to translate between human languages, for example, between Danish and Bulgarian. It may be amazing, but the field of machine translation is almost as old as the invention of computer itself [4]. On 15th of July 1949 an American cryptographer Warren Weaver presented his memorandum to The Rockefeller Foundation (American institution supporting the scientific research), in which he demanded starting the research on the automation of translation between natural languages [1]. Warren Weaver was inspired by cryptographic techniques, which were developed very strongly during the years of The Second World War, and he thought that there existed some fundamental similarities between these cryptographic techniques and the process of translation between human languages [21].

The author is with Department of Automatic Control, Academy of Mining and Metallurgy, Al. Mickiewicza 30 30-059 Crawow, Poland. e-mail: mgajer@ia.agh.edu.pl

This work was supported from the KBN grant No. 8T1101917 "The formal methods and tools of databases and knowledgebases analysis and design".

Received 28.01.2002, revised 23.05.2002.

2. History of the development of machine translation systems

The first research group dedicated especially to machine translation was established in Austin in the United States in 1951 [20]. The first public demo of an operating machine translation system was given also in the USA in 1954. During this demo the system translated 49 pre-selected sentences from Russian into English. The system was using a vocabulary of 250 words and only six simple grammatical rules. The possibilities of early machine translation systems were very far from this, what had been expected, and many scientists connected with the field started to be disappointed. In 1966 the ALPAC (Automatic Language processing Advisory Committee) published its famous report, concluding that machine translation was slower, less precise, and more expensive than human translation. As a result, funding of this type of projects was cut [20]. The renaissance came in the late 1970s. The United States Air Force funded work on the METAL system at the University of Texas in Austin, and the results of the work of TAUM group led finally to the installation of the METEO system, which was a great commercial success. It is worth to notice that the METEO machine translation system is still in use, and it translates every day from English into French more than 50,000 words of weather forecast bulletins.

Now, a still growing interest of machine translation systems can be observed in many countries, especially in Japan, the United States, the European Union, and India [3], but after so many years of an intensive scientific research high-quality machine translation between human languages for unrestricted text is still a long-term scientific dream of enormous political, social, and scientific importance [14]. Machine translation was also one of the earliest applications suggested for the computers, but turning this scientific dream into reality has turned out to be much harder, and much more interesting than it had first appeared [1].

3. Difficulties with translation

To answer the question about the origin of difficulties with automation of translation between human languages, let us consider the differences which we can discover when we compare some of the human languages. First of all, when we study grammatical systems of any natural languages that are not closely related with each other, we easily can see that there exist much more differences than similarities between them [23].

For example let us compare the systems of personal pronouns of Arabic and Hungarian languages.

Personal pronouns system of Hungarian:

Singular	Plural
1. <i>én</i>	1. <i>mi</i>
2. <i>te</i>	2. <i>ti</i>
3. <i>ő</i>	3. <i>ők</i>

Personal pronouns system of Arabic:

Singular	Double	Plural
1. <i>ana</i>	1. <i>nahnu</i>	1. <i>nahnu</i>
2. (m.) <i>anta</i>	2. <i>antum</i>	2. (m.) <i>antum</i>
2. (f.) <i>anti</i>		2. (f.) <i>antunna</i>
3. (m.) <i>huua</i>	3. <i>huma</i>	3. (m.) <i>hum</i>
3. (f.) <i>hija</i>		3. (f.) <i>hunna</i>

It's clear that personal pronouns system of Arabic is much more complicated than this one of Hungarian. It is caused by the fact that Hungarian language doesn't know such invention as grammatical gender of the words. Also grammatical number in Hungarian can be only singular or plural, whereas in Arabic it can be singular, plural, or double.

So, one can easily see that translating Hungarian personal pronouns into their Arabic equivalents is a hard task. For example, if we want to translate Hungarian pronoun *ők* (in English *they*) into Arabic we must additionally know how many persons are involved with this pronoun *ők*. If exactly two persons are considered we will use the Arabic word *huma*. But, if there are more than two persons we must additionally know, whether they are men or women. If they are men we will use the Arabic word *hum*, in other case *hunna*.

Where do we know from how many persons are involved, and whether they are men or women, while Hungarian word *ők* states nothing about it? The answer is that we know this from the context of the utterance. A human translator can in most cases very easily extract such context information, but the full automation of this process is still a pure science-fiction. A quite big differences between human languages can also be noticed when we study their vocabularies. In fact, the vocabulary of each language is an independent and very compound system. If we want to translate, for example, from Chinese into Croatian it's a hard work to find in Croatian the equivalents of Chinese words that preserve their original meanings. While doing that a human translator have to cope with a enormous number of lexical holes, it is, words that don't have their equivalents in other language, and as such can be translated only by the medium of a long description that clears up their semantics. This situation is illustrated in Fig. 1. In Fig. 1 each rectangle is a symbol of some physically existing object or some abstract entity. The rectangles are numerated from 1 to 6. Further, we have two different natural languages: language A and language B. We can see that in language A objects 1 and 2 are described only by one common lexical entity, whereas in language B there exist two different lexical entities, separate for object 1 and object 2. Further, we can notice that the object 3 doesn't have any lexical entity in language A, so it is a lexical hole, whereas in the language B it has its own lexical item. Objects 4 and 5 in language A are grouped together in one lexical entity and object 6 is a separate lexical entity, while in language B it is otherwise. We can notice that objects 5 and 6 form one lexical entity.

A very good example of this divagation (maybe a bit too abstract) comes from Swedish language. If we want to translate English word *grandfather* into Swedish, we

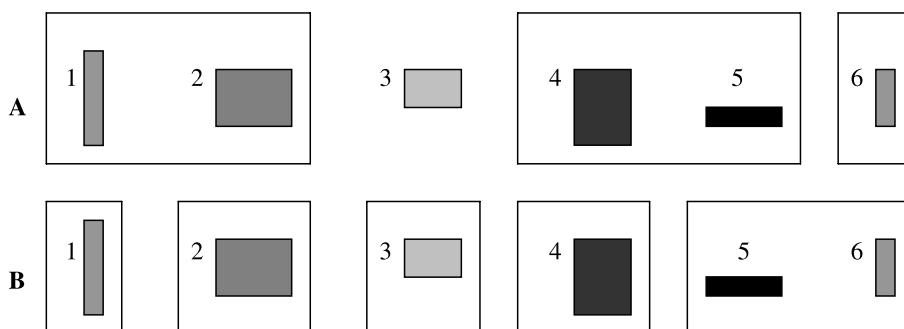


Figure 1. The illustration of the way in which different languages divide the reality into lexical items

must additionally know whether this grandfather is a father of a father or a father of a mother. In the first case we should use the Swedish word *farfar* in the other *morfar*, which is illustrated in Fig. 2.

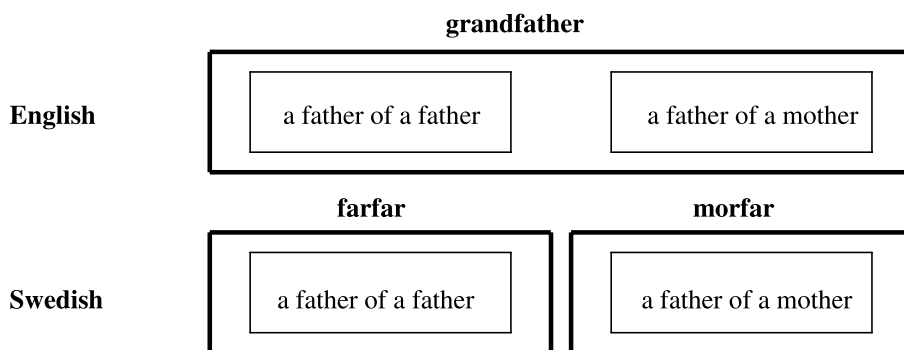


Figure 2. An English word *grandfather* versus Swedish words *farfar* and *morfar*

Another similar example comes from French language. Namely, if we want to translate an English word *river* into French, we must know whether it is a main river, which is directly connected with the sea, or it is only a tributary of some bigger river. We absolutely must have this information because in the first case we have to use French word *fleuve*, and in the second the correct choice is French word *rivièr*. This situation is illustrated in Fig. 3.

But the perhaps most serious problem, which the computer has to cope with in machine translation is the ambiguity of any human language [2]. We can distinguish syntactic ambiguity when there exist at least two alternative ways of syntactic analysis of a sentence. One of the examples is an English sentence:

I see a man in the park with a telescope.

This sentence is threefold ambiguous because we don't know if the phrase with the *telescope* should be interpreted in connection with a verb *to see*, or with the noun *a man*,

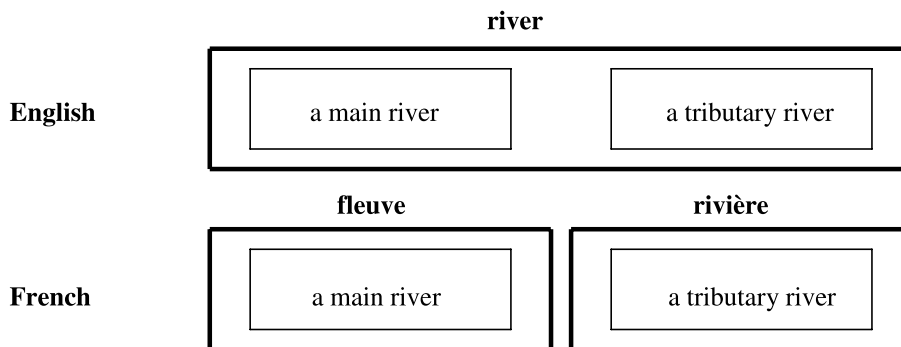


Figure 3. An English word *river* versus French words *fleuve* and *rivière*

or with the noun *the park*. In each of these cases the meaning of the sentence is totally different. The problem is that this ambiguity cannot be preserved during the translation because in order to translate, one has to understand the sentence being translated. For example, in the case of translating this sentence into Polish three different translations are possible, depending on the interpretation of the English sentence:

Widzę człowieka w parku za pomocą teleskopu.
Widzę w parku człowieka z teleskopem.
Widzę człowieka w parku z teleskopem.

Another example of the ambiguity on the syntactic level is an English phrase:

old man and women

Analyzing this phrase we don't know whether it is:

old man and old woman

or

old man and woman at any age

Another kind of ambiguity is at the semantic level. Semantic ambiguity appears when one sentence can be understood in at least two different manners [22]. A good example is an English sentence:

She threw the vase at the window and it broke.

This sentence is ambiguous because we don't know what is broken, *a window* or *a vase*? If we would like to try to translate this sentence into Polish we would have a hard choice to make. If we decided that the window is broken, the Polish translation would be:

Ona rzuciła wazę w okno i ono pękło.

In the other case we would obtain the following Polish translation:

Ona rzuciła wazę w okno i ona pękała.

Another kind of ambiguity is the ambiguity at the lexical level of language analysis. Lexical ambiguity is such a serious problem in the case of machine translation systems because it exists in every natural language and it is really ubiquitous. In fact, if we open any bilingual dictionary, for example, *The Great English-Polish Dictionary*, it's very hard to find a word that would have only exactly one meaning. In fact, most of English words have at least two completely different Polish equivalents. So, the question is, which one of them the computer should choose while translating, and where can computer know from, which one of them is the correct one?

Let us suppose that we have a sentence built from ten different word, and let each of these words has exactly two different meanings. If the computer chose the equivalents of these words at random this sentence could be translated in 1024 different ways. The probability that acting this way we obtain a correct translation of a whole document built from many such sentences is equal to zero in practice. Moreover, no efficient algorithm that allows for solving this problem is known, and lexical ambiguity can be found in abundance in any human language - bellow are listed some examples of possible Polish translations of lexically ambiguous words taken from several languages of the world.

Polish equivalents of French word **perle** are: 1. perła, 2. paciorek, 3. kapsułka.

Polish equivalents of Spanish word **fondo** are: 1. dno, 2. głębia, 3. tło.

Polish equivalents of Italian word **stufa** are: 1. piec, 2. ciepłarnia.

Polish equivalents of German word **Absatz** are: 1. ustęp, 2. obcas, 3. osad, 4. złoże, 5. osadzenie, 6. zbyt.

Polish equivalents of English word **butt** are: 1. beczka, 2. pień, 3. pniak, 4. grubszy koniec, 5. kolba karabinu, 6. płastuga, 7. nasyp za strzelnicą, 8. półmiewisko, 9. uderzenie głową.

Polish equivalents of Dutch word **boodschap** are: 1. poselstwo, 2. polecenie, 3. wiadomość, 4. zakupy.

Polish equivalents of Swedis word **tomten** are: 1. parcela, 2. plac, 3. krasnoludek.

Polish equivalents of Norwegian word **hytte** are: 1. chata, 2. szałas, 3. buda, 4. huta, 5. kabina.

Polish equivalents of Danish word **lober** are: 1. biegacz, 2. dywanik.

Polish equivalents of Finnish word **kanta** are: 1. podstawa, 2. obcas, 3. stanowisko, 4. baza.

Polish equivalents of Greek word **σκοπός** (skopos) are: 1. zamiar, 2. melodia, 3. wartownik.

Very problematic for machine translation systems are also complex nominal groups like for example:

adult toy manufacturer

This nominal group can be understood as:

manufacturer of toys for adults

or

an adult manufacturer of toys

Another kind of difficulties making the automation of translation process so hard are all idiomatic phrases. The problem is that these idiomatic phrases can be also interpreted literary. For example, an idiom taken for the Hausa language:

Gari ya yi kyau.

means that:

It is a beautiful weather.

But when we treat this sentence literary it means that:

The town is beautiful.

So, which of this two meanings is the correct one? A human translator basing on the context of this idiom can probably make the right decision, but the automation of such inference is still far beyond the possibilities of any computer system.

Taking under consideration all above mentioned factors translation between natural languages can be seen as a highly creative process. A human translator must have a lot of invention and must know how to deal with the situations he had never met before. So, the right question is, whether it is possible to replace a human by a computer?

A prominent physicist Roger Penrose in his famous books on artificial intelligence, entitled "New Caesar's Mind" [19] and "The Shadows of the Mind" gave very strong arguments supporting his thesis that the human brain operates in a non-algorithmic manner and because of this fact a human mind cannot be fully simulated by computer. Thus, if we cannot replace a human by a computer does it also mean that a fully-automatic high-quality machine translation for unrestricted text is impossible [10], [16], [18], [14]? A philosopher Alan Melby in his paper [13] states that machine translation is headed in the right direction as far as domain-specific approaches using controlled languages are concerned. But further work on fully-automatic high-quality machine translation of unrestricted text is a waste of time and money. To build such systems a real breakthrough in natural language processing (and maybe in the whole filed of information processing) is required. Moreover, such breakthrough will not be based on any extension of currently known techniques, as electric bulb was not invented just because the research on the candle had been conducted [13].

4. Example-based machine translation technique

The arguments given by Roger Penrose are very strong and it's not possible to ignore them any further. So, probably Alan Melby is right that replacing a human translator

totally is not possible basing only on the currently known techniques. But, by using these currently known techniques we can still try to approach as close as possible to this unattainable goal, which is a fully-automatic high-quality machine translation for unrestricted text. Suppose that during the intensive scientific research we built a machine translation system, which gives a translation of 99% accuracy, while operating on an unrestricted text (only 1% of this text need to be approved by a human translator). So can we really say, like Alan Melby, that we had wasted time and money on this research?

Up till now, many totally different approaches to machine translations have been developed. These are, among others: syntactic transfer, interlingua-based machine translation, knowledge-based machine translation, systems based on statistics or neural nets, etc. [17], [8], [11]. Among these example-based machine translation is becoming a serious alternative paradigm, but in most cases it is still an unproven technique, which is in its early research phase [9].

But it is not always so. One prominent example comes from Spain. The case of the magazine entitled *Periódico de Catalunya* is interesting because it is probably the first fully operational machine translation system for translation of unrestricted text that has ever been built, which produces nearly hundred percent satisfactory results while translating from Spanish into Catalan. It is really amazing that this machine translation system is not based on any of the currently known computational linguistics theories. Moreover, it does not analyze the sentence in any way it only replaces source words (or groups of words) by their target equivalents, just like spelling-checker would do. The system has a huge dictionary that effectively replaces all linguistic analysis of the source text. The development of the systems requires a lot of work, in fact a quite big team of trained linguists constantly updates the dictionary with new terms, verbs in their different forms and sequences of words of up to six elements. Up till now, it has been probably the only practical implementation of a purely unsophisticated machine translation system basing only on a pattern-matching scheme [20].

So, can this Spanish-Catalan system be an example showing the way how to mysteriously solve the problem of building a fully-automatic high-quality machine translation system? The answer to this question is not so obvious as somebody may think. We can not omit the fact that this Spanish-Catalan system benefits to the great measure from the similarities of the two languages involved in the machine translation process. In fact, the differences between Spanish and Catalan languages are rather minor and have in most cases only phonological nature and more rarely morphological or grammatical.

The results obtained during the development of Spanish-Catalan machine translation system can be obviously applied to any system, which translates between closely related languages. Probably acting this way an effective machine translation systems for unrestricted text can be built for such pair of languages like Swedish and Norwegian, Norwegian and Danish, Swedish and Danish, Spanish and Portuguese, German and Dutch, or Finnish and Estonian. But it is a bit doubtful if acting this way a high-quality machine translation system for unrestricted text can be built, which would be able to translate between a pair of totally typologically different and genetically unrelated languages like, for example, Chinese and French.

5. Example-based machine translation for the Polish language

The main idea of example-based machine translation technique is that, if we have a bilingual text in which some phrase was translated in a certain way, there exist a high probability, but not certainty, that using this translation of the above mentioned phrase in some other text we will obtain the correct translation again. The example-based machine translation technique was proved to be very useful for such language pairs as: English-Spanish and English-French [7], [6], [5], but it has never been implemented for the Polish language.

The Polish language, which belongs to the group of the Slavonic languages, differs very much in its grammar from the West-European languages. This is a reason why a direct implementation of the example-based machine translation technique for the Polish language is not so easy and probably wouldn't bring the desired effects. In order to use the example-based machine translation for the system, which translates from West-European languages into Polish, the example-based translation technique must be modified a bit. The system proposed by this author is based on the following observations:

- Perhaps, in every human language we can distinguish the first, the second, and the third grammatical person.
- Perhaps, in every human language we can distinguish such elements of the sentence as: a subject group S, a verb group V, and an object group O. In the majority of the Indo-European languages the most common word order in the sentence is SVO (subject-verb-object).
- In the Polish language the grammatical person, number, and gender of the verb group must agree with these of the subject group. Further, the grammatical case of the object group must agree with the one required by the verb group.

Taking into account the abovementioned observations the architecture of the propose example-based machine translation system is the following. The system is based on the database in which the translation examples are stored. The database records can have different attributes, such as: <case>, <number>, <person>, and <gender>. During the process of translation the values of these attributes are set respectively to the grammatical rules of the Polish language, so as the subject and verb agreed in the grammatical person, number, and gender. Also the grammatical case of the object must agree with the one required by the verb. There exist three types of translation examples:

1. Noun group translation examples. These translation examples play the role of the subject or the object of the sentence, which is to be translated. The database record for a noun group translation example has the following form:

NG_source	<case>	
NG_target_1	1	
NG_target_2	2	
NG_target_3	3	
NG_target_4	4	
NG_target_5	5	
NG_target_6	6	
<person>	<number>	<gender>
PERSON	NUMBER	GENDER

- NG_source - a noun group of the source language that is to be translated into the target language (Polish) in a manner depending on the value of the attribute <case>
 - NG_target_1 - translation of the source language noun group when <case> = 1
 - NG_target_2 - translation of the source language noun group when <case> = 2
 - NG_target_3 - translation of the source language noun group when <case> = 3
 - NG_target_4 - translation of the source language noun group when <case> = 4
 - NG_target_5 - translation of the source language noun group when <case> = 5
 - NG_target_6 - translation of the source language noun group when <case> = 6
 - PERSON - the value to which the attribute <person> is set, so as the subject and verb groups agreed in their grammatical person
 - NUMBER - the value to which the attribute <number> is set, so as the subject and verb groups agreed in their grammatical number
 - GENDER - the value to which the attribute <gender> is set, so as the subject and verb groups agreed in their grammatical gender
3. Verb group translation examples. These translation examples play the role of the verb of the sentence, which is to be translated. The data base record for verb group translation examples has the following form:

VG_source	<person>	<number>	<gender>
VG_target_1	1	1	1
VG_target_2	1	1	2
VG_target_3	1	2	1
VG_target_4	1	2	2
VG_target_5	2	1	1
VG_target_6	2	1	2
VG_target_7	2	2	1
VG_target_8	2	2	2
VG_target_9	3	1	1

VG_source	<person>	<number>	<gender>
VG_target_10	3	1	2
VG_target_11	3	1	3
VG_target_12	3	2	1
VG_target_13	3	2	2
VG_target_14	3	2	3
<case>			
CASE			

- VG_source - verb group of the source language that is to be translated into the target language (Polish) in a manner depending on the value of the attributes <person>, <number>, and <gender>
- VG_target_1 - translation of the source language noun group when <person> = 1, <number> = 1, and <gender> = 1
- VG_target_2 - translation of the source language noun group when <person> = 1, <number> = 1, and <gender> = 2
- VG_target_3 - translation of the source language noun group when <person> = 1, <number> = 2, and <gender> = 1
- VG_target_4 - translation of the source language noun group when <person> = 1, <number> = 2, and <gender> = 2
- VG_target_5 - translation of the source language noun group when <person> = 2, <number> = 1, and <gender> = 1
- VG_target_6 - translation of the source language noun group when <person> = 2, <number> = 1, and <gender> = 2
- VG_target_7 - translation of the source language noun group when <person> = 2, <number> = 2, and <gender> = 1
- VG_target_8 - translation of the source language noun group when <person> = 2, <number> = 2, and <gender> = 2
- VG_target_9 - translation of the source language noun group when <person> = 3, <number> = 1, and <gender> = 1
- VG_target_10 - translation of the source language noun group when <person> = 3, <number> = 1, and <gender> = 2
- VG_target_11 - translation of the source language noun group when <person> = 3, <number> = 1, and <gender> = 3
- VG_target_12 - translation of the source language noun group when <person> = 3, <number> = 2, and <gender> = 1
- VG_target_13 - translation of the source language noun group when <person> = 3, <number> = 2, and <gender> = 2

- VG_target_14 - translation of the source language noun group when <person> = 3, <number> = 2, and <gender> = 3
 - CASE - the value to which the attribute <case> is set, so as the grammatical case of the object agreed with the one required by the verb group
3. Other translation examples. These translation examples are the same as in the classical technique of example based machine translation. In the sentence they can play the role of particles, junctions, etc. The data base record for this kind of translation examples has the following form:

EX_source
EX_target

- EX_source - a phrase of the source language that is to be translated into the target language (Polish)
- EX_target - translation of the source language phrase

6. The Spanish-to-Polish example-based machine translation system

The proposed machine translation technique was implemented by this author for the system, which translates from Spanish into Polish. The Spanish language belongs to the group of Romanic languages, and it differs much from Polish, which is a Slavonic language. What is important is that both languages belong to the Indo-European family of languages that implies their grammatical structures to be similar enough, so that the example-based machine translation could be used.

This author developed a database of translation examples, according with the proposed by himself methodology, which allowed to translate simple texts from Spanish into Polish. The manner in which the system operates is illustrated on the following example. The purpose of the proposed system is to translate into Polish the following Spanish text composed of a few simple sentences:

Las ciudades españolas son grandes y modernas. Pero en estos viejos pueblos las casas son de tipo tradicional. Esta clase de pueblo tiene en el centro una plaza grande. Los edificios municipales y comerciales son bajos y blancos. Las casas típicas de un pueblo andaluz tienen balcones de piedra igual que las calles. Muchos de los habitantes son labradores. Las viejas tradiciones de los labradores parecen a ser una combinacion de influencias católicas árabes y judías.

First, at the beginning of the sentence the value of attribute <case> is set to 1, because the subject in the Polish sentence is always in a nominative case.

<case> =1;

The translation examples are taken from the database in the order of their occurrence in the translated sentence:

1) noun group translation example

las ciudades españolas	<case>	
miasta hiszpańskie	1	
miast hiszpańskich	2	
miastom hiszpańskim	3	
miasta hiszpańskie	4	
miastami hiszpańskimi	5	
miastach hiszpańskich	6	
<person>	<number>	<gender>
3	2	3

2) verb group translation example

son grandes y modernas	<person>	<number>	<gender>
-	1	1	1
-	1	1	2
-	1	2	1
-	1	2	2
-	2	1	1
-	2	1	2
-	2	2	1
-	2	2	2
jest duży i ładny	3	1	1
jest duża i ładna	3	1	2
jest duże i ładne	3	1	3
są duzi i ładni	3	2	1
są duże i ładne	3	2	2
są duże i ładne	3	2	3
<case>			
-			

At the beginning of a new sentence the attribute <case> is set to 1.

<case> = 1;

3) translation example

pero
ale

4) translation example

en estos viejos pueblos
w tych starych miastach

5) noun group translation example

las casas	<case>	
domy	1	
domów	2	
domom	3	
domy	4	
domami	5	
domach	6	
<person>	<number>	<gender>
3	2	3

6) verb group translation example

son de tipo tradicional	<person>	<number>	<gender>
-	1	1	1
-	1	1	2
-	1	2	1
-	1	2	2
-	2	1	1
-	2	1	2
-	2	2	1
-	2	2	2
jest typu tradycyjnego	3	1	1
jest typu tradycyjnego	3	1	2
jest typu tradycyjnego	3	1	3
są typu tradycyjnego	3	2	1
są typu tradycyjnego	3	2	2
są typu tradycyjnego	3	2	3
<case>			
-			

At the beginning of a new sentence the attribute <case> is set to 1.

<case> = 1;

7) noun group translation example

esta clase	<case>	
ten rodzaj	1	
tego rodzaju	2	
temu rodzajowi	3	
ten rodzaj	4	
tym rodzajem	5	
tym rodzaju	6	
<person>	<number>	<gender>
3	2	3

8) translation example

de pueblo
miasta

9) verb group translation example

son de tipo tradicional	<person>	<number>	<gender>
-	1	1	1
-	1	1	2
-	1	2	1
-	1	2	2
-	2	1	1
-	2	1	2
-	2	2	1
-	2	2	2
ma	3	1	1
ma	3	1	2
ma	3	1	3
maja	3	2	1
maja	3	2	2
maja	3	2	3
<case>			
-			

10) translation example

en el centro
w centrum

11) noun group translation example

una plaza grande	<case>	
duży rynek	1	
dużego rynku	2	
dużemu rynkowi	3	
duży rynek	4	
dużym rynkiem	5	
dużym rynku	6	
<person>	<number>	<gender>
3	1	1

At the beginning of a new sentence the attribute <case> is set to 1.

<case> = 1;

12) noun group translation example

los edificios municipals y comerciales	<case>	
budynki miejskie i handlowe	1	
budynków miejskich i handlowych	2	
budynkom miejskim i handlowym	3	
budynki miejskie i handlowe	4	
budynkami miejskimi i handlowymi	5	
budynkami miejskimi i handlowymi budynkach miejskich i handlowych	6	
<person>	<number>	<gender>
3	2	3

13) verb group translation example

on bajos y blankos	<person>	<number>	<gender>
-	1	1	1
-	1	1	2
-	1	2	1
-	1	2	2
-	2	1	1
-	2	1	2
-	2	2	1
-	2	2	2
jest niski i biały	3	1	1
jest niski i biały	3	1	2
jest niski i biały	3	1	3
są niskie i białe	3	2	1
są niskie i białe	3	2	2
są niskie i białe	3	2	3
<case>			
-			

At the beginning of a new sentence the attribute <case> is set to 1.

<case> = 1;

14) noun group translation example

las casas típicas	<case>		
typowe domy	1		
typowych domów	2		
typowym domom	3		
typowe domy	4		
typowymi domami	5		
typowych domach miejskich i handlowych	6		
<person>	<number>	<gender>	
3	2	3	

15) translation example

de un pueblo andaluz
andaluzijskiego miasta

16) verb group translation example

tiene	<person>	<number>	<gender>
-	1	1	1
-	1	1	2
-	1	2	1
-	1	2	2
-	2	1	1
-	2	1	2
-	2	2	1
-	2	2	2
ma	3	1	1
ma	3	1	2
ma	3	1	3
maja	3	2	1
maja	3	2	2
maja	3	2	3
<case>			
-			

17) noun group translation example

balcones de piedra		<case>
balkony z kamienia		1
balkonów z kamienia		2
balkonom z kamienia		3
balkony z kamienia		4
balkonami z kamienia		5
balkonach z kamieniah		6
<person>	<number>	<gender>
3	1	3

18) translation example

igual que las calles
tak samo jak ulice

At the beginning of a new sentence the attribute <case> is set to 1.

<case> = 1;

19) noun group translation example

muchos de los habitantes		<case>
wiele mieszkańców		1
wiele mieszkańców		2
wiele mieszkańcom		3
wiele mieszkańców		4
wieloma mieszkańcami		5
wielu mieszkańcach		6
<person>	<number>	<gender>
3	1	3

20) verb group translation example

son	<person>	<number>	<gender>
-	1	1	1
-	1	1	2
-	1	2	1
-	1	2	2
-	2	1	1
-	2	1	2
-	2	2	1
-	2	2	2
jest	3	1	1
jest	3	1	2
jest	3	1	3
są	3	2	1
są	3	2	2
są	3	2	3
<case>			
5			

21) noun group translation example

labradores		<case>
rolnicy		1
rolników		2
rolnikom		3
rolnikom		4
rolnikami		5
rolnikach		6
<person>	<number>	<gender>
3	1	1

At the beginning of a new sentence the attribute <case> is set to 1.

<case> = 1;

22) noun group translation example

las viejas tradiciones		<case>
stare tradycje		1
starych tradycji		2
starym tradycjom		3
stare tradycje		4
starymi tradycjami		5
starych tradycjach		6
<person>	<number>	<gender>
3	1	1

23) translation example

de los labradores
rolników

24) verb group translation example

parecen a ser	<person>	<number>	<gender>
-	1	1	1
-	1	1	2
-	1	2	1
-	1	2	2
-	2	1	1
-	2	1	2
-	2	2	1
-	2	2	2
wydaje się być	3	1	1
wydaje się być	3	1	2
wydaje się być	3	1	3
wydają się być	3	2	1
wydają się być	3	2	2
wydają się być	3	2	3
<case>			
5			

25) noun group translation example

una combinacion	<case>	
kombinacja	1	
kombinacji	2	
kombinacji	3	
kombinacje	4	
kombinacją	5	
kombinacji	6	
<person>	<number>	<gender>
3	1	2

26) translation example

de influencias católicas árabes y judías
wpływów katolickich arabskich i żydowskich

The effect of the work of the proposed example-based machine translation system is Polish translation of the original Spanish text:

Miasta hiszpańskie są duże i ładne. Ale w tych starych miastach domy są typu tradycyjnego. Ten rodzaj miasta ma w centrum duży rynek. Budynki miejskie i handlowe są niskie i białe. Typowe domy andaluzyjskiego miasta mają balkony z kamienia tak samo jak ulice. Wiele mieszkańców jest rolnikami. Stare tradycje rolników wydają się być kombinacją wpływów katolickich arabskich i żydowskich.

It must be stressed that the Polish translation is both correct from the grammatical point of view and it is an exact translation of the original Spanish text. Moreover, the obtained Polish text seems to be natural, and thus very similar to the one produced by a human translator. These facts point out that the modified example-based machine translation technique proposed by this author is headed in the right direction.

Some further comments must be made about, why for some Spanish phrases that are in a plural form also Polish translation in a singular form are proposed. It is a bit strange fact but it sometimes occurs that during the process of translation singular form is changed into a plural one and vice versa. For example Polish noun "drzwi" is plural, but its Spanish equivalent "la puerta" is singular. Also Spanish word "la gente" is singular, but its Polish translation "ludzie" is plural, etc. So, this is the reason why in the above machine translation system the possible translations were taken into account both for singular and plural.

7. Final remarks

The high-quality machine translation system for unrestricted text has always been an unachievable goal for the computer scientists working in the field of automatic translation between human languages. And maybe, because of the reasons of fundamental nature (the lack of possibility of constructing an algorithm equivalent to the creativeness of the human mind) human translators will never be eliminated by computers totally, and high-quality machine translation for unrestricted text will forever remain the Holy Grail of scientific research [15]. But, by using various machine translation techniques we can of course try to approach as close as possible to this unattainable goal [11]. Quite recently the example-based machine translation technique has emerged as a very serious and tempting alternative to the existing systems that are mainly based on the knowledge developed in the field of computational linguistics.

In the paper the implementation of example-based machine translation technique in the system, which translates from Spanish into Polish is proposed. In order to use the example-based machine translation technique for the Polish language, which possesses very specific grammatical features, so different from the West-European languages, this author proposed a thorough modification of this technique that allows to take into account the flexion nature of the Polish language.

The results obtained so far are very promising and show that the usage of example-based machine translation technique for Polish and Spanish language pair is a step made in the right direction. But, we cannot forget that the final success depends strongly on the dimension of the database of translation examples. The effective constructing of such database requires a lot of work and time. In fact, it is a task for a quite big team of trained linguists and computer scientist, who basing only on the great bilingual corpus would be able to extract all necessary and most frequently used translation examples.

Last but not least, this author would like to mention that according with its knowledge the proposed Spanish-to-Polish machine translation system is the first system of this kind that has ever been built, thus the obtained by this author results have a totally pioneer character.

References

- [1] D. ARNOLD, L. BALKAN, S. MEIJER, R. L. HUMPHREYS and L. SADLER: *Machine Translation: An Introductory Guide*. NCC Blackwell, London, 1994.
- [2] K.L. BAKER, A.M. FRANZ, P.W. JORDAN, T. MITAMURA and E.H. NYBERG: *Coping with Ambiguity in a Large-Scale Machine Translation System*. Center for Machine Translation, Carnegie Mellon University, Pittsburgh, USA, 1998.
- [3] S. BANDYOPADHYAY: State and role of machine translation in India. *Machine Translation Review*, **11** (2000), 1-3.

-
- [4] M. BLEKHMA and B. PEVZNER: First Steps of Language Engineering in The USSR: The 50s Through 70s. *Machine Translation Review*, **11** (2000), 5-7.
- [5] R.D. BROWN: Adding linguistic knowledge to a lexical example-based translation system. Carnegie Mellon University, Pittsburgh, USA, 2000.
- [6] R.D. BROWN: Example-based machine translation in the Pangloss system. Carnegie Mellon University, Pittsburgh, USA, 2000.
- [7] R.D. BROWN: Transfer-rule induction for example-based translation. Language Technologies Institute, Carnegie Mellon University, Pittsburgh, USA, 2000.
- [8] R. CANALS, A. ESTEVE, A. GARRIDO, M.I. GUARDIOLA, A. ITURRASPELLVER and S. MONTSERRAT: InterNOSTRUM: A Spanish-Catalan Machine Translation System. *Machine Translation Review*, **11** (2000), 21-25.
- [9] J.G. CARBONELL, T. MITAMURA and E.H. NYBERG: The KANT Perspective: A Critique of Pure Transfer (and Pure Interlingua, Pure Statistics. Center for Machine Translation, Carnegie Mellon University, Pittsburgh, USA, 1998.
- [10] O. FUKUTOMI: Report on Commercial Machine Translation in a Manufacturing Domain. *Machine Translation Review*, **11** (2000), 16-25.
- [11] N.V. LOUKACHEVITCH and B.V. DOBROV: Thesaurus-Based Structural Thematic Summary in Multilingual Information Systems. *Machine Translation Review*, **11** (2000), 10-20.
- [12] A.F. MAJEWICZ: The languages of the world and their classifying. PWN, Warsaw, Poland, 1989, (in Polish).
- [13] A. MELBY: Machine Translation and Philosophy of Language. *Machine Translation Review*, **9** (1999), 6-17.
- [14] T. MITAMURA: Controlled Languages for Multilingual Machine Translation. Center for Machine Translation, Carnegie Mellon University, Pittsburgh, USA.
- [15] T. MITAMURA, E.H. NYBERG and J. CARBONELL: An Efficient Interlingua Translation System for Multi-Lingual Document Production. Center for Machine Translation, Carnegie Mellon University, Pittsburgh, USA, 1998.
- [16] D. MURPHY: Keeping Translation Technology under Control. *Machine Translation Review* **11** (2000), 7-10.
- [17] H. NEY, S. NIESSEN, F.J. OCH, H. SAWAF, C. TILLMANN and S. VOGEL: Algorithms for Statistical Translation of Spoken Language. *IEEE Trans. on Speech and Audio Processing*, **8**(1), (2000), 24-36.

- [18] E. NYBERG, T. MITAMURA and J. CARBONELL: The KANT Machine Translation System: From R&D to Initial Deployment. Center for Machine Translation, Carnegie Mellon University, Pittsburgh, USA, 1999.
- [19] R. PENROSE: *New Caesar's Mind*. PWN, Warsaw, Poland, 1995, (in Polish).
- [20] C. RICO: From Novelty to Ubiquity: Computers and Translation at the Close of Industrial Age. <http://www accurapid.com/journal/15mt2.htm>
- [21] A. WAIBEL, P. GEUTNER, L.M. TOMOKIYO, T. SCHULTZ and M. WOSZCZYNA: Multilinguality in Speech and Spoken Language Systems. *Proc. IEEE*, **88**(8), (2000), 1297-1313.
- [22] P. WHITELOCK and K. KILBY: *Linguistic and Computational Techniques in Machine Translation System Design*. UCL Press, London, UK, 1995.
- [23] V.W. ZUE and J.R. GLASS: Conversational Interfaces: Advances and Challenges. *Proc. IEEE*, **88**(8), (2000), 1166-1180.