The European Proceedings of
Social & Behavioural Sciences
EpSBS

Future Academy

ISSN: 2357-1330

http://dx.doi.org/10.15405/epsbs.2017.12.29

ICPE 2017
International Conference on Psychology and Education

INTERLINGUA-BASED NUMERAL TRANSLATION IN WEB-APPLICATION WITH KNOWLEDGE-TESTING

Alexander Prutzkow (a)*
*Corresponding author

(a) Department of Computational and Applied Mathematics, Ryazan State Radioengineering University, 390005, Gagarina str., 59/1, Ryazan, Russia, mail@prutzkow.com

Abstract

The purpose of this research is developing an Interlingua-based technique of the natural language numeral processing and translation. We propose a three-level generalized numeral model as Interlingua-representation. The model formal grammar describes the natural language numeral structure. The first level grammar rules define that a numeral consists of sign, integer part, separation symbol, and fractional part. The second level decribes numeral integer part as a triad sequence. The third level defines the triad structure. We developed number-into-numeral, numeral-into-number, and translating algorithms based on the model. The algorithms are implemented in the Markov normal algorithms. We realized the model and the algorithms in web-application in the Internet. The web-application has a knowledge-testing function. The function allows to users test numeral convering and translating knowledge. Users from more than 100 countries visit web-application and convert numerals. The largest number of users resides in the US and the Russian Federation. The web-application log contains more than 200,000 records. The largest number of user requests related to the conversion of cardinal numerals of Spanish. The web-application is integrated in toolbox of a complex linguistic web-portal for translators as well. We conclude the Interlgua-based technique is effective for numeral processing and translation and realization in web-applications.

Keywords: Natural Language Processing, Numeral Translation, Markov Normal Algorithms, Linguistic Web-Application

http://dx.doi.org/10.15405/epsbs.2017.12.29
Corresponding Author: Alexander Prutzkow
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

1. Introduction

We use the following terms in this paper. A numeral is a cardinal numeral having symbolic notation in the text, e.g. «three hundred fifty two». A number is a cardinal numeral having digital notation, e.g. «352».

In the process of text translation program application also converts numerals of the source language into numerals of the target language. However, the numeral translation rules aren't the same to the language-into-language text translation rules. Also there are number-into-numeral and numeral-into-number converting tasks in the text processing.

In this paper we describe how to process numbers and numerals in the text using the Interlingua representation. We present a web-application for numeral converting and translation, user request statistics. The web-application can be used in the natural language learning as a knowledge-testing system.

2. Problem Statement

Machine translation is used by many users especially in the Internet. Popular web-translators have many language directions of text translation and demonstrate quality result.

There are two basic text translating technique:

1) translation with rules and small bank of translating equivalents

2) memory translation (or translation memory) (Planas & Fruse, 1999; Dillon & Fraser, 2006) with only huge bank of translating equivalents.

However, no one of techniques can't get the text meaning and translates numerals perfect. The process of numeral translation has differences from language to language and isn't similar to text translation. So it is necessary additional tools for numeral translation.

We try to solve a problem of numeral translation using the Interlingua-based technique.

2.1. Interlingua-based translation

The Interlingua is an intermediate representation of translating text (Dorr, Hovy & Levin, 2004; Lampert, 2004; Lee & Seneff, 2005). There are two steps in the Interlingua-based translation: 1) converting the source language text into the Interlingua; 2) converting the Interlingua into the target language text. The translation is easy-extending. To add new language in multi-lingual translating system you need to develop the language-Interlingua converting algorithms.

We use such Interlingua-based translation to process numerals in the text. We describe the Interlingua representation with the formal grammar.

2.2. G. Hardegree grammars

Gary Hardegree proposed grammars (Hardegree, 1999) for representation of the numeral structure and number transformation into English numerals. The grammars can be intended only for numeral building and used only in English. The grammars also describe the rules of building only for integer parts

http://dx.doi.org/10.15405/epsbs.2017.12.29
Corresponding Author: Alexander Prutzkow
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

of numerals, but don't consider numeral case inflection as there is no case grammatical category in English.

3. Research Questions

The following research questions guide the current study:

Question 1: Is the Interlingua-based technique effective for the numeral processing and translating?

Question 2: What structure has the Interlingua representation for the numeral processing and translating?

Question 3: How realize the Interlingua-based numeral processing and translating in application accessible for users in countries of the World?

4. Purpose of the Study

The purpose of this study is to describe an Interlingua representation for the numeral processing and translating using the formal grammar and develop a web-application realized this representation.

5. Research Methods

Our research has the following stages:

- To develop the Interlingua representation and descibe it with the formal grammar.
- To develop algorithms of numeral converting and translation.
- To develop a web-application based on the Interlingua representation and the algorithms.

6. Findings

In the result of our research we have got the following findings.

6.1. Three-level generalized numeral model

To describe a generalized numeral structure for the Interlingua representation by the formal grammar, we use the following terms.

Number terms are:

- digits – $Z = \{Z_0|0, Z_1|1, Z_2|2, \ldots, Z_9|9\}$;
- negative – S_-;
- integer and fractional parts separation symbol – J.

The numeral terms are:

- digit names (simple atomic numerals) – $C = \{C_0, C_1, C_2, \ldots, C_9\}$;
- the tens and hundreds names – $D = \{D_1, D_2\}$ (D_1 is the tens name; D_2 is the hundreds name);
- triad order names: thousands, millions, billions, etc. – $M = \{M_0, M_1, M_2, M_3, \ldots\}$ (M_i is the name of $i + 1$ numeral triad order.),

http://dx.doi.org/10.15405/epsbs.2017.12.29
Corresponding Author: Alexander Prutzkow
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

- numeral sign name – $P = \{P_-, P_+, P_0\}$ (P_- is a negative sign name; P_+ is a positive sign name; P_0 is the zero mark.);
- integer and fractional parts separation symbol name – E;
- fractional part ending name – B.

Example 1. In this terms, the number

1 000 400 973 = $Z_1\ Z_0Z_0Z_0\ Z_4Z_0Z_0\ Z_9Z_7Z_3$

presents as:

$P_+C_1M_3C_4D_2M_1C_9D_2C_7D_1C_3M_0EB$. □

The terms are necessary for model generalization and language independence.

We analyze the natural languages numeral building rules and develop a three-level generalized numeral model (or *model*) as the Interlingua. It consists of the following levels.

Level 1 renders numeral sign and integer and fractional parts. Part delimiters are patch words «*comma*», «*point*».

Level 2 renders three-digit ingredients (triads). Each part is divided into three-digit ingredients beginning with integer and fractional parts separating char. Three-digit ingredients part delimiters are patch words «*thousand*», «*million*», etc.

Level 3 renders three-digit ingredients items. Part delimiters are patch words «*tens*», «*hundreds*».

Example 2. Figure 01 illustrates the three-level generalized numeral model structure by the example of number 34 567.89. □

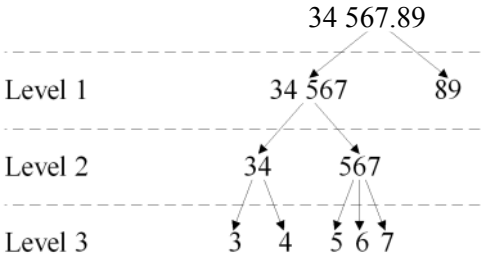

Figure 01. Three levels of number 34 567.89 decomposition

The model grammar rules are shown below.

Level 1:

$K = P + N_1 + E + \{\varnothing|N_2\} + B$

$P = P_-|P_+|P_0$

Level 2:

$N_1 = C_0|N_{10}|N_{11}|N_{12}|...|N_{1i}|...$

$N_2 = (\{T_1|C_0\} + N_2)|T_1|C_0$

$N_{10} = T + M_0$

$N_{11} = T + M_1 + \{\varnothing|N_{10}\}$

$N_{12} = T + M_2 + \{\varnothing|N_{10}|N_{11}\}$

...

$N_{1i} = T + M_i + \{\varnothing|N_{11}|N_{12}|...|N_{1(i-1)}\}$

...

Level 3:

$T = T_1|T_2|T_3$

$T_1 = C_1|C_2|C_3|C_4|C_5|C_6|C_7|C_8|C_9$

http://dx.doi.org/10.15405/epsbs.2017.12.29
Corresponding Author: Alexander Prutzkow
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

$T_2 = T_1 + D_1 + \{\varnothing | T_1\}$

$T_3 = T_1 + D_2 + \{\varnothing | T_1 | T_2\}$

The model is a generalized numeral structure using in programming of number-into-numeral, numeral-into-number, and translating algorithms.

6.2. Converting algorithms

We use Markov normal algorithms (Markov, 1954) to implement converting algorithms.

We add new operations and symbols to simplify the algorithm normal scheme:

$\rightarrow +\gamma$ – to add symbol γ at the end of the word;

$\rightarrow \gamma +$ – to add symbol γ at the beginning of the word;

$[Z_0]_k$ – k zeros sequence;

_ (underlining) – space symbol.

In this section we present some basic converting algorithms.

Number-into-model integer part converting algorithm replaces digits by numeral terms.

1) $Z_k \gamma_1^j \rightarrow \gamma_2^j C_k M_j; j = 0, 1, 2, \ldots; k = 1, 2, \ldots, 9$

2) $Z_k \gamma_2^j \rightarrow \gamma_3^j C_k D_1; j = 0, 1, 2, \ldots; k = 1, 2, \ldots, 9$

3) $Z_k \gamma_3^j \rightarrow \gamma_1^{j+1} C_k D_2; j = 0, 1, 2, \ldots; k = 1, 2, \ldots, 9$

4) $Z_0 Z_0 Z_0 \gamma_1^j \rightarrow \gamma_1^{j+1}; j = 0, 1, 2, \ldots$

5) $Z_0 \gamma_1^j \rightarrow \gamma_2^j M_j; j = 0, 1, 2, \ldots$

6) $Z_0 \gamma_2^j \rightarrow \gamma_3^j; j = 0, 1, 2, \ldots$

7) $Z_0 \gamma_3^j \rightarrow \gamma_1^{j+1}; j = 0, 1, 2, \ldots$

8) $\gamma_2^0 M_0 E \rightarrow P_0 C_0 E B \bullet$

9) $S_- \gamma_i^j C_k \rightarrow P_- C_k \bullet; i = 1, 2, 3; j = 0, 1, 2, \ldots; \quad k = 0, 2, \ldots, 9$

10) $\gamma_i^j C_k \rightarrow P_+ C_k \bullet; i = 1, 2, 3; j = 0, 1, 2, \ldots; k = 0, 2, \ldots, 9$

11) $J \rightarrow \gamma_1^0 E$

12) $\rightarrow + \gamma_1^0 E B$

An index symbol γ marks processed symbol in the number.

Model-into-number integer part converting algorithm transforms a numeral integer part into a number integer part.

1) $M_j \gamma_i^L \rightarrow \gamma_1^j [Z_0]_{(4-i)+3\cdot(j-L-1)}; i = 1, 2, 3;$
$L = 0, 1, 2, \ldots; j = L+1, L+2, \ldots$

2) $M_j \gamma_1^j \rightarrow \gamma_1^j; j = 0, 1, 2, \ldots$

3) $C_k \gamma_1^j \rightarrow \gamma_2^j Z_k; j = 0, 1, 2, \ldots; k = 1, 2, \ldots, 9$

4) $C_k D_1 \gamma_i^j \rightarrow \gamma_3^j Z_k [Z_0]_{2-i}; i = 1, 2; j = 0, 1, 2, \ldots; k = 1, 2, \ldots, 9$

5) $C_k D_2 \gamma_i^j \rightarrow \gamma_1^{j+1} Z_k [Z_0]_{3-i}; i = 1, 2, 3; j = 0, 1, 2, \ldots; k = 1, 2, \ldots, 9$

6) $P_+ \gamma_i^j Z_k \rightarrow Z_k \bullet; i = 1, 2, 3; j = 0, 1, 2, \ldots; k = 1, 2, \ldots, 9$

7) $P_- \gamma_i^j Z_k \rightarrow S_- Z_k \bullet; i = 1, 2, 3; j = 0, 1, 2, \ldots; k = 1, 2, \ldots, 9$

8) $P_+ \gamma_i^j Z_0 \rightarrow Z_1 Z_0 \bullet; i = 2, 3; j = 0, 1, 2, \ldots$

http://dx.doi.org/10.15405/epsbs.2017.12.29
Corresponding Author: Alexander Prutzkow
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

9) $P\_\gamma_1^j Z_0 \rightarrow S\_Z_1 Z_0 \bullet$; $i = 2, 3$; $j = 0, 1, 2, \ldots$

10) $P_0 C_0 EB \rightarrow Z_0 \bullet$

11) $EB \rightarrow \gamma_1^0$

12) $E \rightarrow \gamma_1^0 J$

The algorithm also use the index symbol γ.

The algorithm using the model checks numeral for errors as well. 6-9 replacements execution is a correct algorithm ending. None of the replacements execution means that a numeral contains an error.

Number-into-model fractional part converting algorithm is shown below.

1) $\lambda Z_k \rightarrow C_k \lambda$; $k = 0, 1, \ldots, 9$

2) $\lambda \rightarrow B \bullet$

3) $E \rightarrow E\lambda$

In this algorithm the index symbol is a λ Greek alphabet letter.

Model-into-number fractional part converting algorithm replaces numeral terms while all terms will processed.

1) $\lambda C_k \rightarrow Z_k \lambda$; $k = 0, 1, \ldots, 9$

2) $\lambda C_k B \rightarrow Z_k \bullet$; $k = 0, 1, \ldots, 9$

3) $EB \rightarrow EB \bullet$

4) $E \rightarrow E\lambda$

6.3. Numeral translation order

The model terms have general nature. In each language, they have unique types. For example, in the Russian language C_0 = «*ноль*», C_1 = «*один*», etc.

In some languages, numerals are formed by the rules that are different from the model rules. In this case, numeral is converted by algorithms. Two algorithms are necessary:

1) numeral-into-model converting algorithm transforming a numeral in the model representation into a numeral of the target language;

2) model-into-numeral converting algorithm transforming a numeral of the source language into a model numeral.

These algorithms carry out the opposite actions.

Number-into-numeral converting includes four steps:

1) to execute the number-into-model integer part converting algorithm;

2) to execute the number-into-model fractional part converting algorithm;

3) to execute the model-into-numeral converting algorithm if it exists;

4) to replace model terms by language symbols in required case.

Numeral-into-number converting consists of the following steps:

1) to replace language symbols by model terms;

2) to execute the numeral-into-model converting algorithm if it exists;

3) to execute the model-into-number fractional part converting algorithm;

4) to execute the model-into-number integer part converting algorithm.

Using the model, we can translate numerals. Translation of a numeral from the L1-language into the L2-language includes four steps:

1) to replace the L1-language symbols by the model terms;

2) to execute numeral-into-model converting algorithm if it exists for the L1-language;

3) to execute model-into-numeral converting algorithm if it exists for the L2-language;

4) to replace the model terms by the L2-language symbols in required case.

http://dx.doi.org/10.15405/epsbs.2017.12.29
Corresponding Author: Alexander Prutzkow
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

All algorithms in this paper are implemented in web-application.

6.4. Web-application with knowledge-testing function

In 2012 we have developed (with Dmitriy Tsybulko) a web-application for processing of the natural language cardinal numerals. The web-application is available to users of the Internet at http://prutzkow.com/en-us/numbers/ (the English-language version) or http://prutzkow.com/ru-ru/numbers/ (the Russian-language version).

The web-application has the following functions:

1) translation of numerals of Russian, English, German, Spanish and Finnish languages in any direction (because of the Interlingua-based technique);

2) number-into-numeral and numeral-into-number converting;

3) declination of numerals of the Russian language;

4) numerals converting and translating knowledge testing.

The web-application was integrated in the toolkit of the My-Polyglot.com expert network for translators.

6.5. User request statistics

Each request to the web-application is recorded in the log. The log uses for web-application debugging and analyzing of request statistics. A record of the log includes the following request data:

- date and time;
- user IP address;
- query string;
- translating direction;
- type of the query string (number or numeral, the natural language is also analyzed for numerals);
- result of translation;
- UserAgent string sending by the web-browser.

There are more than 200,000 records in the log in this moment.

We analyzed the log and present numeral translating direction statistics in Table 01.

Table 01. Percentage of numeral processing requests by translating directions and by years

The target language / The source language	Percentage of numeral processing requests in each year						
	Russian	**English**	**Spanis**	**German**	**Finnish**	**Number**	**Total**
Russian	< 1	< 1	< 1	< 1	< 1	< 1	1.2
	< 1	< 1	< 1	< 1	< 1	< 1	0.7
	< 1	< 1	< 1	< 1	< 1	< 1	1.1
English	< 1	< 1	< 1	< 1	< 1	< 1	0.7
	< 1	< 1	< 1	< 1	< 1	< 1	0.6
	< 1	< 1	< 1	< 1	< 1	< 1	0.5
Spanish	< 1	< 1	< 1	< 1	< 1	< 1	1.1
	< 1	< 1	< 1	< 1	< 1	< 1	1.4
	< 1	1.3	< 1	< 1	< 1	< 1	1.8
German	< 1	< 1	< 1	< 1	< 1	< 1	0.4
	< 1	< 1	< 1	< 1	< 1	< 1	0.1
	< 1	< 1	< 1	< 1	< 1	< 1	0.1

http://dx.doi.org/10.15405/epsbs.2017.12.29
Corresponding Author: Alexander Prutzkow
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

Finnish	< 1 < 1 < 1	< 1 < 1 < 1	< 1 < 1 < 1	< 1 < 1 < 1	< 1 < 1 < 1	< 1 < 1 < 1	< 0.01 < 0.01 < 0.01
Number	15.7 8.4 15.5	6.9 4.8 6.6	58.9 76.9 63.7	13.7 6.0 9.4	< 1 < 1 < 1	< 1 < 1 < 1	96.6 97.2 96.5
Total	16.7 8.8 16.1	7.8 6.0 8.0	59.5 77.6 64.3	14.1 6.3 10.1	0.7 0.5 0.6	1.1 0.9 0.9	100.0 100.0 100.0

In Table 01, each cell contains three values. The upper value corresponds to 2014, the middle value corresponds to 2015, and the lower value corresponds to 2016.

The most of users of the web-application reside in the US and Russia (Table 02). The trend has not changed for the last three years. Users of the web-application live in more than 100 countries on all permanently inhabited continents. The data for Table 02 includes all requests, even erroneous.

Table 02. Percentage of the web-application users by country and by year

Countries of the World	2014	2015	2016
USA	53.7	68.9	52.2
Russia	26.6	14.3	21.9
Ukraine	4.8	2.6	3.3
UK	1.3	1.8	2.6
Canada	1.5	1.3	1.5
India	0.8	1.0	1.5
Mexico	0.3	0.6	1.4
Belarus	1.9	0.8	1.1
Other countries	9.1	8.7	14.5

7. Conclusion

In the result of this research, we have got the following answers to research questions:

1. The Interlingua-based technique is effective for the numeral processing and translating. The technique allows adding numerals of the new natural language easy writing two converting algorithms: numeral into the Interlgua and the Interlingua into numeral.

2. The Interlingua representation for the numeral processing and translating has a numeral-like structure. We have developed the three-level generalized numeral model describing the Interlingua representation structure.

3. To make practical results of our research accessible for users in countries of the World we have developed the web-application for numeral processing and translation with knowledge-testing function. The web-application is based on the model as the Interlingua representation.We have received many good comments from users of our web-application. The web-application was integrated in a complex linguistic web-portal for translators.

We have published the result of this research in Russian scientific journals as well.

http://dx.doi.org/10.15405/epsbs.2017.12.29
Corresponding Author: Alexander Prutzkow
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

Acknowledgments

We are grateful to our co-author Dmitriy Tsybulko for collaboration in this research.

References

Dillon, S., Fraser, J. (2006). Translators and TM: An Investigation of Translators' Perceptions of Translation Memory Adoption. *Machine Translation*, 20 (2), pp 67-79.

Dorr, B. J., Hovy, E., Levin, L. (2004). Machine Translation: Interlingual Methods, Encyclopedia of Language and Linguistics. 2nd ed., Brown, Keith (ed.).

Hardegree, G. (1999). Symbolic Logic, First Course, 3rd ed. McGraw-Hill.

Lampert, A. (2004). Interlingua in Machine Translation, Technical Report.

Lee, J., Seneff, S. (2005). Interlingua-Based Translation for Language Learning Systems. In Proc. of ASRU, Cancun, Mexico.

Markov, A. A. (1954). Theory of algorithms (in Russian). Akad. Nauk SSSR (English trans. published by the Israel Program for Scientific Translation, Jerusalem, Vol. XLII, 1962).

Planas, E., Furuse, O. (1999). Formalizing Translation Memories. In Proc. of MT Summit VII, Singapore, September 13-17, 1999, pp. 331-339.