Neoslavonic Language
Zonal Language Constructing: Challenge, Experience, Opportunity to the 21st Century

Vojtech Merunka
Czech University of Life Sciences
Czech Technical University in Prague
Prague, Czech Republic
E-mail: vmerunka@gmail.com

Martin Molhanec
Czech Technical University in Prague
Prague, Czech Republic
E-mail: martin.molhanec@gmail.com

Abstract—This paper describes the project of artificial zonal language construction and first experiences with its use. The paper presents the design principles of this language and the sources of it: Old (Church) Slavonic, Interslavic project, and the False Friend of the Slavic project. Finally, the opportunities for the language practical use and first experiences are discussed. This article also proposes a new approach of analysis and more accurate machine translation between fusion free-word-order languages.

Keywords—Neoslavonic zonal constructed language; Slavic languages; machine translation of free-word-order languages

I. INTRODUCTION

Our European civilization is formed by a very diverse mix of peoples, among which are the largest group Slavic nations (around 250 million, which is about 32% of the European population), Romanesque nations (approximately 242 million, which is about 31% of the European population) and Germanic peoples (approximately 212 million, which is about 27% of the European population). These three language groups make up more than 90% of the total population of Europe. It should also be noted that many more millions of speakers are living in other continents, who are (just thanks to their very same languages) closely linked with the European culture and civilization. It makes up about 1/4 of the total population on Earth, of which the Slavic languages are about 400 million.

In the European history, there were always used many local languages, but there was always some attempt for the standardization. In the ancient age it was Greek language, then Latin language, and now it is the dominance of English.

II. ZONAL CONSTRUCTED LANGUAGES

Zonal constructed languages are artificially constructed languages made to facilitate communication between speakers of a certain group of closely related languages.

Because all our Slavic languages derive from that common Proto-Slavic tongue, knowledge of one Slavic language will often allow one to have at least a rough understanding of text written in another Slavic language - but not sufficiently enough to achieve a strong comprehension. [3] This fact has inspired linguists and others over the centuries to attempt to create a universal zonal Slavic language that would be more understandable to all Slavs. [9] Among these include Old Church Slavonic, developed in the 9th century by two Byzantine Greek missionaries and saint co-patrons of the Europe, the brothers Constantine the Philosopher (Cyril) and Methodius of Thessalonica [10], as well as dozens of other projects since today. What they have in common is that they are all based on the assumption that the Slavic languages are similar enough to make such an auxiliary language possible at all.

The oldest known example (except the Old Church Slavonic from 9th century, of course) is Ruski jezik (1665) by the Croatian priest Juraj Križanić, who is often regarded as the first recorded Pan-Slavist. Other notable examples of early Pan-Slavic language projects are Universalis Lingua Slavica by Ján Herkeľ (1826), Uzajemni Pravopis Slavjanski by the Slovene Matija Majar (1865), and the project of a Czech translator and writer Václav (Vyacheslav) Bambas (1861). In the past, the authors were motivated by the belief that all Slavic languages were dialects of only one single Slavic language rather than separate languages. [9]

There are also known similar projects from the Germanic area (Volapük, for example) and Romanesque area (Interlingua, for example). Some scientists also argue that the Sanskrit language has been designed on the zonal principles.

Zonal constructed languages are based on the different concept than fully artificial languages like Esperanto. Zonal constructed languages are comprehensible for people of certain language group without the need of prior learning. This means that zonal constructed language appears to be almost the same as real spoken languages. For example, if people from certain language zone hear zonal language, they understand and feel this artificial language as only some unknown dialect of their native language. This is why, in our opinion, reduction of grammar and dictionary ad absurdum is not the right way. A universal language should primarily have a logical structure - a clear, understandable, but morphologically rich set of rules with minimum exceptions and idioms. Here, there is no space for unnatural structures made from a combination of random consonants and vowels. A universal language should not even have too abbreviated words and unnatural grammatical building blocks. This means that it should for better clarity prefer words in natural,
full and euphonic forms. Here, in our opinion, Esperanto has failed (e.g. the Esperanto system of unnatural pronouns and adverbs mutually differing in only one vowel or consonant without any natural analogy to some real language).

III. NEOSLAVONIC LANGUAGE

Neoslavonic is created as an auxiliary language that appears to be almost the same as real spoken languages. As such, it is ideal for being used for the following purposes:

- To share grammar principle and a common vocabulary with modern spoken Slavic languages in order to build a universal language that Slavic people can understand without any - or very minimal - prior learning.
- To be an easy-learned language for those who want to use it actively. Non-Slavic people can use this language as the door to the "big Slavic world". We believe that knowledge of Neoslavonic enables both Slavic and non-Slavic people greater passive (e.g. receptive) understanding and better learning of the living Slavic languages.
- To be a basic conversation platform for Slavs and non-Slavs alike, without having to learn a specific real Slavic language.
- Neoslavonic continues the tradition of Church Slavonic. It is designed as the modernized and simplified - but still sufficiently compatible - version of Old Church Slavonic.

In this perspective we cannot ignore the existence of the Russian language. Russian is the largest spoken Slavic language. The amount of its users together exceeds the holders of all other Slavic languages. Russian also holds a strong international position. It is the language of the UN, the language of the astronauts, etc.; it is taught in the USA, China, India, Iran and Arabic countries, and is often preferred as the only Slavic language under “Slavic Studies” programs at foreign universities. Also, we must acknowledge past and present performances of Russian science, technology, culture and sport. Moreover, we know that about one half (maybe more) of the total number of Slavic-speaking people use the Russian language. If the Russian language is sufficiently simple and understandable to other Slavs without learning, our project would be unnecessary; but unfortunately it is not. Russian is far from the imaginary linguistic balanced center of Slavic languages. It has a specific alphabet, phonetics, grammar and significant portion of its vocabulary without the universal Slavic validity. Of course, the same holds true in varying instances for all of our modern national Slavic languages (Polish, Serbian, Czech ...).

IV. NEOSLAVONIC CONSTRUCTION PRINCIPLES

Neoslavonic language can be considered as a dialect of Old Church Slavonic which has evolved into a modern Slavic standard. Neoslavonic can serve as a substitute for Church Slavonic. The biggest differences between "Old" and "New" Slavonic are in some grammatical endings, vocalization and the presence of words from modern Slavic languages. It should be noted that the “Old” Slavonic also has its own dialects in particular Churches (e.g. Russian, Ukrainian, Czech, Croatian, Serbian, Bulgarian, etc.) which mutually differ in pronunciation and details of grammar. Neoslavonic is not so far from all of them, as it has been approved in orthodox monasteries in the south-eastern Europe. The main characteristics of Neoslavonic are:

- Neoslavonic is better balanced between western, southern and eastern Slavic languages. Therefore, it may sometimes act surprisingly "Old-Serbian" for Serbian speakers, "Old-Czech" for Czech speakers, "Old-Slovenian" for Slovenian speakers etc.
- Neoslavonic follows phonetic changes which have occurred in the evolution of modern Slavic languages. This is the loss of nasal vowels, for example.
- Neoslavonic replaces reduced vowels ъ and ъ by standard or loosing vowel e and o.
- Neoslavonic uses semivowels l and r as they are known in Sanskrit, Old Church Slavonic, and some modern Slavic languages (Bulgarian, Serbian, Croatian, Slovenian, Czech, Slovak, and Sorbian).
- Neoslavonic has a limited and regularized collection of declension and conjugation patterns. Moreover, it has numerous symmetries within its grammatical rules – i.e. the same or analogous element shares the same rule in all its occurrences. For example, there are three noun genders of seven cases, each of two patterns (soft and hard) plus some additional exceptional patterns and palatalization rules for euphonic reasons. In total, the overall level of irregularity and complexity of Neoslavonic is significantly much lower than of standard languages.
- Pronoun and adjective declination is united into the same only two declension patterns: soft and hard.
- Neoslavonic unites the concept of verb conjugation to only two regular patterns. All verbs - except for only three irregular verbs byti (to be), iti (to go), and jasti (to eat) belong to either a hard or soft pattern.
- There is active-passive voice symmetry between the verb byti (to be) and the verb imeti (to have). These two verbs create a two voice forms of the same ontological concept. The lost-less voice transformation is formally defined.
- There are 6 symmetrical verb tenses in three levels of time: two past, two present and two future tenses. There is defined lost-less clause transformation among these levels.
- There are both adjective and adverb verbal participles. Total number of participle species is four: active present, passive present, active past and passive past. There is defined lost-less clause
transformation from elements of a complex sentence into these participles.

- Neoslavonic can be written in more alphabets and orthographies. The most important orthographies are accented Latin (e.g. člověk / a man), non-accented Latin (e.g. člověk / a man), Cyrillic (e.g. чловěк / a man) and simplified Cyrillic (e.g. чловек / a man). Of course, the pronunciation is the same in all orthographies.

- In addition, there are designed some additional orthographies for Glagolitic, Greek, Hebrew, Arabic, and Persian alphabets. They are proposals of the universal standard of writing Slavic words and names.

Thus, although Neoslavonic is an artificial language and is much more regular than contemporary spoken Slavic languages, it is not reduced to an unnatural level like Esperanto or Slovio (Slavic clone of Esperanto). Neoslavonic contains only natural grammatical forms which must be the same or very similar to those in contemporary Slavic languages. Like natural Slavic languages, because Neoslavonic does not reduce it grammar and phonetics to Esperanto-like style, Slavic speakers using Neoslavonic can make without problems the same errors in conjugation patterns (different vowels or accent, for example) as people of various Slavic nations today normally do when they speak in non-native Slavic language with each other (for example, when a Czech speaks Russian or a Pole speaks Croatian, etc.)

V. SOURCES OF NEOSLAVONIC

There are three main sources of Neoslavonic:

A. Old Church Slavonic Corpus

This is grammar and dictionary of about 6 thousand words of the Old Church Slavonic (OCS) Corpus from the period of 8th-12th century. [11]

The main construction principle from this source is regularization and transformation of the original language.

B. Interslavic Lexicon and the Interslavic Project

The Interslavic Lexicon is a project managed by Steeven Razdikowski in San Diego, USA. This lexicon of modern words (which are not included in the medieval corpus of the OCS) has about 27 thousand words. [5]

Modern Slavic conlang projects Neoslavonic and Slovianski has accumulated today into one collaborative community called Interslavic. Our project members write and read in both language forms, have their discussions on the same Internet forums, share the same dictionary, the same Internet newspaper, as well as several other places. Interslavic is a common denominator for this collaborative community.

The main construction principle of the lexicon is the Voting Machine, the statistical algorithm of looking for a balanced Interslavic words from modern Slavic languages (Russian, Belarusian, Ukrainian, Polish, Czech, Slovak, Slovenian, Croat, Serbian, Macedonian, and Bulgarian). Each language has its own importance and comprehensibility of the other languages expressed by different mathematical coefficient. [8]

C. False Friends

This is only a correction tool for solving problems in situations where the same word has different meanings in different modern Slavic languages. It is a table-based document of shared words having different meanings between languages. The author of this free document is Daniel Bunčić who made this document under GNU Free Documentation License; [4, 12]

If the voting machine finds a word that is in the false friends table, it must be eliminated and the voting machine should be restarted.

VI. LANGUAGE EXAMPLE 1 - HOTEL INFORMATION LEAFLET

This Neoslavonic text (here it is shortened at about 1/3 of the original text) has been made for our Interslavic project in the hotel OASIS at the Dead Sea coast for hotel guests coming from the Slavic countries. This real and successful project was organized with the kind assistance of Ms Viktorija Ajziković from the hotel management and Mr. Andrej Teterevov from Ben Gurion University in Be’er Sheva, Israel.

A. English Text

Dear Guests! Check in time after 15:00. Check out time by 11:00

Breakfast: 07:00-10:30 Dinner: 18:00-20:30

The lobby bar is open daily from 11:00 until 22:30. Lobby bar offers light snacks meals all day with live entertainment in the evening. Free entrance in the external swimming pool hours are 08:00-18:00.

In addition, our hotel has a private beach. The beach hours are 07:30 - 18:00. Hotel transport to the beach goes each 15 minutes from the central entrance.

Sincerely, the hotel management.

B. Neoslavonic Text

Dragi gosti! Do hotela se možete priglasiti ot 15:00. Iz hotela jest tréba se otići do 11:00

Jutrenica: 07:00 - 10:30 Večerja: 18:00 - 20:30

Lobby-Bar je otvoren vsekakor den ot 11:00 do 22:30. Lobby-Bar vam vsekakor dan predlagaje male jedenije. Večerom vas priglaševamo na veseljenje s živo muzikojo. Bezplatni vstup na otkryty hotelovy bazen jest ot 8:00 do 18:00.

Hotel takože imaje svojo privatnu plažu. Hotelova plaža jest otvorena ot 7:30 do 18:00. Na plaži ide hotelovy transport vsekakih 15 minut ot centralnega vhoda.
S uvaženjem, personal hotela.

VII. LANGUAGE EXAMPLE 2 - NAŠE SÉLO (OUR VILLAGE)

Here is the original English text by Mary Russel Mitford, written in the first half of the XIX century. This text is frequently used by conlangers to test the quality of their constructed languages.

A. English Text

Of all situations for a constant residence, that which appears to me most delightful, is a little village far in the country, a small neighborhood, not of fine mansions finely peopled, but of cottages and cottage-like house, such ignoble and non-descript dwellings, with inhabitants whose faces are as familiar to us as the flowers in our garden; a little world of our own, close-packed and insulated like ants in an ant-hill, or bees in a hive, or sheep in a fold, or nuns in a convent, or sailors in a ship; where we know every one, are known to every one, interested in every one, and authorized to hope that every one feels an interest in.

How pleasant it is to slide into these true-hearted feelings from the kindly and unconscious influence of habit, and to learn to know and to love people about us, with all their peculiarities, just as we learn to know and to love the nooks and turns of the shady lanes and sunny commons that we pass every day. And a small neighborhood is as good in sober waking reality as in poetry or prose; a village neighborhood, a long winding street at the bottom of a fine eminence, with a road through it, always aboundings in carts, horsemen, and carriages. Will you walk with me through our village, dear reader? The journey is not long. We will begin at the lower end, and proceed up the hill.

B. Neoslavonic Text (Latin Accented Orthography)

Из всих моžних идеалних мест, в коториž живулю люди, најболше живулю мале село, далеко от шумнога града, с њега малој обичној. Оно не имеје преплјени блоки, само мале д’рене будникы. Јест то прости и слићне мesta до љубова, с људима, којими лица су тут ровно знаме како цивити у нашем саду. То јест затворени свећ с немного људима, близко у љубимим како мравки в мравковој купи, пчеле в пчелинику, овце в овчей сталг, монашки в монастиру или мораки на корабу, кад већи знаме свега и все јест знане од свега, кад всеј се интересуе у свим и сеље може имати надђо, же нјако се у њем интересуе.

Како драго би је било задржити се у то ђурдече чувство любови и незначних потрђу, двијући се и бити колегама с тјеме људима, њима интересују. Јест то, чemu поезија и проза благодарје најбоље. Иде тамо узак дрога, јаде у дивој улици дни прђходи велику широку дрогу пљну аут и тиров. Не хотиш ли путовати с мној, драги читатели? Пут не буде дълга. Ми будемо путовати од почетка долине и оттамо идемо нагору.

C. Neoslavonic Text (Cyrillic Full Orthography)

Из всих можних идеалних мест, в кторых живут люди, найболш люблю маль село, далеко от шумнага града, с њега малю обичноно. Оно не имееме прегрязаны блоки, само малы дыврене буднники. Іест то просто и слично мєсто до житва, с жителами, кторых лица суть ровно знаеми како цивити в нашем саду. То іест затворенна свєт с немного людим, близко у љубимими како мравки в мравковы кугт, пчели в пчелинику, овце в овяй свталг, монашки в монастырю или мораки на корабу, кад вељи знане свега и все іест знаны от свега, кад весь свей се интересує в всим и весь може имати надђо, же нєкто сє в њем интересує.

Како драго би је било задржити се у то ђурдече чувство любови и незначних потрђу, двијући се и бити колегами с тĳеме људима, њима интересују. Іест то, чemu поезија и проза благодарје најбоље. Иде тамо узак дрога, јаде у дивој улици дни прђходи велику широку дрогу пљну аут и тиров. Не хочеш ли путовати с мною, драги читатели? Пут не буде дълга.

Ми будемо путовати од початку долини и оттамо идемо нагору.

VIII. ISSUE OF THE MACHINE PROCESSING

In the Neoslavonic language (as well as in all Slavic languages), the word order in a sentence is rather flexible. In English, the position of words in sentences is necessary to inform whether it is a noun, adjective, verb, subject, object or something else. Neoslavonic words have their own endings (declension and conjugation), in which the complete information about their grammatical category is stored, so there is no need for any word to respect a specific position in the sentence. This system of endings is specific for the family of fusional (or inflecting) languages.

Fusional languages do not need to use position of words for expressing grammatical categories. Another examples of fusional Indo-European languages include Sanskrit (and many of the modern Indo-Aryan languages), Greek (both classical and modern), Latin, Lithuanian, Latvian and Albanian. Another notable group of fusional languages is the Semitic languages group (Hebrew, Arabic, Aramaic). Some degree of fusion is also found in many other languages in the world.

We can say that Neoslavonic does not need to use the fixed position of words in sentences in order to express information about whether a word is subject, object, verb or something else. It is obvious that fusional languages operate with words containing unambiguous grammatical information without the need to use fixed positions. Free word order can then be used to express the finer details of communication in these languages.

Of course this does not mean that Neoslavonic words can be mixed in any haphazard way. For example, if adjectives
belong to a specific noun, they must be positioned either in front of or behind that corresponding noun, and other elements of the sentence cannot intervene between them. Metaphorically, a Neoslavonic sentence is like a branched tree, whose branches represent particular sentence components. Mutually branches may have flexible order, but elements within each branch must not be mixed with elements of another branch. From the theoretical perspective, the Neoslavonic sentence is a multidimensional oriented graph rather than a linear sequence of words.

IX. LANGUAGE MODEL OF TRANSLATION

We believe that Neoslavonic language is a useful intermediary tool for the translation between various Slavic languages. Therefore, it is necessary to create a correct model of the language for parsing and translation purposes. In our case, this means to create three formal models: model of the source language, the reference model, and the model of the target language.

The reference model is de facto an ontological model of the real world around us, as seen in general. In the Slavic world, the Neoslavonic language can be used as a common standard of this reference model for translation of all Slavic languages.

Unlike more conventional approaches in linguistics, we do not use the standard language model of a linear sequence of words. The author of the standard model is the most famous contemporary linguist Noah Chomsky [1], but it is not suitable for Slavic free-word-order languages. That's why we propose to use a different language model based on a conceptual model of linguistic ontology expressed by a directed graph. [2]

This new model expresses the content of a sentence independently on a word order, because it is based on a form of a multidimensional graph. Of course, there does exist a way how to linearize this graph into a sequence of words.

The basic principle of language interpretation and translation is very simple. It is based on graph transformations. In the source model expressed as a graph, we are looking for a subgraph that corresponds to a transformation rule. If we find such a subgraph in the source model, we replace it with another subgraph by this rule. Step-by-step, we apply sequentially all the transformation rules to the source model. The whole process will end if there is no rule to apply. Finally, we obtain a new graph that represents our target model. The graph transformations are more general than transformations of linear sequences to another linear sequences. We demonstrate our method on an example of the simple sentence from “Fig. 1”.

Fig. 1. First example - the sentence linearized by the S-V-O order.

Fig. 2. Second example - the same sentence linearized by a different order.

A conceptual model of this statement is shown in “Fig. 3”. This model corresponds to more linear models of the same sentence, for example “Fig. 1” and “Fig. 2”.

To create these two linear representations (or translations into another free-word-order Slavic language), we use a basic graph algorithm, which subset of rules and restrictions are shown in “Fig. 4”.

Fig. 3. Conceptual model of the sentence.
XI. RESULTS

A. Internet News - http://IZVIESTIJA.info

During the first year of operation since October 2011, this news has read ¼ million article readings by 14,000 unique Internet readers from 93 countries. Today, there are readers from 130 countries and 53 readers from internet satellite connection (e.g. astronauts and polar explorers).

Articles are written by more authors in various dialects of Slavic conlang. Interslavic news are operated by the Interslavic Community. [6]

B. Neoslavonic in Film and Video

Czech screenwriter, producer and director Václav Marhoul signed three preliminary coproduction agreements in the International Film Festival in Cannes for the adaptation of Jerzy Kosinski's Second-World-War novel "The Painted Bird" published in 1965. The feature film of 5 million EUR budget will be shot beginning in March 2016 and ending in June 2017. The Neoslavonic language will be used as the unspecified Slavic language of some film characters in this story.

Neoslavonic is also used for movie subtitles. They are: Russian film the Island from 2006, Czech film Stolen Kosovo from 2012, series of Polish TV cabaret show from 2015, and yet miscellaneous amateur YouTube video channels.

C. Neoslavonic in the Internet

There are two main Facebook groups directly managed by the Interslavic community: Neoslavonic and Interslavic. Together, they include about 1,100 members.

Likewise, Interslavic zonal conlangs are occasionally used by members of other groups in Facebook and VKontakte.

D. Neoslavonic in the Public Life

The Slavic Union belongs to the oldest Slavic societies operating in the Czech Republic. It was founded in 1992 by the Czech writer Alexej Pludek and derives its tradition from the first Slavic Congress in 1848 in Prague. It is a separate legal entity #48133396 and operates in the Czech Republic under the Czech law. The purpose of the Slavic Union is cultural, educational, and training activities. Based on the statutes of the Slavic Union, the official languages of all stored documents are Czech or Slovak or Neoslavonic written in the Latin alphabet. Documents in any other languages, including Russian and English, must be accompanied by a translation into one of these official languages. The website of the Slavic Union is regularly archived by the National Library of the Czech Republic for its cultural, educational, scientific, research or other values with the aim of documenting an authentic sample of the Czech web. It belongs to a collection of Czech websites to be preserved by the National Library and made available for posterity. Its record is included in the Czech National Bibliography and the National Library catalogue.

The research team of prof. Emil Heršak from the University of Zagreb in Croatia started teaching a new course called "Linguistic relations and Neoslavonic" for students of the Faculty of Philosophy.

XI. CONCLUSION

Modern people are travelling much more than in the past. Six Slavic languages are official languages of the European Union member states. Except for the long time separation of Slavic nations, Slavic people still feel their mutual cohesiveness and affinity. This is often well manifested during tourism, business and social events, when Slavic people from miscellaneous nations tend to speak together not in English but in some randomly assembled common inter-Slavic language, which is sufficient for many situations. The Neoslavonic language is therefore a professional response to this phenomenon as follows:

- A universal common auxiliary language usable in business, international transport (information texts and labels in trains, buses, planes, ...), marketing (product manuals and description), tourism (leaflets, news, ...) and social events.
- An auxiliary language tool to facilitate rapid knowledge acquisition of Slavic languages using a receptive method, by which a person is able to gradually receive some real Slavic language using his/her native Slavic language background combined with Neoslavonic. During this learning process, Neoslavonic language is gradually replaced by the knowledge of the new adopted language.
- An auxiliary language demonstrating cultural and historical reality for interested people in experimental archeology or cultural anthropology in order to bring together people from various Slavic nations.
- A practical usable auxiliary language for multinational Slavic groups in touristic destinations, historical and cultural places, exhibitions, and religious communities.
- A practical tool to demonstrate the similarity of today’s Slavic languages for teaching and learning
purposes. This covers also historical themes where Neoslavonic imitates the original Old Slavonic language (e.g. movies, theatre, fine art, artistic literature, ...).

- A theoretical tool that would enable easier and more accurate machine translation between modern Slavic languages. Also, easier and more accurate tool of machine translation between Slavic and other languages.

The theoretical idea, we work with, is the interpretation of a sentence in a form of a multi-dimensional graph rather than a linear sequence of words. Of course, each spoken language is a linear sequence, but it is not the basic formal structure but only a necessary way of sound-based information transmission. The reference model of many languages can be multi-dimensional. We believe that our approach can be used not only for machine processing of spoken free-word-order languages, but also for another multi-dimensional languages such as sign languages for deaf people, information of road signs, or information stored in various cockpit panels, for example.

ACKNOWLEDGMENT

The authors would like to acknowledge the use of the yED Graph Transformation Tool [12], the support of the research grant SGS14/209/OHK4/3T/14, and the support of the Slavic Union, which is a voluntary, independent, and non-political association operating as a legal entity #48133396 the Czech and E.U. law (http://slovane.org).

After all, we would like to thank for the thorough and dedicated work of all our colleagues in the project Interslavic, especially the coordinator Mr. Jan van Steenbergen.

REFERENCES
