[Mechanical Translation, vol. 8, No. 1, August 1964]

Slavic Languages: Comparative Morphosyntactic Analysis*

by Milos Pacak, Westchester Laboratory, Itek Corporation

This paper discusses the results of a comparative study of distributional equivalences among adjectivals in four Slavic languages, namely, Russian, Czech, Polish and Serbo-Croatian. A procedure for determining equivalence is defined, and is applied to the results of analyzing the adjectivals of each language with respect to gender, animateness, and case and number.

A appropriate goal for present-day linguistics is the development of a general theory of relations between languages. Classification which is based on common origin is fundamental for historical and comparative linguistics. A group of four major Slavic languages—Russian, Czech, Polish and Serbo-Croatian—was selected for comparative investigation because of the similarities stemming from their common origin and from subsequent parallel development. The comparative computer-oriented analysis of this group of Slavic languages was conducted in order to ascertain whether the similarities in structure of a group of related languages might permit of developing a common system of morphology and syntax which would facilitate machine translation to and from those languages. The research might also indicate whether a core system of morphology and syntax is useful for groups of languages which are not related. The possibility of a common general syntax for a group of related languages was suggested by L. E. Dostert.[2] It should be stressed that this report refers only to a small part of a major problem and is not intended to assert general conclusions about the results of an overall linguistic analysis.

Morphosyntactic Analysis

The first stage of our investigation was concentrated on the identification and classification of inflected forms in terms of their morphosyntactic properties. An attempt was made to set up classifications by choosing criteria which are common for all four Slavic languages mentioned above. First of all, a computer-oriented transliteration system was established. The total number of Cyrillic and Latin characters in the four languages is 80. These are represented in the transliteration by 51 signs, of which 25 consist of single symbols and 26 are digraphs. The objective of our comparative research was limited to the establishment of the patterns of the distributional identity of two major classes of morphological components: (a) the class and subclass of adjectival stem morphemes, and (b) the class of inflectional morphemes which are automatic in respect to the class of stem morphemes. The relationship between the major classes of stem morphemes and inflectional morphemes is defined as the functional dependence of the dependent variables upon the independent constant:

$$f(x,y),$$

where 'x' is the distributional class of the derived stem morpheme (which is a constant) and 'y' is the class of inflectional morpheme (which is a variable). The morphosyntactic (grammatical) value of inflected forms is the logical sum of the class or subclass value of the stem morpheme and the class or subclass value of the inflectional morpheme:

$$\sum (X_n\ Y_m),$$

where X is the class of stem morpheme and subscript n denotes a subclass of X and Y is the class of inflectional morphemes with subscript m denoting a subclass of Y. The morphosyntactic value of the stem and inflectional morpheme combination is either single (the given inflected form has an unambiguous morphosyntactic function) or multiple (the given inflected form is ambiguous).

Comparative Procedure

The tentative comparative procedure was based on the establishment of patterns of (a) absolute equivalence, (b) partial equivalence, and (c) difference. Absolute equivalence exists when the distribution, and consequently the morphosyntactic function, of the members of a class or subclass of inflected forms is identical in all four of the languages mentioned above. Partial equivalence exists when an identical morphosyntactic function is shared by some, but not all, of the languages under consideration. A difference exists when a certain morphosyntactic function is found in one language only (unique distribution).

Comparison of Adjectivals

In the previous part, the general methodological approach to synchronic comparative linguistic analysis was discussed. The applicability of this procedure was

* This work was accomplished at the Georgetown University Machine Translation Research Project and was supported in part by EURATOM and in part by the U.S. Atomic Energy Commission. The author wishes to thank Dr. R. R. MacDonald and B. Henisz-Retman for their valuable suggestions at various stages in this study.

tested on the class of adjectivals in Russian, Czech, Polish, and Serbo-Croatian. After analyzing the adjectivals in the languages mentioned above independently, a comparative distributional analysis was made. The results obtained are as follows:

The number of inflectional morphemes for adjectivals in each of the four languages is:

Russian	39
Czech	49
Polish	27
Serbo-Croatian	18

The length of the inflectional morphemes (not transliterated) ranges from one to four graphs.

The number of subclasses that were established within the class of adjectivals is:

Russian	11 subclasses
Czech	9 subclasses
Polish	9 subclasses
Serbo-Croatian	9 subclasses

Three morphosyntactic properties of the class of adjectivals were considered and compared: (a) category of gender; (b) category of animateness; and (c) category of case and number. The results of the comparison are:

Category of Gender

ABSOLUTE EQUIVALENCES

All three genders (masc., fem., neuter) are always distinguished by inflectional morphemes in the nominative and accusative singular in all four languages.

Examples:

NOV+Y1 (M) ; +A4 (F) ; +OE (NTR)— Russian

NOV+Y (M) ; +Á (F) ; +É (NTR)— Czech

DOBR+Y (M) ; +A (F) ; +E (NTR) — Polish

ZELEN+Ø (M) ; +A (F) ; +O (NTR) — Serbo-Croatian

The contrast between the masculine and neuter on the one hand as against the feminine on the other is marked in the accusative case of the singular in all four languages.

Examples:

NOV + UH — Russian
NOV + OU — Czech
DOBR + A — Polish*

ZELEN + U — Serbo-Croatian

The gender is not marked in the genitive, dative, prepositional or instrumental plural in any of the languages compared.

* —A is also the marker of the instrumental singular, feminine.

PARTIAL EQUIVALENCES

All genders are distinguished in the nominative and accusative plural in Serbo-Croatian and Czech—with the exception of one paradigmatic subclass in Czech.
Examples:
Serbo-Croatian:

ZELEN + —I (nom. pl. masc.)
—E (acc. pl., masc.)
—A (nom. + acc. pl., neuter)
—E (nom. + acc. pl., fem.)

Czech:

ZELEN + —I /—É (nom. pl., masc.)
— É (acc. pl. masc.)
—Á (nom. + acc. pl., neuter)
—É (nom. + acc. pl., fem.)

DIFFERENCES

In Polish, the distinction in gender in the nominative and accusative plural is connected with the personal and non-personal aspects of the noun which is modified. Gender is not distinguished in any case of the plural in Russian.

Category of Animateness

The category of animateness as against inanimateness is characterized in general by the morphological identity of the nominative and the accusative case if the adjectival modifies an inanimate noun; if the modified noun is animate then the genitive and the accusative case of the adjective are morphologically identical.

However, in Polish the category of animateness is subdivided into two sub-categories in the masculine gender only; personal and non-personal are marked by morphological contrast in the masculine plural only (A=D non-personal; B=D personal).*

ABSOLUTE EQUIVALENCES

a. If A modifies N / G1 / $A2^2$ in the singular or the plural, the nominative and accusative case are identical in Russian, Czech and Polish. In Serbo-Croatian, there is a morphological contrast between the inflectional morpheme —I in the nominative plural and the inflectional morpheme —E in the accusative plural.
b. If A modifies N / G3 / A1 v A2 / SG v PL, the nominative and accusative are identical in the singular and plural in Russian, Czech, Polish and Serbo-Croatian.

DIFFERENCES

There are nine differences which are unique for the Slavic Languages under consideration. Three of them

* See the appendix for a list of symbolic notations.

are unique for Czech, three for Russian, two for Polish and one for Serbo-Croatian.

Category of Case and Number

The total number of single and multiple morphosyntactic values which refer to case and number is 78 in the four languages under consideration. The distribution of equivalences and differences is as follows:

Absolute equivalences	6
Partial equivalences (3 languages)	6
Partial equivalences (2 languages)	13
Differences	10

However, it must be noticed that the total morphosyntactic value is a logical product of all three categories mentioned. If all three categories are compared simultaneously the number of distributional patterns which are identical in all four languages is four (absolute equivalences) as compared with 11 patterns of partial equivalence and 87 patterns of difference.

An example of an absolute morphosyntactic equivalence is the following formation rule:

$$[(Ax)_{R,C,P,SC.}]\ [(EGO/OGO)_R \vee (—ÉHO/—IHO)_C \vee (-EGO/-IEGO-)_P \vee (-EGA/-OGA)_{SC}] \bullet [(G_1 \bullet A_1) \supset (B \vee D)] \vee [(G_1 \bullet A_2) \supset (B)] \vee [(G_3) \bullet (A_1 \vee A_2) \supset (B)]_{R,C,R,SC.}$$

If there is an adjectival stem morpheme A belonging to the distributional subclass x in all four languages (R, C, P, SC) and if it occurs with the set of inflectional morphemes —EGO/—OGO in Russian, —ÉHO/-IHO in Czech, -EGO/-IEGO in Polish, or -EGA/—OGA in Serbo-Croatian, then if that adjectival modifies a noun which is masculine and inanimate (G_1A_1) it marks the genitive or accusative singular (BvD); if the modified noun is masculine and inanimate, the adjective marks the genitive singular only (B); if the modified noun is neuter animate or inanimate, the adjectival marks the genitive singular only (B).

The other morphosyntactic patterns of absolute equivalences are:

1. (G1.A1) $\supset$ (A) v (G1.A2) $\supset$ (AD), exhibited by the inflectional morphemes -Y1/-I1/-1/-Ø/-OT in Russian, -Y/-Ø/-EN/-UJ/ in Czech, -Y/-Ø/-EN in Polish, and -I/-0 in Serbo-Croatian;
2. (G2.A1 v A2) D (D), exhibited by the inflectional morphemes -U/-H/-UH/-HH/-OE in Russian, -OU in Czech, -E in Polish, and -U in Serbo-Croatian;
3. (G3.A1 v A2) $\supset$ (AD), exhibited by the inflectional morphemes -E/-O/-EE/-OE in Russian, and -E/-É/-O/I in Czech, and -E in Polish and Serbo-Croatian.

The largest number of differences was found in Serbo-Croatian and the smallest in Polish. The high number of morphosyntactic values which are different is due to the multiplicity of morphosyntactic properties (category of case and number, category of gender, category of animateness) which are conveyed by adjectival inflectional morphemes functioning as markers of syntactic relations. However it seems possible to reduce the number of multiple syntactic values partially by an additional subclassification of adjectivals.

Adjectivals can be classified on the basis of their syntactic function, namely those which function as: (a) modifiers only, (b) nominals only, or (c) both modifiers and nominals. An additional useful subclassification could be based on the admissible agreement with animate nouns only, inanimate nouns only, or both. The semantic classification of adjectivals is another large field which must be studied.

Katz and Fodor in their recent article, "The Structure of a Semantic Theory,"[6] defined the semantic relationship between the modifier and the modified element as the process of creating a semantic unit, compounded from a modifier and a head, except that the meaning of the compound is more specific than that of the head alone. We attempted experimentally to identify and classify a group of adjectivals which can function as semantic modifiers of a subclass of nouns. For example, the basic meaning of the adjectival form CERNY 1 in Russian is "black." If CERNY 1 modifies a certain subclass of nouns (METALLURGIYA; RABOTA), it loses its basic meaning and becomes a member of a larger conceptual unit (A denotes N):

CERNAYA METALLURGIYA = ferrous metallurgy
CERNAYA RABOTA = manual work

An example in English is the unit 'hot dog,' in which both elements have lost their basic meaning and form a new conceptual unit. However, this is only a very small part of a much larger problem which will have to be studied more extensively.

Conclusions

If single categories are considered and compared, the number of absolute and partial equivalences is higher than if all categories are compared simultaneously.

The multiplicity of morphosyntactic properties might lead to mismatchings, which would produce meaningful combinations which are valid for one language but which are not permissible in other languages.

The multiplicity of morphosyntactic properties affects proportionally the quantitative comparison between related languages.

It is assumed that the set of formation rules will be less complex for syntactic constructions because the syntactic properties of elements that function as initial markers of syntactic constructions exhibit a high degree of similarity in the Slavic languages.

The comparative research might be of interest to scientists who study the laws of similarity which reveal the relationship between the qualitative and quantitative aspects of certain phenomena and its applicability to computing methods.

APPENDIX

SYMBOLIC NOTATIONS

N	noun	A	nom. sg.
AD	adjectival	B	gen. sg.
A1	animate	C	dat. sg.
A2	inanimate	D	acc. sg.
Non-pers	non-personal	E	Instr. sg.
Pers	personal	F	Prep. sg.
S	inflectional morpheme	G	nom. pl.
R	Russian	H	gen. pl.
CZ	Czech	1	dat. pl.
P	Polish	J	acc. pl.
SC	Serbo-Croatian	K	Instr. pl.
G1	masculine gender	L	Prep. pl.
G 2	feminine gender		
G 3	neuter gender		
SG	singular		
PL	plural		

References

1. DeBray, R. G. A., *Guide to the Slavonic Languages,* J. M. Dent and Sons, Ltd. (London), 1951.
2. Dostert, L. E., *An Experiment in Mechanical Translation: Aspects of General Problems,* American Chemical Society, 1954.
3. Church, A., *Introduction to Mathematical Logic,* Vol. 1, pp. 48-61, Princeton University Press, 1956.
4. Greenberg, J. H., *Language as a Sign System: Essays in Linguistics,* pp. 1-17, University of Chicago Press, 1963.
5. Harris, Z. S., *Structural Linguistics,* pp. 299-324, University of Chicago Press, 1963.
6. Katz, J., and Fodor, J., "The Structure of a Semantic Theory," *Language* 39, pp. 170-210, 1963.
7. Lehmann, W. P., and Pendergraft, E., "Structural Models for Linguistic Automation," a chapter in *Vistas in Information Handling* (pp. 78-102), Spartan Books (Washington, D. C.), 1963.
8. Melchuk, I. A., "On the Standard Form and Quantitative Characteristics of Several Linguistic Descriptions," *Questions of Linguistics* 1.
9. Nikolajeva, T. M., "Opyt Algoritmiceskoi Morfologii Russkogo Jazyka," *Structurno-Tipologiceskie Issledovanija,* Akademija Nauk SSSR (Moscow), 1962.
10. Pacak, M., *Logical Scheme of Russian Morphology in Terms of MT,* Seminar Work Paper No. 74, Georgetown University, 1957.
11. Pacak, M., and Ulatowska, H., *Morphological Abstraction of Adjectivals in Czech,* MT Research Project No. 27, Georgetown University, May, 1962.
12. Pacak, M., "Syntagmatic Limits of Morphological Sets," *Method* 13 (Milano), Numbers 49-50, 1961.
13. Pacak, J. H., *Distributional Classes of Derivational Morphemes in Czech,* Master's Thesis, Georgetown University, 1959.
14. Retman, B., *Morphological Analysis of Polish Nouns,* MT Research Project, Georgetown University, June, 1962.
15. Sgall, P., "Soustava Pádových Koncovek V Češtině," *Acta Universitatis Carolina Slavica Pragensia* 11, pp. 65-84, 1960.
16. Vaillant, A., "Grammaire Comparée des Langues Slaves," *Les Langues du Monde* 12, pp. 495-541, 1958.