Order of Subject and Object in Scientific Russian When Other Differentia Are Lacking

D. G. Hays, The Rand Corporation, Santa Monica, California

The order of subject and object is an adequate criterion for distinguishing between them when other grammatical properties are ambiguous.

HARPER[1] AND LEHISTE[2] have discussed the order of subjects and predicates in Russian scientific text. Lehiste concludes that "form and function" should be used to distinguish the subject from the predicate of a Russian sentence; although her conclusion may be accepted (subject to assumptions about the value of maintaining customary English order in the output), her dictum must be converted into programmable instructions.

To a certain extent, the most economical method of distinguishing subject from predicate is obvious and straightforward. Verbs, short-form adjectives and participles, and other potential "fillers of the predicate slot" are marked in the glossary and can be identified when they occur in text. Inasmuch as some glossary entries are marked (in effect) "possibly predicate," some difficulties are involved in finding the predicate, but we wish to pass over these to a specific problem of detail.

The formal characteristics by which a subject can be recognized are, roughly, part of speech, gender, number, person, and case. The subject and predicate of a sentence are, in fact, two of its members of specifiable parts of speech, agreeing in number and either person or gender, while the subject must be of specified case, i.e., nominative. Unfortunately, for example, two nouns in a sentence may be equally good candidates for the role of subject; this is true because the nominative and accusative cases are not always formally distinct. Thus, if two neuter nouns, each nominative or accusative, respectively precede and follow a third-person, singular, non-past verb (which takes an accusative object), the choice between these nouns must be made on grounds other than morphology.

Word order and semantic agreement immediately come to mind. Semantic agreement would require thoughtful, expensive research. The hypothesis that subjects precede their predicates whenever the latter contains a noun that could be mistaken (morphologically) for the subject can be tested rapidly and inexpensively by reference to a body of data already collected at The RAND Corporation.

Method

A large volume of Russian physics text has been keypunched into IBM cards, referred to a glossary, and analyzed by translators[3]; the structure of each sentence has been determined in accordance with a <u>dependency</u> theory, and each dependency relation punched into a card. For a sample of 22, 000 occurrences (running words) of text[4], a special report has been prepared (by machine processes), showing all dependents of every occurrence in the sample; the listing is ordered by the grammatical type of the governor.

Since subject and object are regarded as dependents of the main predicate element in our theory, it is simple to scan the section of this report that is devoted to verbs and their dependents, noting the textual location of every verb with two dependents, of which either could be

1. K. E. Harper, "A Preliminary Study of Russian," in W. N. Locke and A. D. Booth, <u>Machine Translation of Language,</u> New York, Wiley, 1955.

2. Ilse Lehiste, "Order of Subject and Predicate in Scientific Russian," <u>MT,</u> 4, 1957, 66-67

3. H. P. Edmundson and D. G. Hays, "Research Methodology for Machine Translation," <u>MT,</u> 5, 1958, 8-15.

4. H. P. Edmundson, K. E. Harper, D. G. Hays, and A. K. Koutsoudas, <u>Studies in Machine Translation</u> - - 9: Bibliography of Russian Scientific Articles, The Rand Corporation, Research Memorandum RM-2069, October 16, 1958. (Corpus 2 was used in the present study.)

Table 1

INSTANCES OF MORPHOLOGICALLY INDISTINGUISHABLE SUBJECT AND
OBJECT IN A SAMPLE OF RUSSIAN PHYSICS TEXT

Subject	Verb	Object	Order	Frequency
Nom./acc. sing. noun	3rd person sing., non-past	Nom./acc. sing. noun	SVO	27
do.	do.	do.	VOS	4
Nom./acc. sing. pronoun	do.	do.	SVO	1
Non-Cyrillic	do.	do.	SVO	6
Nom./acc. sing. noun	do.	Non-Cyrillic	SVO	3
Nom./acc. plural noun*	3rd person plural non-past	Nom./acc. plural noun	SVO	7
do.	do.	Non-Cyrillic	SVO	1
Conjunction of proper names	do.	Nom./acc. plural noun	SVO	1
Nom./acc. sing masc. noun	Masc. sing past	Nom./acc. sing. masc. noun	SVO	1
Name (indeclinable, masc.)	do.	do.	SVO	1
Nom./acc. sing. neut. noun	Neuter. sing. past	Nom./acc. sing. neut. noun	SVO	2
Conjunction of proper names	Plural past	Nom./acc. plural noun	SVO	1
do.	do.	Non-Cyrillic	SVO	1
		Subtotal	SVO	52
		Subtotal	VOS	4
	Total			56

* Three subjects are in apposition with conjunctions of Non-Cyrillic occurrences.

subject. All doubtful cases were noted as well. A 3x5 card was prepared for each such occurrence, and the cards (about 100 in number) were sorted into textual order.

Examination of all 100 occurrences required only about 3 hours. Doubtful cases were resolved, situations in which a modifier of either noun distinguished its case were recognized and discarded, subject and object were differentiated by careful human judgment, and their order was noted on each card.

Results

Just 56 instances of true ambiguity were found in 22, 000 occurrences.[5] They are summarized in Table 1. The subject precedes the verb 52 times; the object follows the verb 56 times. When both object and subject follow the verb, the object precedes the subject 4 times.

The 4 sequences V-O-S are:

Обращает внимание наличие (The presence [of..] calls attention [to..])

_____

5. If an adjectival modifier forms an unambiguous noun phrase with either subject or object, or if negation of the verb calls for a genitive object, the instance is irrelevant to the present study.

Имеет место состояние (a state that occurs)

Имеет место правило (a rule occurs)

Имеет место уменьшение (a decrease occurs)

Note that the verb-object pair might be regarded as idiomatic on grounds other than those of the present study; neither is translated literally.

Conclusions

On the basis of a preliminary study of the 56 relevant instances in 22, 000 running words of text, we conclude that: If two nouns in a sentence cannot be distinguished as subject and object of a transitive verb by their morphological properties, and if one precedes the verb while the other follows, the first noun is the subject. This rule, together with adequate coverage of idioms, appears entirely effective. The study should be repeated on a larger sample of text, however.

_____