

The Problem of the Syntactic Head in Polish and Russian Constructions with Numerals (Sidenotes to Igor Mel'čuk's Book "The Surface Syntax of Russian Numeral Expressions")

Zygmunt Saloni
University of Varmia and Masuria
Olsztyn, Poland
saloni@mercury.ci.uw.edu.pl

Mots-clefs – Keywords

polonais, russe, numéraux
Polish, Russian, numerals

Abstract - Résumé

This paper deals with Polish and Russian constructions that contain numerals. In some of them the case requirements of the verb (nominative or accusative) are not clearly fulfilled. Similar properties are also characteristic of constructions in which alien elements (e.g. mathematical formulas) occur in the position of a numeral form. The surface syntactic analysis of such constructions leads to the conclusion that what should be regarded as the head of the whole construction is the alien element. The same interpretation also fits constructions with regular numerals.

Le problème du noeud syntaxique principal dans les constructions polonaises et russes avec les adjectifs numéraux. Remarques en marge du livre d'Igor Mel'čuk «Syntaxe de surface des expressions numériques russes»

Dans l'article sont analysées les constructions polonaises et russes qui contiennent des adjectifs numéraux. Il arrive que dans de telles constructions les exigences casuelles de la forme verbale (nominatif ou accusatif) ne soient pas clairement réalisées. On observe une situation semblable dans les constructions où apparaît en position de forme numérique un élément étranger, c'est-à-dire, non-linguistique (p. ex. une formule mathématique). L'analyse syntaxique de surface de telles constructions montre que l'élément étranger doit être considéré comme le noeud principal de la construction. La même interprétation peut s'appliquer aux constructions contenant un adjectif numéral régulier.

The aim of this paper is to discuss some difficult and controversial problems of interpretation of some Polish and Russian constructions (consisting of a noun and a numeral) whose head is not clearly visible at first sight, e.g.:

- | | | | |
|-----|------------------|----------------------------|---|
| (1) | Widzę pięciu | chłopców. | |
| | see-I | five[<i>m-human</i>] | boys[<i>gen</i> ∨ <i>acc</i>] 'I see five boys.' |
| (2) | Widzę pięć | dziewczynek. | |
| | see-I | five[¬ <i>m-human</i>] | girls[<i>gen</i>] 'I see five girls.' |
| (3) | Widzę pięć | okien. | |
| | see-I | five[<i>non-m-human</i>] | windows[<i>gen</i>] 'I see five windows.' |
| (4) | Vižu pjat' okon. | Вижу пять окон. | |
| | see-I five | windows[<i>gen</i>] | 'I see five windows.' |

The point of departure is the analysis of Russian numeral expressions provided by Igor Mel'čuk (Mel'čuk 1985)¹. His analysis is founded on the stable and consistent methodological basis of "Meaning $\iff$ Text" Theory (Mel'čuk 1974). My remarks are formulated from the point of view of syntactic analysis and its relation to morphological representation of word forms in Polish (and, partially, in Russian).

The problem with sentences (1)-(4) consists in the contrast between the government of the verb (accusative) and the form of the substantive (noun) which for semantic reasons is normally treated as the head of the noun phrase. If we omit numeral forms in these sentences the resulting expressions are semantically full (the quantity is not introduced into the message), but the results are problematic:

- | | | |
|------|---------------------|--|
| (1') | Widzę chłopców. | |
| | see-I | boys[<i>gen</i> ∨ <i>acc</i>] 'I see boys.' |
| (2') | *Widzę dziewczynek. | |
| | see-I | girls[<i>gen</i>] 'I see girls.' |
| (3') | *Widzę okien. | |
| | see-I | windows[<i>gen</i>] 'I see windows.' |
| (4') | *Vižu okon. | *Вижу окон. |
| | see-I | windows[<i>gen</i>] 'I see windows.' |

The sentences (2')-(4') are incorrect because the case requirement of the verb form is not fulfilled as noun forms occurring in them are genitive; the sentence (1') is correct as genitive and accusative forms of the masculine-human nouns (gender *MI*) are always syncretic. We can easily correct all these sentences by changing the case into the accusative:

- | | | |
|-------|--------------------|---|
| (2'') | Widzę dziewczynki. | |
| | see-I | girls[<i>acc</i>] 'I see girls.' |

¹I announced this discussion in my previous paper, inspired by Mel'čuk's book and devoted to syntactic properties of Polish numerals (Saloni 1992). I also quote here several examples from that paper.

- (3'') Widzę okna.
see-I windows[acc] 'I see windows.'
- (4'') Vižu okna. Вижу окна.
see-I windows[acc] 'I see windows.'

On the contrary, the omitting of the noun in sentences (1)-(4) leads to sentences which are elliptic (it is not said what is the nature of the object, only the number) but correct:

- (1''') Widzę pięciu.
see-I five[m-human] 'I see five.'
- (2''') Widzę pięć.
see-I five[¬m-human] 'I see five.'
- (3''') Widzę pięć.
see-I five[¬m-human] 'I see five.'
- (4''') Vižu pjat'. Вижу пять.
see-I five 'I see five.'

Such sentences can be used by native speakers of Polish or Russian, if the missing information is interpolated from the outside.

Therefore the problem what is the head of the construction Num + S (numeral + substantive) arises. Mel'čuk characterizes it as follows: "The construction Num + S in Russian is a linguistic platypus, i.e. an egg-laying mammal—a contradictory being, and, as I suppose, it simply cannot be described to the satisfaction of all interested people in all respects. For instance, we observe here a discrepancy between the two main criteria for deciding what is the direction of SSynt-dependencies [surface syntactic dependencies]: the criterion of imposing the SSynt-valences and the criterion of the contact point. [...] In all other cases in Russian the contact point of a phrase, i.e. that element by which the phrase is formally (= morphologically) connected with its context, is its SSynt-head. However, in the construction Num + S in the nominative and the accusative the contact point is the daughter, i.e. Num: *Ja vižu dva* [acc] *stola* [gen] 'I see two [acc] tables [gen]'. This Num gets the case (nominative or accusative), induced by the syntactic context for the construction as a whole." (authorized translation)².

In the question of the direction of the surface-syntactic dependency in Russian expressions Num + S (numeral + substantive) Mel'čuk argues that the only acceptable point of view is that the head is always the substantive (Mel'čuk 1985, p. 59 ff.). My previous consideration of Polish

² „Конструкция Num + S в русском языке — это своеобразный утконос, т.е. лингвистическое яйцекладущее млекопитающее, и, как я полагаю, ее просто невозможно описать так, чтобы удовлетворить всех и во всех отношениях. Так, например, здесь вступают в противоречие два основных критерия определения направления ПС-зависимостей: критерий навязывания ПС-валентностей и критерий контактной точки. [...] Во всех остальных случаях в русском языке контактная точка словосочетания, т.е. тот его элемент, через который она формально (= морфологически) взаимодействует с внешним контекстом, является его вершиной, однако в конструкции Num + S в именительном и винительном падежах контактной точкой является слуга, т.е. Num: *Я вижу два* [вин] *стола* [род], которое и получает падеж (именительный или винительный), индуцируемый синтаксическим контекстом для всей конструкции в целом.” (Mel'čuk 1985, p. 80)

constructions with numerals (Saloni 1977) and (Gruszczyński, Saloni 1978) impelled me to presume that in the description of Polish surface syntax it is more fruitful to decide that in analogous constructions the head is the numeral (the features of many Russian constructions are in this respect similar to Polish). Let us discuss the arguments for such a decision.

In the consideration we will use the results of the morphological analysis of the words of the sentence, whose dependence tree we want to obtain, and a general intuitive understanding of the meaning (sense). We assume that a deeper formal representation of the structure of a sentence should be obtained after its syntactic analysis.

According to our view many features of the units considered in the analysis should be caught in a preliminary informal way. A linguist endeavors to describe some phenomena as suitable to the structure whose general picture he has in mind. When he analyzes difficult phenomena, he wants to interpret them by analogy to others, sufficiently close to those discussed but simpler and easier to interpret.

So we will start with the problem of some other atypical constructions whose interpretation should be coordinated with the analysis of numeral expressions shown above.

In the syntactic analysis of a text it is necessary to take into account that the sequence of characters may contain some signs which do not belong to the vocabulary of the language in which this text is written. It is impossible to anticipate all proper nouns we can meet in the analysis, although it is necessary to provide their presence in the text — on the position of noun phrases. On the same basis other elements in text may occur, even consisting of symbols not belonging to the set of characters of the analyzed language. The typical examples are quotations in other languages, using a different alphabet.

This phenomenon has important consequences for syntactic analysis. The ways of using alien expressions in Polish and Russian are similar. Normally, I work with Polish texts and have Polish examples prepared for the occasion of a presentation. However, during the discussion with the solutions taken in MTT it will be better to illustrate my theses with Russian examples, especially that alien insertions (in Latin alphabet) are more frequent in Russian texts than in Polish ones.

Let us take an example from “Eugene Onegin” (Chapter II, XXXIII):

И русский <i>H</i> как <i>N</i> французский Произносить умела в нос.	‘And the Russian <i>H</i> , like the French <i>N</i> , she could pronounce nasally.’
---	---

Let us analyze it:

- (5) I russkij *H* kak *N* francuzskij
 and Russian[sg&m&acc] *H* like *N* French[sg&m&acc]

 Proiznosit’ umela v nos.
 pronounce[inf] could-she in nose

It is possible to consider *H* as an element of Russian vocabulary; it would be rather difficult to treat the Latin *N* in similar way. However, this *n* is an element of a Russian text. It is not only

used in the position determined by elements occupying higher levels of the syntactic hierarchy, but also has its own syntactic subordinates (французский). It is easy to call this position: the position of the noun phrase in the case determined by the head. The alien element gets the properties of a noun in this case. Its gender and number are dependent on various factors which we won't consider here.

There is no sense to introduce into the lexicon of the analyzed language such elements as *N* inserted in a Russian text, unless they are really frequent in some sort of texts. Generally, it is impossible to predict which elements may occur in the texts of analyzed language. However, their occurrence and their potential properties should be taken into account in any advanced syntactic analyzer of texts in a given language.

Mel'čuk discusses such alien elements of the texts included in the names of numbers. He treats them as a special "part of speech", formula, to which belong "all non-alphabetic expressions occurring in texts" (Mel'čuk 1985, p. 40). They are the secondary way of realization of a quantifier, standing with a noun. Some syntactic properties of the name of a number used in Russian (and Polish) texts are more visible when alien elements occur in a sentence instead of regular numerals. In order to show them let us consider an example given by the author himself³:

На дне глубокого сосуда	'On the bottom of a deep vessel
Лежало ровно <i>n</i> шаров;	Exactly <i>n</i> balls lay.
Попеременно их оттуда	They were alternately taken from there
Таскали двое дураков.	By two fools.
Ввиду события такого	For the sake of such an event
Сколь вероятность велика,	How big is the probability
Что был один глупей другого,	That one of them was more stupid than the other
Когда шаров он вынул <i>k</i> ?	When he took <i>k</i> balls.'

Let us analyze the first clause, up to the semicolon:

- (6) Na dne glubokogo sosuda
 On bottom[loc] deep vessel[gen]
 ležalo rovno *n* šarov
 lay[sg&n] exactly *n* balls[pl&gen&M]

The center of the clause is the verb form *лежало* (*ležalo*); other segments should be treated as (directly or indirectly) subordinated. The expression in the first verse is a local modifier, easily

³We won't analyze more complicated names of numbers, consisting of symbols not belonging to the alphabet of the analyzed language, e.g.:

Женщин можно положить $2 \cdot (n!)$ способами. (Mel'čuk 1985, p. 41)

Он преобразуется в

$$\frac{\Gamma_1 \rightarrow \Theta_1, \exists b \exists b, \Gamma_2 \rightarrow \Theta_2}{\frac{\Gamma_1, \Gamma_2^* \rightarrow \Theta_1^*, \Theta_2}{\Gamma_1, \Gamma_2 \rightarrow \Theta_1, \Theta_2}} .$$

(Математическая теория логического вывода. Moskva 1967, p. 36)

They have the same syntactic properties as simpler expressions analyzed here.

understood and analyzed automatically, it does not have any immediate relation to any word in the sentence besides the head. This head, however, demands the subject (semantically, the necessary condition of lying is something or somebody who is lying). This complement of the verb should be in the nominative case. But in the analyzed clause there is no nominative of any Russian noun.

The subject (in potential nominative) should be contained in the second verse, which is semantically and syntactically complete. It contains an adverb *ровно* (*rovno*), a substantive *шаров* (*šarov*) whose natural interpretation is genitive plural, and an element of indefinite character, an alien element *n*. The adverb evidently is not a subject. It can be joined as a subordinate directly either backward, to the verb, or forward, to the alien element.

We understand the verse, we know that the balls lay on the bottom of a deep vessel. But the noun referring to them is in genitive plural, although the verb demands the nominative. If an existing Russian word form, e.g. *пять* (*pjat'*), had been used instead of *n* we could have analyzed the whole construction Num + S and treated *шаров* (*šarov*) as its head put in untypical case and joined with subordinate in the nominative. Moreover, in one (and more natural and frequent) interpretation, we would have decided that this *пять* (*pjat'*) has its own subordinate (*ровно*) (*rovno*): there were exactly five balls there. Such a solution for an element not belonging to the vocabulary would be artificial.

In the last verse of the quoted passage the situation is still more complicated because the construction Num + S is discontinuous and, in consequence, the immediate-constituent tree of the whole sentence is not projective. It is a real problem how to analyze the following clause:

- (7) *šarov* *on* *vynul* *k* шаров он вынул *k*
 balls[pl&gen] he[nom&sg&M] took[sg&m] *k* 'he took *k* balls'

The verb form governs two cases: the nominative and the accusative. It is easy to find the first, *он* (*on*); instead of the accusative we have a word whose natural interpretation is genitive *шаров* (*šarov*) and—separately—an alien element (*k*). In Mel'čuk's interpretation these two elements would create a syntactic unit subordinated to the verb, non-continuous, and having the head in the a case inconsistent with the syntactic structure of the whole sentence (determined by the verb form). It is worth noticing that such an interpretation agrees with our intuitive understanding of the sentence (and its representations on the deeper levels of the linguistic analysis): those were balls that were taken by somebody and the number of them was *k*.

However, we do not have these intuitions during automatic syntactic analysis. All we consider are words having some morphological interpretation and an element not belonging to the lexicon. The analysis postulating the realization of the noun phrase in accusative (governed by the verb) as a construction with a head in genitive (governed by its subordinate whose form influences the head and, additionally, in the considered sentence has a textual exponent not belonging to the vocabulary) would demand very complicated mechanisms. The natural method of the analysis of this sentence is to formulate the hypothesis that the alien element included in it takes the position of the accusative (we should provide such mechanisms for other noun phrases) and as an element of a Russian syntactic construction acquires some properties of Russian word forms. Exactly speaking, it acquires syntactic properties of the form *пять* (*pjat'*), cf.:

- (7') *šarov* *on* *vynul* *pjat'* шаров он вынул пять
 balls[pl&gen] he[nom&sg&M] took[sg&m] *pjat'* 'he took five balls'

The difference is that *пять* (*pjat'*) is an element of the lexicon, and constructions containing it should be included in the syntactic description of Russian. It would be rather difficult to take into account the syntactic constructions with alien elements influencing other members. They can acquire such abilities occasionally when they are introduced into the text.

The consequence of this decision is to assign to the alien element the function of the head of a construction Num + S.

It seems to me (as far as I understand the methodology of the theory) that SSynt-valence of an alien element (its capability to be used in certain syntactic positions) is similar to SSynt-valence of a noun. In particular, they can have similar subordinates. As it has been shown above, an alien element can occur in a text in the position of a noun phrase, and if this phrase is complex, it takes the position of its head. It concerns also the Russian and Polish constructions in which alien elements should be semantically treated as quantifying modifiers of the noun.

The examples of alien elements, used in Russian texts as names of numbers in other cases were given in Mel'čuk's book (Mel'čuk 1985, p. pp. 31–32). I will not analyze them here as less disputable. However, it is worth to mention one difference in treating linguistic data by Mel'čuk and the linguists who try to build effective procedures of syntactic analysis of natural-language texts (at least, Polish). He tries to analyze the texts as occurring in the most natural form: spoken; they analyze written texts. As a consequence, in Mel'čuk's book numerical insertions are analyzed in their spoken realizations, read by somebody (sometimes such reading is not natural), although the analysis of a written text would suggest that all names of numbers written in numeric form, especially in figures, are alien elements too. And these numbers have properties analyzed above. It is much more fruitful to treat them as heads of noun phrases.

We can return to the constructions we started with:

- (4) *Vižu pjat' okon.* *Вижу пять окон.*
 see-I five windows[*gen*] 'I see five windows.'

We stated that they are analogous to the constructions containing alien elements. And we should analyze them in the similar way.

The same is true also for Polish examples:

- (2) *Widzę pięć dziewczynek.*
 see-I five[*¬m-human*] girls[*gen*] 'I see five girls.'
- (3) *Widzę pięć okien.*
 see-I five[*non-m-human*] windows[*gen*] 'I see five windows.'

It is much more natural to decide that the head is the numeral form.

Similar relations are observed in the nominative:

- (8) *Pięciu chłopców bawiło się.*
 five[*m-human*] boys[*gen*∨*acc*] played 'Five boys played.'
- (9) *Pięć dziewczynek bawiło się.*
 five[*¬m-human*] girls[*gen*] played 'Five girls played.'

The noun forms in (8) and (9) cannot be interpreted as nominative, characteristic of the subject. They are genitive forms, governed by numerals which fulfil the nominative requirement of the verbal form. On the other hand, the verb forms are caused by their subjects, i.e., the numeral forms, to appear in the neuter singular. It is worth noting that the nominative numeral-noun phrases in (8) and (9) are the only ones possible for the given noun lexemes (CHŁOPIEC or DZIEWCZYŃKA) and a numeral form corresponding to the number 5. In particular it is impossible to connect a construction of the type *pięciu chłopców* with a predicate in the plural (constructions with such characteristics do exist in Russian).

In consequence in Polish we should analyze all constructions of the type Num + S where the case of the noun is in discrepancy with the demands of the constituent of higher rank in the syntactic hierarchy as having the numeral as the head. Consequently, the next step is similar interpretation (in surface-syntactic analysis) of all names of numbers in considered. Such a solution, after discussing many types of similar constructions (consisting of various numerals and nouns of all genders in various cases) we have taken in our handbook of Polish syntax (Saloni Z., Świdziński M., 1987).

References

- Gruszczyński W., Saloni Z. (1978), Składnia liczebników we współczesnym języku polskim, *Studia gramatyczne*, Vol. II, 17-42.
- Mel'čuk I. 1974, *Opyt teorii lingvističeskich modelej "Smysl $\Longleftrightarrow$ Tekst"* (*Опыт теории лингвистических моделей «Смысл $\Longleftrightarrow$ Текст»* 'Toward a Theory of "Meaning $\Longleftrightarrow$ Text" Type Linguistic Models'), Moscow, "Nauka".
- Mel'čuk I. 1985, *Poverhnostnyj sintaksis russkih čislovyh vyraženij* (*Поверхностный синтаксис русских числовых выражений* 'The Surface Syntax of Russian Numeral Expressions'), Wien, Wiener Slawistischer Almanach.
- Saloni Z. (1977), Kategorie gramatyczne liczebników we współczesnym języku polskim, *Studia gramatyczne*, Vol. I, 145-173.
- Saloni Z. (1992), New remarks on Polish numerals, *Le mot, les mots, les bons mots*. Ed. André Clas, Montréal, 169-182.
- Saloni Z. and Świdziński M. (1987), *Składnia współczesnego języka polskiego*, 3rd ed., Warsaw, PWN.