Russian Numerals:
An example of case treatment for Slavic

Term Paper for the Seminar "HPSG for Slavic"

Tatjana Scheffler
tasc@coli.uni-sb.de

August 24, 2003

Contents

1 Introduction

In this paper I will discuss the analysis of case assignment phenomena with respect to Russian numeral phrases in HPSG.

1.1 Goals

In the first part of this paper, I will review the treatment of Polish numeral phrases that Adam Przepiórkowski proposed in his dissertation. I will show how the language-specific HPSG CASE PRINCIPLE is developed.

Then, the main goal is to explore how this principle and the underlying analysis can be transferred to Russian numeral phrases, which differ from their Polish counterparts in many interesting respects. I will sketch some possible analyses of these differences. The formalization of the analysis, although a desirable goal, must remain for further work.

1.2 Structuring of the paper

Section 2 will introduce a non-configurational treatment of case proposed by [Przepiórkowski, 1999]. Various phenomena will be discussed that need to be observed when formalizing the case analysis.

Section 3 then shows how the developed formalization in Head-driven Phrase Structure Grammar (HPSG) can be used to correctly analyse case phenomena in Polish. A short overview about the interesting phenomena in Polish is presented.

Section 4 concentrates on numeral phrases in Polish: the interactions with case assignment and case concord, and how they can be accounted for in HPSG.

The remaining sections constitute the contribution of this paper. I give an overview of numeral phrases in Russian, which differ from their Polish counterparts in interesting respects. Section 6 then shows, how the Russian numeral phrases can also be treated in HPSG using the case principle from [Przepiórkowski, 1999].

2 Case in Slavic with HPSG

There is a quite vast literature on case in Slavic languages. This is because they have a relatively rich morphology which distinguishes several cases. In this paper, we will mostly concentrate on the work by Przepiórkowski [1999]. He proposes a new treatment of case assignment in the framework of HPSG.

2.1 What is Case?

This first question is not the easiest one to answer. In a first dichotomy, case can be seen as either:

a grammatical category: cases are just formal markings of structural positions or arguments, in order to distinguish them from one another; or

a semantic category: cases have meanings (see, e.g., [Jacobson, 1971] and his *Gesamtbedeutung*)

Some accounts claim that case is a phenomenon on both of these scales. Thus, some of the particular cases of a language are more syntactic, formal categories, while others perform mainly semantic tasks.

In newer accounts in HPSG, individual cases are distinguished into *structural* (or *direct*) and *lexical* (or *oblique*, *inherent*) cases. Structural cases are those that are assigned structurally, by the grammatical function of the position the case-bearing element is located at. On the other hand, oblique cases are assigned idiosyncratically in the lexical entry of the element that assigns case. This dichotomy will be explored in more detail for Polish, below.

2.2 Previous Approaches to Case in HPSG

The first approaches to case in HPSG note, that not all cases can be assigned directly in the lexical entries. Examples that triggered this realization are Exceptional Case Marking (ECM) constructions in English such as:

(1) She let him clean the table.

In these sentences, the embedded subject stands in accusative case, because the matrix verb is a special Exceptional Case Marking verb.

More evidence comes from languages as Icelandic, and the famous quirky subject problem. Subjects in Icelandic can bear oblique case, and they retain these cases under raising:[1]

(2) **Hann** *virdist* *elska* *hana.*
He$_{nom}$ seems love$_{inf}$ her$_{acc}$.
He seems to love her.

(3) **Hana** *virdist* *vanta* *peninga.*
Her$_{all}$ seems lack$_{inf}$ money.
She seems to lack money.

Configurational case assignment Thus, case assignment in HPSG is split: While oblique cases are assigned lexically (or semantically), direct cases are assigned by a language-specific case principle. For example, in an account by Heinz and Matiasek [1994], structural NPs are assigned case by their position on the DTRS list of a phrase. In this sense, the assignment is configurational, because the DTRS attribute in HPSG reflects the tree structure. Raising issues are a problem for this account, because one needs to avoid multiple case assignment on the same object.

2.3 Non-configurational treatment of case

As the configurational approach to case assignment is problematic in the light of various phenomena, [Przepiórkowski, 1999] proposes a non-configurational treatment.

2.3.1 Problems with configurational case assignment

The first argument against configurational case assignment is a rather philosophical one. Pollard is cited to note a "traditional aversion within HPSG theory to tree-configurationally-based notion." This is of course meant in contrast to transformational grammars, where all statements are made with respect to positions on the syntax tree.

A big linguistic problem for configurational case assignment are fronted constituents in German:

(4) *[***Ein** **Außenseiter** *gewinnen]* *wird* *hier* *nie.*
an$_{nom}$ outsider win$_{inf}$ will here never.
An outsider will never win here.

[1] The examples are taken from [Przepiórkowski, 1999], and appear there as examples (3.19) to (3.20).

(5) [**Einen Außenseiter** *gewinnen]* *läßt* *Gott* *hier* *nie.*
an$_{acc}$ outsider win$_{inf}$ lets god here never.
God never lets an outsider win here.

This is because the configurational approach can not explain why the realized subject of the infinitive stands in accusative in one example. If the infinitive assigns accusative case, the first example would be ungrammatical. If it doesn't assign any case, it is still unexplained where the case of the subjects comes from. This must then be explained by using a non-configurational mechanism.

Configurational approaches express their constraints non-locally, which is undesirable. The non-configurational case assignment will avoid this.

Another serious problem arises with traceless accounts of extraction, as they are now widely assumed in the HPSG community. The extracted element usually bears the same case as it would have in the place where it was extracted from. In the configurational accounts, case assignment on extracted elements can not easily be explained when no traces are introduced in the syntactic structure.

Then, the extraction mechanism (e.g. the different kinds of lexical rules) must each specify a part of the case assignment. This makes the case assignment distributed in the grammar, and much less easily changeable. It also contradicts the desirable modularity of the grammar.

2.3.2 Non-configurational Case Principle

In the non-configurational treatment of case, case assignment operates on the ARG-ST-list. The first version of the CASE PRINCIPLE states that:

- Initial NPs on the ARG-ST bearing a structural case get a CASE value of *snom* (*structural nominative*).
- Non-initial structural NPs are assigned *sacc* (*structural accusative*).

In further refinements, it is found out that this CASE PRINCIPLE should only to those elements that are realized locally (when they are not raised any further). This makes it possible to correctly treat the raising examples given below:

(6) It would be possible for him (*he) to be promoted.

(7) It was decided that he (*him) be promoted.

(8) Him (*He) being promoted made us all glad.

2.3.3 Extraction

One has to show, now, that the CASE PRINCIPLE can also deal with case assignment on extracted phrases.

Trace-based extraction: The case assignment works straightforwardly, by assigning the case to the trace locally; and then ensuring that the extracted element receives the same case as its trace.

Traceless extraction: For traceless extraction, it must be ensured that extracted elements are [REALIZED +], i.e. that they get assigned their case locally on the ARG-ST of the element where they are extracted from.

It is therefore necessary to insert appropriate constraints into the lexical rules or principles that are used in the respective extraction mechanism.

2.3.4 One further revision

One further revision to the language-universal Case Principle is necessary. The relevant example (from German) is the following:

(9)

Ich	*habe*	**den**	**Burschen**	*geraten,*	*im*	*Abstand*	*von*	*wenigen*	*Tagen*
I	have	[the	lads]$_{acc}$	advised,	in	interval	of	few	days

einer	**nachdem**	**anderen**	*zu*	*kündigen.*
[one	after	the	other]$_{nom}$	to	quit.

I advised the lads to hand in their notice one after the other at intervals of a few days.

The German phrase "einer nach dem anderen" (*one after the other*) always agrees with the subject. It can therefore be seen, that unrealized subjects in German are nominative, as the phrase stands in nominative in the above example.

However, controlled subjects are never realized, so they would be never subject to the Case Principle as stated above. It is not clear, then, how the phrase "einer nach dem anderen" gets assigned its nominative case in the example sentence.

Therefore, instead of the feature [realized $-/+$] that was used previously, a feature [raised $+/-$] is introduced. An element of arg-st is [raised $+$], if and only if there is a higher arg-st which contains both an that element, as well as a projection of this sign (i.e.: An element is raised iff itself and the element introducing it are both contained on some higher argument structure. It is then said to be raised to this higher argument structure.)

This new version of the Case Principle provides a uniform analysis of differently realized arguments. It assigns case correctly not only to those arguments that are directly realized in the position they are introduced at, but also to those that are extracted (the idiosyncratic treatment of extraction as it is explained above is not necessary any more), as well as elements that are never overtly realized, such as the controlled subjects.

These facts make the new treatment of case assignment general, modular, conceptually coherent and non-redundant (in the sense that the mechanism needs not be specified distributedly over the grammar).

3 Case in Polish

This section explains some case phenomena in Polish and how they can be treated using the above formalization.

Polish distinguishes 7 cases:

- nominative,
- accusative,
- genitive,
- dative,
- instrumental,
- locative, and
- vocative.

In this paper, locative and vocative will be throughout ignored.

3.1 Tests for case structurality

In Polish, nominative and accusative are generally seen as the structural cases, as well as the genitive of negation. The other cases (lexical genitive, dative and instrumental) are the lexical cases. There is also a lexical accusative, which only occurs on very few certain verbs.

[Przepiórkowski, 1999] developed a series of tests for the structurality of cases in Polish. The most important are:

- The structural case changes to genitive under negation. This holds for accusative, and in some cases maybe also for nominative subjects (of the existential copula). An example of the Genitive of Negation is:

  (10) *Lubie Marie.*
  like$_{1sg}$ Mary$_{acc}$.
  'I like Mary.'

  (11) *Nie lubie Marii (*Marie).*
  NM like$_{1sg}$ Mary$_{gen}$ (*Mary$_{acc}$).
  'I don't like Mary.'

- Structural NPs are realized as a genitive argument to nominalized verbs. This holds for the nominative subjects and accusative direct objects of declarative sentences. Arguments in other cases retain their case marking if they are realized as the argument of a nominalized verb.

  (12) *lubienie Marii$_{gen}$* (liking)

  (13) *uderzenie Marii$_{gen}$* (hitting)

- Polish has a word for "many", "dużo", which can only occur in structural positions. Positions where *dużo*-type phrases can occur are therefore necessarily structural.

  (14) *Dużo ludzi przyszło.*
  Many people came.
  Many people came.

  (15) *Janek pomaga [dużo osób / osobom]$_{dat}$. (Jan helps many people.)

  (16) Widziałem dużo osób. (I saw many people.)

  (17) Nie widziałem dużo osób. (I didn't see many people.)

  But "dużo"-phrases can not occur in ad-nominal positions:

  (18) *zobaczenie dużo osób ("seeing many people")

- NPs bearing structural case can license distributive *po*-phrases.

- Some such structural positions may be predicated by instrumental NPs.

Structural vs. Inherent Case – Results

The following table shows the results of the case structurality tests for Polish cases. A + indicates that this effect is an argument for structurality, a − is an argument against it. All cases for which the table contains at least one + should be considered structural cases.

	GoN	Nominalization	dużo	po	NP[*ins*]
nom	−	+	+	+	+
acc (standard)	+	+	+	+	+
gen. of negation	+	NA	+	+	+
other genitive	NA	NA	−	−	−
dat	−	−	−	−	−
ins	−	−	−	−	−
acc ("boleć")	+/−	NA	NA	NA	NA

4 Numeral Phrases in Polish

Case assignment in Slavic numeral phrases is complex and has been subject to discussions. For numerals that end in numbers higher than or equal to five, the declension pattern for Polish is shown in table 1.

	non-virile declension		virile declension	
	five	*women*	*five*	*guys*
NOM	pieć$_{?nom/?acc}$	kobiet$_{gen}$	pieciu$_{?nom/?acc/?gen}$	facetów$_{gen}$
ACC	pieć$_{acc}$	kobiet$_{gen}$	pieciu$_{?acc/?gen}$	facetów$_{gen}$
GEN	pieciu$_{gen}$	kobiet$_{gen}$	pieciu$_{gen}$	facetów$_{gen}$
DAT	pieciu$_{dat}$	kobietom$_{dat}$	pieciu$_{dat}$	facetom$_{dat}$
INS	pieciu/-ioma$_{ins}$	kobietami$_{ins}$	pieciu/-ioma$_{ins}$	facetami$_{ins}$
LOC	pieciu$_{loc}$	kobietach$_{loc}$	pieciu$_{loc}$	facetach$_{loc}$

Table 1: Declension of Polish numeral phrases.

The declension distinguishes between the virile (i.e., human masculine) and the non-virile (all others) pattern. For the nominative and accusative cases, it is not quite clear in the linguistic discussions as what case the numeral should be analysed.

4.1 Numeral Phrases – Problems

There are many problems with the phenomena belonging to numeral phrases. The two biggest problems can be exemplified with a sentence like:

(19) *Pieciu facetów weszło do kina.*
Five guys entered$_{3rd,sg,neut}$ into cinema.
Five guys entered the cinema.

The subject of the sentence (*five guys*) stands in accusative. And furthermore, the finite verb does not agree with the subject, although subject-agreement in person, number, and gender is otherwise always obligatory in Polish.

These two phenomena can only be accounted for together. I.e., one has to decide that the subjects are indeed accusative and not some strange form of nominative, because nominative subjects always agree with the verb. Thus, the fact that the verb occurs in a kind of "default" 3rd person neuter form can be explained by the fact that the subject is accusative.

4.2 Government-Agreement schema in numeral phrases

[Przepiórkowski, 1999] shows that numeral phrases in Polish exhibit a mixed government-agreement schema with their noun arguments, that resembles the government-agreement mixture of finite verbs. Thus, he concludes that the noun arguments of numerals must be their subjects. In that way, the numeral assigns accusative case on its subject, if it occurs in a structural position. In a lexical position, the numeral agrees with its noun argument in the lexical case.

4.3 Case Principle for Polish (final version)

The CASE PRINCIPLE must therefore be revised again. The final CASE PRINCIPLE for Polish is:

- Structural arguments of nonfinite verbs are caseless.
- Structural initial arguments of finite verbs, that are not numerical phrases, are assigned *snom*.
- Structural initial numerical arguments of finite verbs are assigned *sacc*.
- Structural non-initial arguments of non-negated verbs are assigned *sacc*.
- Structural non-initial arguments of negated verbs are assigned *sgen*.
- Structural arguments of prepositions are assigned *sacc*.

5 Numeral phrases in Russian

The case patterns for numeral phrases in Russian differ from those in Polish. However, as in Polish, one has to distinguish between so-called *paucal* numerals (numerals that end in 2, 3, or 4) and main numerals (that end in 5 or higher).

In oblique cases, the numeral always agrees with its noun in case. In structural positions, the declension is as depicted in table 2.

	feminine	masculine / neuter
paucal	genitive singular *tri dorógi* three roads	genitive singular *dva bráta* two brothers
main	genitive plural *sem' dorόg* seven roads	genitive plural *pjat' brat'ev* five brothers

Table 2: Declension of Russian numerals

5.1 Finite verb agreement

Other than in Polish, the finite verb does not stand in a "default" form after subjects that are numeral phrases:

(20) *Èti tri studenta priexali včera.*
These three students came-here yesterday.
These three students arrived yesterday.

(21) *Pjat' studentov sideli v komnate.*
Five students sat in room.
Five students were sitting in the room.

If the number is in focus, and the verb's meaning can be interpreted in an existential sense, the verb can also stand in the "default," 3rd person singular neuter form:

(22) *Segodnja v klasse prisutstvovalo*$_{3rd,sg,neu}$ *tol'ko dvenadcat'*$_{nom}$
Today in class be-present only twelve
malčikov$_{gen.pl}$.
boys.

Today there were only twelve boys present in class.

(23) *Itak, do Berlina ostalos'*$_{3rd,sg,neu}$ *tol'ko 896*$_{nom}$ *kilometrov*$_{gen.pl}$.
So, to Berlin remained only 896 kilometres.

And so, to Berlin there remained only 896 kilometres.

Another possibility is the partitive reading:

(24) *Včera priexalo*$_{3rd,sg,neu}$ *tri6*$_{nom}$ *studenta6*$_{gen.sg}$.
Yesterday came-here three students.

Three students arrived yesterday.

5.2 Attributes

Attributes modifying the noun in a numeral phrase can be ordered either before the numeral, or between the numeral and the noun. They stand usually in genitive plural:

(25) *tri molodýx*$_{gen.pl}$ *petuxá*$_{gen.sg}$
three young chickens

three young chickens

(26) *molodýx*$_{gen.pl}$ *pjat' studentov*$_{gen.pl}$
young five students

five young students

One exception are paucal numeral phrases with a feminine head noun. There, the attribute need not agree with the noun, it can rather stand in nominative plural.

(27) *Ja verju, čto pridët den', kogda dve*$_{nom}$ *naši*$_{nom.pl}$ *velikie*$_{nom.pl}$
I believe, that arrives day, when two our great
strany$_{gen.sg}$ *budut doverjat' drug drugu.*
countries will trust each other.

I believe that the day will come, when our two great countries will trust each other.

Some speakers even judge the nominative grammatical for some masculine or neuter noun modifiers:

(28) *Pri ètom zabyvajutsja dva*$_{nom}$ *važnye*$_{nom.pl}$ *obstojatel'stva*$_{gen.sg}$.
By this are-forgotten two important circumstances.

Thereby, two important circumstances are forgotten.

Often, genitive singular and nominative plural of feminine nouns are just distinguished by the stress: compare rúki – rukí(*hand*). This makes these forms almost syncretic, and it is sometimes even judged grammatical (at least by some speakers) to use the nominative plural also for the noun.

6 Treatment of numeral phrases in Russian with HPSG

The general treatment of Russian numeral phrases can be taken over from [Przepiórkowski, 1999]. That is, the distinction between paucal and main numerals through different lexical entries will be retained here. Furthermore, the classification of numerals as the heads of numeral phrases and of the noun arguments as their subjects are also both valid for Russian.

However, other than in Polish, the finite verb is normally required to agree with the subject, even if this subject is a numeral phrase. Is is only in some (semantically determined) cases that a "default" form can be used for the finite verb. For these cases, it would maybe suffice to introduce additional lexical entries for those verbs that can have an existential reading, and specify that they always occur in the default form, regardless of their subject.

Attributes

The case pattern of attributes is quite interesting.

For modifiers that stand *before* the numeral, an elegant account is possible: The modifier can either be extracted from the NP, and then combine with the NP after the numeral has combined with the noun. The extraction case can be handled in the same way as proposed in [Przepiórkowski, 1999]: the modifier is assigned case locally on the argument structure from which it is extracted. It will therefore agree with the noun and be assigned genitive singular.

However, one could allow the modifier to alternatively combine with the whole numeral phrase and modify it (as numerals are also nouns, this should be possible). Then it agrees with the whole NP, which stands in nominative plural.

It is not quite as clear, however, how modifiers inside of the NP can be assigned nominative plural. The agreement with the head noun is of course straightforward. The only possibility I see right now would be to allow a flat structure of the NP. Then, the numeral would be the head (in this *numeral phrase*, as it argued at length in [Przepiórkowski, 1999]). The head noun is the subject of the numeral, and is assigned genitive singular. If the head noun is feminine, the numeral can then optionally assign nominative plural on all other elements of its argument structure.

7 Conclusion

In this paper I have discussed some problems concerning case assignment in Slavic numeral phrases.

I have first presented the work of Przepiórkowski on the treatment of case in Polish in the framework of HPSG. I concentrated on his discussion of Polish numeral phrases.

Then, after giving a short account on some interesting data for numeral phrases in Russian (with a focus on those facts that differ from their counterparts in Polish), I have discussed how they can be analysed using an approach similar to Przepiórkowski's. The formalization, unfortunately, stays for future work.

References

Dina B. Crockett. *Agreement in Contemporary Standard Russian*. Slavica Publishers, Inc., 1976.

W. Heinz and J. Matiasek. Argument structure and case assignment in German. In J. Nerbonne, K. Netter, and C. Pollard, editors, *German in Head-Driven Phrase Structure Grammar*. CSLI Publications, Stanford, 1994.

R. O. Jacobson. *Selected Writings II*, chapter Beitrag zur allgemeinen Kasuslehre. Gesamtbedeutungen der russischen Kasus, pages 23–71. Mouton, The Hague, 1971.

Herbert Mulisch et al. *Die russische Sprach der Gegenwart*, volume 2. Morphologie. VEB Verlag Enzyklopädie Leipzig, 1985.

Adam Przepiórkowski. *Case Assignment and the Complement/Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. PhD thesis, Tübingen University, 1999.

B. O. Unbegaun. *Russische Grammatik*. Vandenhoeck & Ruprecht, Göttingen, 1969.