6

Case Assignment in Polish: Towards an HPSG Analysis

Adam Przepiórkowski

Edinburgh Working Papers in Cognitive Science, Vol. 12: *Studies in HPSG*, pp. 191–228.
C. Grover and E. Vallduví, eds.

Abstract

We formulate within Head-driven Phrase Structure Grammar a theory of case which is based on the structural/lexical case dichotomy as developed within the Government and Binding framework. We argue that this dichotomy holds for Polish, a relatively inflectional Slavic language, and give examples of tests which allow to determine which morphological cases belong to which syntactical (i.e., structural or lexical) cases. On the basis of these observations we present a Case Principle for Polish responsible for morphological realization of structural cases. In the second part of the paper we give an account of the infamous problem of Slavic numerals, concentrating however solely on Polish. We show that the distribution and inflection of numeral phrases confirms the case dichotomy in a striking way. We come up with HPSG lexical entries for all main classes of Polish numerals, including the paucal and the quirky indefinite numerals. In the last section of this paper we argue that in Polish, unlike in German and Russian, passivization does not seem to depend on the structural/lexical case dichotomy.

1 Introduction

In this paper, we will present a basic theory of case for Polish. This theory, developed in the framework of Head-driven Phrase Structure Grammar, is deeply embedded in the tradition of generative grammar (in the broad sense of the term) and it borrows freely from Chomsky's Goverment and Binding Theory.

However, we will not deal here with the relationship between meaning and case (if any); our approach will be — to use the terminology of Mel'čuk (1986) — syntagmatically (rather than paradigmatically) oriented. We will also not succumb to the temptation of defining the notion of case. This (by no means trivial!) theoretical task is well outwith the scope of this paper.[1]

1.1 A Historical Note

It will not be an exaggeration to say that everything that has been written about case within the framework of HPSG stands in a strong relationship with the Case Theory of Government and Binding (GB). For Pollard and Sag (1994) this relation is wholesale rejection, while for Heinz and Matiasek (1994) it is development. This subsection will present some of the most prominent assumptions regarding case which exist within GB.

In GB, Case[2] is mainly a structural phenomenon: the Case an NP receives in a sentence depends on its position in the derivational tree of this sentence. Thus, as shown in figure 1, complements of a verb get the accusative Case, while the subject receives the nominative Case.

This is explained by the fact that verbs which govern (i.e., are close to, in a certain configurational sense) their complements assign the accusative, while INFL nodes (representing a bundle of features related to tense and agreement), which govern subjects, assign the nominative Case.

[1] See, however, Mel'čuk (1986) and Comrie (1986) for some attempts.

[2] GB distingushes between morphological case (written with the small 'c') and abstract Case (capital 'C'). Languages differ in the extent to which they exhibit case, but they are all assumed to have abstract Case.

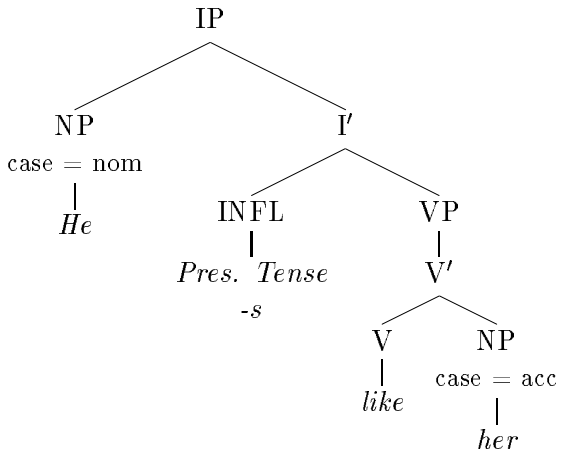

Figure 1: Derivational tree of *He likes her*

This theory works quite well for English — it neatly explains some syntactic phenomena, in particular the phenomenon of passivization exemplified below.

(1) He_{nom} likes her_{acc}. $\stackrel{pass}{\longrightarrow}$ She_{nom} is liked.

According to GB, passivization is — simplifying a little — nothing more than adding the passive morpheme (*-en*) to the verb. This morpheme, in turn, has some very peculiar properties: it *absorbs* both the case assigned by the verb in question, and the so-called external theta role. The latter property simply means that thus affected verb does not sanction a subject, while the former that the complement of the verb does not receive Case. However, according to another principle of GB, the Case Filter, every (overt) NP *must* be assigned Case. Hence, in order to get Case, the complement has to move to a vacant position (leaving a trace behind) where Case can be assigned. Such a position is here the subject position which, according to the Extended Projection Principle, has to be present in the tree even if there is no subject (as in the case of passive verbs). But the Case it receives is not accusative anymore; as the complement is now governed by INFL, it receives the nominative Case. This is illustrated by figure 2.

Even though the Case Theory of GB might have worked well for English, it soon became clear that this purely configurational notion of Case cannot provide explanation for, e.g., more sophisticated course of passivization in German. The problem German poses is that passivization in this language affects Case assignment only in some instances. As the reader can easily verify, GB's Case Theory outlined above correctly accounts for (2), but fails in (3) below.[3] In the latter example *ihm* clearly retains its dative Case assigned by the verb.

(2) a. *Sie sieht ihn.*
She_{nom} sees him_{acc}.

[3]These examples come from Haider (1985), cited here after Haegeman (1991).

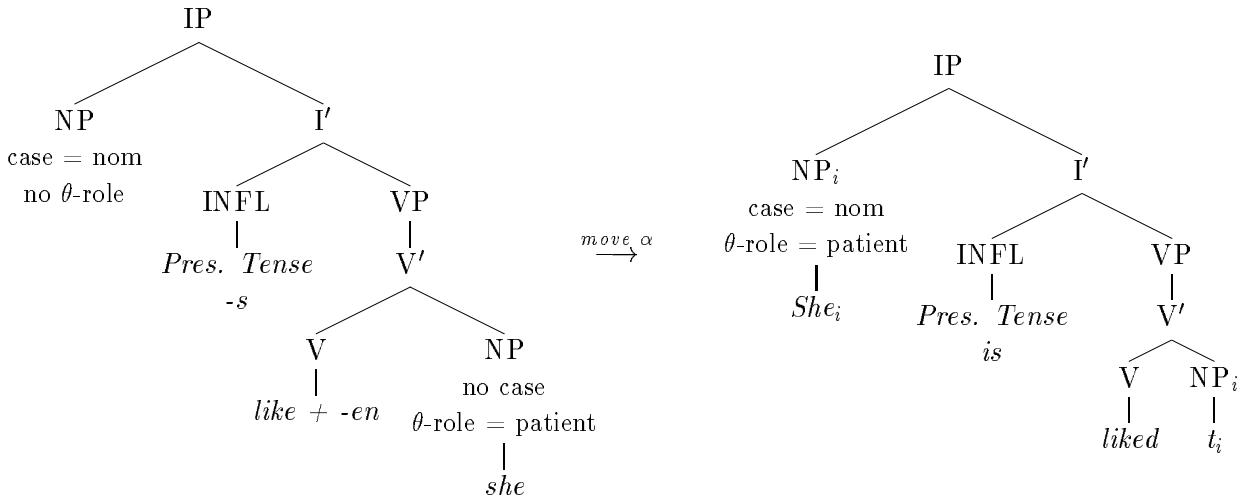

Figure 2: Passivization in GB

'She sees him.'

b. *Er wird gesehen.*
He$_{nom}$ is seen.
'He is seen.'

c. * *Ihn wird gesehen.*
He$_{acc}$ is seen.
'He is seen.'

(3) a. *Sie hilft ihm.*
She$_{nom}$ helps him$_{dat}$.
'She helps him.'

b. * *Er wird geholfen.*
He$_{nom}$ is helped.
'He is helped.'

c. *Ihm wird geholfen.*
He$_{dat}$ is helped.
'He is helped.'

These, and many other Case-related problems with GB's account (also in English, e.g., the problem of the so-called *of*-insertion) lead to substantial changes in the Case Theory and, as a result, Chomsky (1986b) distinguishes between two types of Case assignment: structural, based — as before — on the position of NP at S-structure, and inherent, assigned by the lexical element at D-structure. Unlike structural Case, inherent Case is characterized by its stability: its morphological realization does not change with syntactic environment. Dative case in German (as well as, e.g., genitive in English) is analyzed as an instance of inherent

case, though the extent to which a particular language realizes inherent case is a parameter of the theory; languages can differ significantly in this respect.

Many languages with rich declensional paradigms support this Case dichotomy in interesting ways. We will give examples of such supporting phenomena in the following sections.

1.2 Case in HPSG

There has been no separate theory of case within the framework of HPSG until very recently. Pollard and Sag (1994) put considerable effort into rejecting GB's Case Theory and, by doing so, they announce that no theory of case (or Case) whatsoever is necessary: case is treated in HPSG only as a part of subcategorization requirements. Thus, for example, the value of the SUBCAT feature of the verb *like* would be:

(4) like: ⟨NP[*nom*], NP[*acc*]⟩

In this approach phenomena such as passivization are assigned to the lexicon. More specifically, the Passive Lexical Rule takes care of permuting the complements within the SUBCAT list (cf. Pollard and Sag (1987), p. 215 and Pollard and Sag (1994), p. 121) and of changing the CASE values.

However, as Heinz and Matiasek (1994) notice, "this approach fits well with arguments exhibiting the same case in all syntactic constructions... but makes it difficult to cope with complements showing variations of case depending on the syntactic context." Such variations exist in German and — appropriately enough — were first (within HPSG) taken into consideration in Nerbonne *et al.* (1994) by Pollard (1994) and, especially, Heinz and Matiasek (1994). In this section, we will present some of the most prominent assumptions of Heinz and Matiasek concerning case; the reader is, however, referred to their article for details and some applications of their theory to German.

Heinz and Matiasek (1994), following Haider (1985), posit two kinds of case: lexical (an analogue of GB's inherent case) and structural. The morphological form of the former is determined via subcategorization requirements. This means that whenever a verb requires its complement to bear a lexical case, it also specifies the morphological realization of this case. Thus, the morphological case form of such a complement cannot vary with the syntactic environment.

Alternatively, a verb (or any other lexical element) can specify its complement as bearing a structural case. In this instance, the morphological form of the complement depends upon (and may alternate with) its syntactic environment, rather than on the lexical item (the verb) in question.[4]

According to Heinz and Matiasek (1994), nominalization is an example of a test checking whether a given complement of a verb has lexical or structural case.[5] This can be illustrated by the following examples:

[4] This lexical/structural case dichotomy parallels, of course, GB's inherent/structural case distinction.

[5] However, as Johannes Matiasek points out (pc.), nominalization in German is a more complex matter than Heinz and Matiasek (1994) would suggest.

(5) a. *Der Mann hilft dem Installateur.*
The man$_{nom}$ helps the plumber$_{dat}$.
'The man is helping the plumber.'

b. *das Helfen dem Installateur*
the helping the plumber$_{dat}$
'the help for/*from the plumber'

c. *das Helfen des Installateurs*
the helping the plumber$_{gen}$
'the help from/*for the plumber'

(6) a. *Der Mann unterstützt den Installateur.*
The man$_{nom}$ helps the plumber$_{acc}$.
'The man is helping the plumber.'

b. *das Unterstützen des Installateurs*
the helping the plumber$_{gen}$
'the help for/from the plumber'

Example (5) shows that the verb *helfen* requires a dative complement. The case of this complement does not change under nominalization, it remains dative. This should be contrasted with the behaviour of the complement of *unterstützen*; the case of this complement changes from accusative to genitive in the process (see (6)). Moreover, the case of the subject changes from nominative to genitive in both cases. These data suggest that dative is a lexical case, while nominative, genitive and accusative are structural: their distribution is a matter of configurational rules, rather than inherent properties of lexical items.

Of course, this does not mean that a given morphological case can only be either lexical or structural. Heinz and Matiasek (1994), developing upon Haider (1985), come up with a type lattice (see figure 3) depicting which morphological cases can be instances of which syntactic (i.e., lexical or structural) cases in German.

In this type lattice we follow the convention used in Heinz and Matiasek (1994) of writing the most general type at the top of the lattice.[6] So, the type *case* has as its subtypes *morph-case* and *syn-case*. The former determines the morhological cases German allows, i.e., its subtypes are *nom* (nominative), *gen* (genitive), *dat* (dative) and *acc* (accusative). On the other hand, *syn-case* determines the syntactic properties of cases; each case marking on an NP is either *structural* or *lexical*. Structural cases are nominative (*snom*), genitive (*sgen*) or accusative (*sacc*), while lexical cases are genitive (*lgen*), dative (*ldat*) and accusative (*lacc*). This, of course, means, that NPs bearing genitive or accusative morphological case are either *structural* or *lexical*, depending on the case assigner. In the sequel of this paper we will usually abbreviate *structural* to *str* and *lexical* to *lex*.

Heinz and Matiasek (1994) do not have much to say about lexical case: its morphological realization is entirely determined within the SUBCAT list, so it does not require a separate

[6] Note that lattice of figure 3 is not a correct inheritance hierarchy in the sense of Carpenter (1992); it is not a bounded complete partial order. When presenting the case lattice for Polish we will do so using both conventions: that of Heinz and Matiasek (1994) and that of Carpenter (1992).

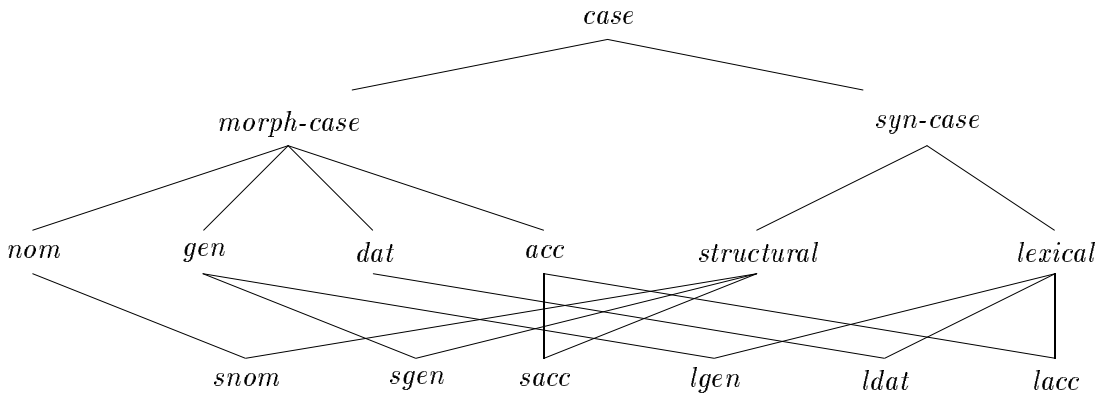

Figure 3: Case hierarchy for German

theory. On the other hand, it is assumed that lexical items do not specify the exact morphological realization of the NP[*str*] they subcategorize for. For example, verbs do not specify their subjects as NP[*nom*], but rather as NP[*str*]. It is the Case Principle, whose scope is somewhat analogous to GB's Case Theory, that determines the exact morphological realization of structural cases. Thus, for instance, values of SUBCAT for *helfen* and *unterstützen* look as follows:

(7) a. helfen: ⟨NP[*str*], NP[*ldat*]⟩

b. unterstützen: ⟨NP[*str*], NP[*str*]⟩

What morphological value a given instance of structural case gets is decided by Case Principle which, in short, says that an NP[*str*] subject of a verb receives nominative (i.e., *snom*) case, structural complements of verbs receive accusative case (*sacc*), while structural complements of nouns receive genitive case (*sgen*). More precisely (cf. Heinz and Matiasek (1994), p. 34), the following constraints have to be present in German grammar:[7]

[7]These are constraints in the sense of Pollard and Sag (1987). Of course, constraints in the sense of Carpenter (1992) would have to be imposed upon types (type *phrase* in this case), not feature structures.

(8)
$$\left[\begin{array}{l} \textit{phrase} \\ \textsc{synsem} \mid \textsc{loc} \mid \textsc{cat} \quad \left[\begin{array}{l} \textit{cat} \\ \textsc{head} \left[\begin{array}{l} \textit{verb} \\ \textsc{vform}\ \textit{fin} \end{array}\right] \\ \textsc{subcat}\ \langle\ \rangle \end{array}\right] \\ \textsc{dtrs} \quad \left[\begin{array}{l} \textit{h-c-str} \\ \textsc{head-dtr} \mid \ldots \mid \textsc{subcat}\ \langle\ \textit{NP}[\textsc{str}\textit{}],\ \ldots \rangle \end{array}\right] \end{array}\right]$$
$$\Longrightarrow \left[\,\textsc{dtrs} \mid \textsc{head-dtr} \mid \ldots \mid \textsc{subcat}\ \langle\ \textit{NP}[\textsc{nom}\textit{}],\ \ldots \rangle\right]$$

(9)
$$\left[\begin{array}{l} \textit{phrase} \\ \textsc{synsem} \mid \textsc{loc} \mid \textsc{cat} \quad \left[\begin{array}{l} \textit{cat} \\ \textsc{head}\ \textit{verb} \\ \textsc{subcat}\ \langle\ \rangle \vee \langle\ \textit{synsem}\ \rangle \end{array}\right] \\ \textsc{dtrs} \quad \left[\begin{array}{l} \textit{h-c-str} \\ \textsc{head-dtr} \mid \ldots \mid \textsc{subcat}\ \langle\ \textit{synsem},\ \textit{NP}[\textsc{str}\textit{}],\ \ldots \rangle \end{array}\right] \end{array}\right]$$
$$\Longrightarrow \left[\,\textsc{dtrs} \mid \textsc{head-dtr} \mid \ldots \mid \textsc{subcat}\ \langle\ \textit{synsem},\ \textit{NP}[\textsc{acc}\textit{}],\ \ldots \rangle\right]$$

(10)
$$\left[\begin{array}{l} \textit{phrase} \\ \textsc{synsem} \mid \textsc{loc} \mid \textsc{cat} \quad \left[\begin{array}{l} \textit{cat} \\ \textsc{head}\ \textit{noun} \\ \textsc{subcat}\ \langle\ \rangle \vee \langle\ \textit{synsem}\ \rangle \end{array}\right] \\ \textsc{dtrs} \quad \left[\begin{array}{l} \textit{h-c-str} \\ \textsc{head-dtr} \mid \ldots \mid \textsc{subcat}\ \langle\ \textit{synsem},\ \textit{NP}[\textsc{str}\textit{}],\ \ldots \rangle \end{array}\right] \end{array}\right]$$
$$\Longrightarrow \left[\,\textsc{dtrs} \mid \textsc{head-dtr} \mid \ldots \mid \textsc{subcat}\ \langle\ \textit{synsem},\ \textit{NP}[\textsc{gen}\textit{}],\ \ldots \rangle\right]$$

The reader will recall that the symbol '$\Longrightarrow$' is used to denote *constraints*, i.e., *conditional feature structures* (cf. Pollard and Sag (1987), p. 43). Thus, for example, (8) should be understood as a condition imposed upon every feature structure of type *phrase*, modelling a finite verb phrase (VFORM *fin*) whose DTRS value is a structure of *head-complement-structure* type, and whose head-daughter specifies its first complement as NP[*str*]. This conditions says, that the first complement of the head-daughter of such a phrase has to bear the nominative case (*snom*).

In the subsequent sections we will follow Heinz and Matiasek (1994) in abbreviating conditional feature structures such as the ones above:

Case Principle
In a *head-complement-structure* of category

- **verb[fin]**: the structural subject has a CASE value of *snom*,
- **verb**: the structural object has a CASE value of *sacc*,
- **noun**: the structural object has a CASE value of *sgen*.

These are the only saturated or almost saturated *head-complement-structures* with structural arguments.[8]

Note, that, given the SUBCATs of *helfen* and *unterstützen* as in (7), this formulation of Case Principle correctly predicts the nominalization facts shown in the examples (5) and (6) on page 196. Of course, the underlying assumption here is that the only change which nominalization (implemented for example as a lexical rule) brings to the SUBCAT list is making each argument optional and possibly specifying that at most one argument is present.

In the remainder of this article we will examine to what extent the lexical vs. structural case dichotomy can be argued for in Polish. We will also attempt to formulate an analogous Case Principle for this language.

2 Structural vs. Lexical Case in Polish

Polish, a language with rich inflectional morphology, shows considerable support for the lexical vs. structural case distinction. The only attempt of formal analysis (in the framework of GB) of how this distinction functions in Polish that we know of is Willim (1990).[9] Our analysis will differ from (the translation into HPSG of) that of Willim in many respects.

2.1 Morphological Case in Polish

There are seven morphological cases in Polish, though vocative can be argued not to be a case in the strict sense: it is used in isolation, mainly for getting attention and for addressing.[10] Of the remaining six, nominative never appears outside sentential subject position,[11] accusative is realized by verbal and prepositional objects, genitive, dative and instrumental occur as arguments of all main lexical categories, and locative is restricted to the prepositional arguments.

[8] Structural subject should be understood as an NP element of the SUBJECT list (in the sense of Pollard and Sag (1994), chapter 9) if it (is present and) bears *str* case. Similarly, by structural object we mean any structural NP element of COMPS. We find these notions more intuitive than, respectively, *external argument* and *internal argument* inherited from GB and used by Heinz and Matiasek (1994). See, however, their article for some motivation for this nomenclature.

[9] Some work has been done on analysis of case dichotomy in other Slavic languages, mainly Russian. The reader is referred to Babby (1986), Franks (1986), Franks (1990), Franks (1994) and references cited therein.

[10] See a.o. Willim (1990), Polański (1993) (p. 578, entry for *vocativus*) and Strutyński (1993) for arguments for this position, but also Saloni and Świdziński (1985) p. 137 for important arguments against it.

[11] This statement, as it stands, is too strong; cf. examples like (i) (pointed to us by Bob Borsley) or (ii) (from Saloni and Świdziński (1985), p. 118) below:

(i) *Jan, rozmawiałem z nim.*
John$_{nom}$, talked$_{1st,sg}$ with him$_{ins}$.
'John, I talked to him.'

(ii) *Przyjaciele wołają go Grubas.*
Friends$_{nom}$ call him$_{acc}$ Fat$_{nom}$.
'The friends call him Fatty.'

2.2 Nominalization

As far as nominalization is concerned, Polish parallels German.[12] Consider the following data:

(11) a. *Janek pomaga Tomkowi.*
John$_{nom}$ helps Tom$_{dat}$.
'John is helping Tom.'

b. *pomaganie Tomkowi*
helping Tom$_{dat}$
'the help for/*from Tom'

c. *pomaganie Janka*
helping John$_{gen}$
'the help from/*for John'

(12) a. *Janek pogardza Tomkiem*
John$_{nom}$ scorns Tom$_{ins}$.
'John scorns Tom.'

b. *pogardzanie Tomkiem*
scorning Tom$_{ins}$
'the scorn for/*from Tom'

c. *pogardzanie Janka*
scorning John$_{gen}$
'the scorn from/*for John'

(13) a. *Janek wspiera Marię*
John$_{nom}$ supports Mary$_{acc}$.
'John is supporting Mary.'

b. *wspieranie Marii (Janka)*
helping Mary$_{gen}$ (John$_{gen}$)
'the help for/from Mary (John)'

Examples (11) and (12) suggest that dative and instrumental cases are here instances of lexical case: they do not change under nominalization. On the other hand, as (13) shows, accusative is structural: the case changes to genitive in the process of nominalization. Of course, nominative and genitive are also structural cases here, just like in German.

[12] Cf. section 4.3 where we actually argue against this statement.

On the basis of the above observations we can postulate the first version of Case Principle for Polish:

Case Principle (First Version)
In a *head-complement-structure* of category

- **verb**: the structural subject has a CASE value of *snom*, the structural object has a CASE value of *sacc*,
- **noun**: any structural argument (subject or object) has a CASE value of *sgen*.

2.3 Objects of Prepositions

It should be emphasized that the foregoing remarks are not relevant for those arguments which are prepositional phrases:[13]

(14) a. *Janek czeka na Marię.*
John$_{nom}$ waits on Mary$_{acc}$.
'John is waiting for Mary.'

b. *czekanie na Marię*
waiting on Mary$_{acc}$
'the waiting for Mary'

As the above example shows, accusative NPs which are arguments of prepositions do not change their case under nominalization.

Heinz and Matiasek analyze prepositions devoid of their inherent (locational) meaning as 'markers'. For them, prepositional phrases are just 'marked' nominal phrases, i.e., NP[+marked]. Hence, Case Principle can be applied to prepositional phrases. This, in turn, means that the accusative complement in example (14) cannot be structural: if it were structural then — according to Case Principle — it would have to change into genitive in the process of nominalization.

On the basis of analogous observations for German, Heinz and Matiasek (1994) posit that prepositional (i.e., 'marked' in their terminology) arguments exhibiting accusative case are instances of lexical accusative (and exactly for this reason such 'marked' NPs do not change case under nominalization). Note that this failure of 'marked' (i.e., prepositional) arguments to change case under nominalization seems to be one of the main reasons for allowing lexical accusatives; all other (i.e., 'unmarked') occurences of accusative phrases are structural. Hence, Heinz and Matiasek miss the generalization that all 'unmarked' (i.e., true NP) accusative phrases are structural, while all 'marked' (i.e., prepositional) accusative phrases are lexical.

On our account there are no such coincidences: we do not follow Heinz and Matiasek (1994) in analyzing prepositional phrases as 'marked' nominal phrases, but rather, traditionally, as

[13] Neither are they relevant for adverbial modifiers but here facts are less clear-cut. Unfortunately, discussion of case assignment to adverbial modifiers is outwith the scope of this paper.

true prepositional phrases. This treatment of prepositional phrases explains the failure of prepositional accusative arguments to become genitive under nominalization (cf. (14a) and (14b)) — according to the Case Principle only *nominal* structural phrases alternate with environment, not prepositional phrases.[14]

We will also postulate that accusative case is always structural in Polish. This position is a consequence of the observation that there are no verbs in Polish subcategorizing for a *stable* accusative (or rather, structural) complement; accusative complements always become genitive under nominalization (and under genitive of negation, see below).[15] Of course, since prepositional arguments can bear accusative case, and we assume (cf. page 197) that lexical items never specify the morphological case of their structural complements, we have to add one more clause to the Case Principle:

Case Principle (Second Version)
In a *head-complement-structure* of category

...	...
• **noun**:	any structural argument (subject or object) has a CASE value of *sgen*,
• **preposition**:	the structural object has a CASE value of *sacc*.

We invite the reader to check that thus revised Case Principle and the lexical entry for the preposition *na* given (partially) below account for the example (14).

(15)
$$\left[\begin{array}{ll} \textit{word} & \\ \text{PHON} & \langle \textit{na} \rangle \\ \text{SYNSEM|LOC|CAT} & \left[\begin{array}{ll} \textit{category} & \\ \text{HEAD} & \left[\begin{array}{ll} \textit{prep} & \\ \text{PFORM} & \textit{'na'+str} \end{array}\right] \\ \text{SUBCAT} & \langle \textit{NP[str]} \rangle \end{array}\right] \end{array}\right]$$

The next section shows that prepositional arguments are also not affected by Genitive of Negation.

2.4 Genitive of Negation

Another phenomenon of case variation is the so-called Genitive of Negation (GoN): an accusative object of a verb appearing in a declarative sentence changes its case marking to genitive under sentential negation. This is illustrated by the following example:

(16) a. *Janek lubi Marię.*
John$_{nom}$ likes Mary$_{acc}$.
'John likes Mary.'

b. *Janek* ***nie*** *lubi Marii.*
John$_{nom}$ not likes Mary$_{gen}$.
'John doesn't like Mary.'

[14] Not much else hinges on this decision, though.
[15] See section 4 for other arguments for the structurality of the accusative of prepositional arguments.

GoN does not exist in German but it is widespread in Slavic and exists also in some other languages (e.g., Finnish). GoN is a very unstable phenomenon: in many Slavic languages the accusative case expands rapidly taking the place of genitive in many constructions, not least in sentential negation. For example, in Czech GoN has practically ceased to exist; only the older generations use it in some restricted environments. In Russian, on the other hand, both accusative and genitive are allowed under sentential negation[16], while in Polish, even though accusative replaces genitive in many syntactic environments, genitive remains the only possibility under sentential negation.[17]

It is worth noticing that Genitive of Negation, just as nominalization, does not affect dative and instrumental complements. The examples below should be compared with (11)–(13) above:

(17) a. *Janek pomaga Tomkowi.*
John$_{nom}$ helps Tom$_{dat}$.
'John is helping Tom.'

b. *Janek* ***nie*** *pomaga Tomkowi.*
John$_{nom}$ not helps Tom$_{dat}$.
'John is not helping Tom.'

(18) a. *Janek pogardza Tomkiem.*
John$_{nom}$ scorns Tom$_{ins}$.
'John scorns Tom.'

b. *Janek* ***nie*** *pogardza Tomkiem.*
John$_{nom}$ not scorns Tom$_{ins}$.
'John doesn't scorn Tom.'

(19) a. *Janek wspiera Marię.*
John$_{nom}$ supports Mary$_{acc}$.
'John is supporting Mary.'

b. *Janek* ***nie*** *wspiera Marii.*
John$_{nom}$ not supports Mary$_{gen}$.
'John is not supporting Mary.'

[16] The reader is referred to Timberlake (1986) for an analysis of distribution of accusative and genitive under negation and for defence of the hypothesis that GoN is in the state of withdrawal in Russian.

[17] Actually, this rule has a few exceptions. Buttler *et al.* (1971) give two conditions when accusative is allowed. The first is semantical in nature: accusative is allowed when the sentence has a positive meaning despite its apparent negation. The second, which is structural, says that accusative is allowed when the complement is "far" from the finite verb. We do not try to model these exceptions in this paper.

Also prepositional arguments are not affected by negation. Again, the examples below parallel (14):

(20) a. *Janek czeka na Marię.*
John$_{nom}$ waits on Mary$_{acc}$.
'John is waiting for Mary.'

b. *Janek* ***nie*** *czeka na Marię.*
John$_{nom}$ not waits on Mary$_{acc}$.
'John is not waiting for Mary.'

These data independently confirm the distinction between structural and lexical case made in the previous section.[18] They also call for splitting one of the clauses of Case Principle:

Case Principle (Third Version)
In a *head-complement-structure* of category

...	...
• **verb**:	the structural subject has a CASE value of *snom*,
• **verb[−neg]**:	the structural object has a CASE value of *sacc*,
• **verb[+neg]**:	the structural object has a CASE value of *sgen*,
...	...

2.5 Case Lattice for Polish

On the basis of the above examples we are able to postulate the case type hierarchy for the Polish case system (Figure 4). In this hierarchy we arbitrarily assume that locative is an instance of the lexical case; such an assumption simplifies the Case Principle. We also assume that — just as in German — genitive can be either structural or lexical. We will present arguments for this stance in section 4.

[18] While avoiding any specific analysis of negation here, we assume an existance of the binary attribute NEG appropiate at least for *verb*. The value of this attribute is '+' if the verb in question is in the scope of negation, or '−' otherwise. A careful account of negation is needed in order to treat examples such as (ii) below (pointed to us by Bob Borsley), where the +NEG value seems to be shared between the matrix verb and its VP[inf] complement.

(i) *Jan chiał widzieć Marię.*
John$_{nom}$ wanted see$_{inf}$ Mary$_{acc}$.
'John wanted to see Mary.'

(ii) *Jan nie chciał widzieć Marii.*
John$_{nom}$ not wanted see$_{inf}$ Mary$_{gen}$.
'John didn't want to see Mary.'

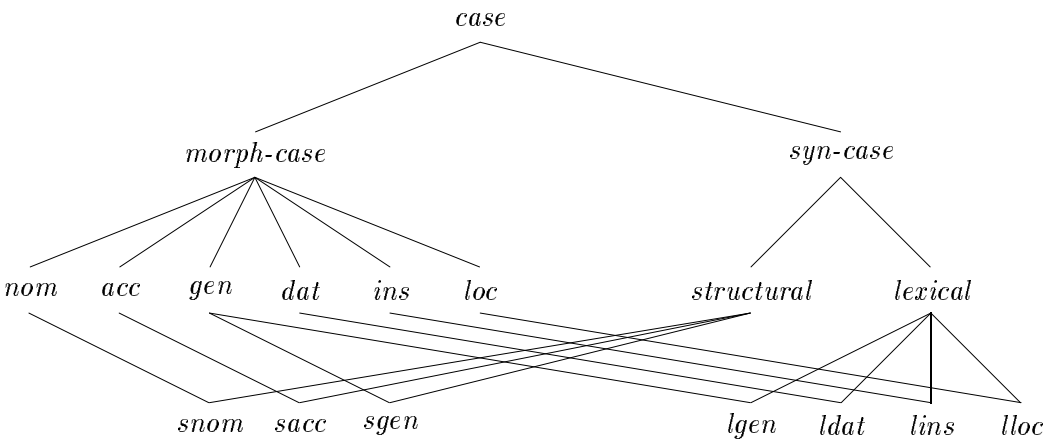

Figure 4: Case hierarchy for Polish

As we have mentioned above, a type hierarchy such as the one in figure 4 is not an inheritance hierarchy in the sense of Carpenter (1992).[19] For example, the consistent types *structural* and *morph-case* have no least upper bound. Figure 5 presents an analogous case hierarchy which is a correct inheritance hierarchy.

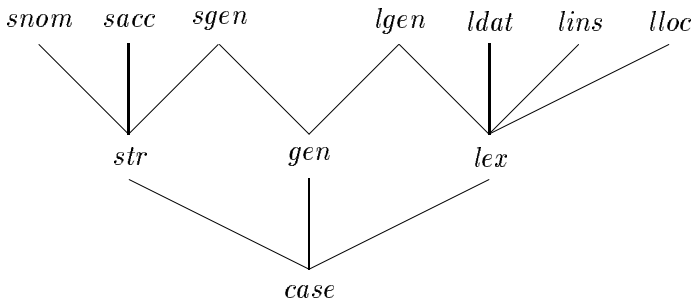

Figure 5: Correct case hierarchy for Polish

In the next sections we will try to analyze within the framework established so far some more ephemeral and idiosyncratic issues concerning case assignment in Polish, namely those of the case of numerals (section 3) and so-called indefinite numerals (section 4). We will also make a few remarks on passivization (section 5).

[19] It is not a BCPO, see Carpenter (1992).

3 Numerals

The complexity of numerals in Slavic languages is really daunting; Polish is no exception here. In general, this complexity is thought to be caused by the transitional character of numerals in Polish.[20] We will try not to forget about these diachronic considerations in what follows. However, our account will be mainly synchronical; we will attempt to analyze the phenomena involved as they stand.[21]

3.1 Basic Facts

In this subsection, we will deal with the most typical relationship basic numerals establish with noun phrases.

3.1.1 There are no nominative numeral phrases!

3.1.1.1 Initial assumptions Consider the declension patterns shown in (21) and (22) below. This is how such patterns should be understood: the NOM row contains these forms which can appear as subjects of typical verbs (such as *jeść*, 'eat', or *lubić*, 'like'); the ACC row contains these forms which can appear in an object position of typical transitive verbs (such as *lubić*, 'like'); the DAT row contains these forms which can appear as second objects of typical ditransitive verbs (such as *dać*, 'give'), etc. It is important to bear this point in mind as we will analyze some sentential subjects (i.e., forms appearing in the NOM row) as *accusative* phrases.[22]

(21) Non-masculine-human declension:

	these	*five*	*women*
NOM	te$_{nom/acc}$/tych$_{gen}$	pięć$_{nom/acc}$	kobiet$_{gen}$
GEN	tych$_{gen}$	pięciu$_{gen}$	kobiet$_{gen}$
DAT	tym$_{dat}$	pięciu$_{dat}$	kobietom$_{dat}$
ACC	te$_{nom/acc}$/tych$_{gen}$	pięć$_{nom/acc}$	kobiet$_{gen}$
INS	tymi$_{ins}$	pięcioma$_{ins}$	kobietami$_{ins}$
LOC	tych$_{loc}$	pięciu$_{loc}$	kobietach$_{loc}$

[20] See for example Buttler *et al.* (1971) for a short assessment of the changes the system of numerals is undergoing currently.

[21] For some analyses of numerals in other Slavonic languages the reader is referred to Corbett (1978) and Franks (1994) and references cited therein.

[22] Hence, there is no contradiction in a pattern containing the following row (cf. (29)):

(i) NOM tych$_{gen/acc}$ pięciu$_{acc}$ mężczyzn$_{gen}$

(22) Masculine-human declension:

	these	*five*	*men*
NOM	tych$_{gen/acc}$	pięciu$_{nom/gen/acc}$	mężczyzn$_{gen/acc}$
GEN	tych$_{gen}$	pięciu$_{gen}$	mężczyzn$_{gen}$
DAT	tym$_{dat}$	pięciu$_{dat}$	mężczyznom$_{dat}$
ACC	tych$_{gen/acc}$	pięciu$_{gen/acc}$	mężczyzn$_{gen/acc}$
INS	tymi$_{ins}$	pięcioma$_{ins}$	mężczyznami$_{ins}$
LOC	tych$_{loc}$	pięciu$_{loc}$	mężczyznach$_{loc}$

The subscripts in these patterns indicate the case values (of the nominal forms in question) which we *initially* deem possible (and relevant). Before we proceed with resolving these case ambiguities, including the crucial one in the NOM row, we have to explicate what exactly we mean by case ambiguities here. Let us start with the NOM row of the non-masculine-human declension (21). The determiner *te* is marked as ambiguous between nominative and accusative. This means that *te* can occur only with nominative or accusative nouns (i.e., only in NOM and ACC rows), e.g.:

(23) a. *te* *kobiety*
these women$_{nom/acc}$

b. * *te* *kobiet/kobietom/kobietami/kobietach*
these women$_{gen/dat/ins/loc}$

Similarly, by marking *pięć* as ambiguous with respect to nominative and accusative case, we mean that it can appear only in the NOM and ACC rows of declension patterns. On the other hand, *tych* is unambiguously marked as genitive in the NOM and ACC rows; by that we indicate that it can co-occur only with genitive nouns (we arbitrarily consider the fact that it can also occur with locative nouns irrelevant here):

(24) a. *tych* *kobiet*
these women$_{gen}$

b. * *tych* *kobiety*
these women$_{nom/acc}$

3.1.1.2 The analysis Note first that in both patterns there is total and unambiguous case concord between the determiner, the numeral and the noun phrase in four cases: genitive, dative, instrumental and locative (i.e., in the lexical cases). Note also that in both declensions the nominative phrase is the same as the accusative one. This is expected as far as non-masculine-human declension is concerned, but quite surprising with respect to the masculine-human declension; in Polish, nominative and accusative cases are normally (i.e., in phrases with no numerals) different in masculine-human declension. Our account explains these facts.

The crux of our analysis concerns the case ambiguities indicated in the NOM and ACC rows of examples (21) and (22). In order to try to resolve these case ambiguities we will make the natural and non-controversial assumption that in both declension patterns the elements in corresponding slots have the same case. For example, since in the non-masculine-human

declension (21) the case of *kobiet* in the NOM and ACC rows is unambiguously genitive, we will assume that the case of *mężczyzn* in the corresponding slots of the masculine-human pattern (22) is also genitive (rather than accusative).

Unlike in (21), the numeral in masculine-human pattern (22) is ambiguous with respect to three case values: nominative, accusative and genitive. Applying the same method that we have already used above, we can establish — again, by analogy with non-masculine-human pattern — that the case values really at issue here are nominative and accusative. The considerations so far are summarized below:

(25) Non-masculine-human declension (extract):

	these	*five*	*women*
NOM	te$_{nom/acc}$/tych$_{gen}$	pięć$_{nom/acc}$	kobiet$_{gen}$
ACC	te$_{nom/acc}$/tych$_{gen}$	pięć$_{nom/acc}$	kobiet$_{gen}$

Masculine-human declension (extract):

	these	*five*	*men*
NOM	tych$_{gen/acc}$	pięciu$_{nom/acc}$	mężczyzn$_{gen}$
ACC	tych$_{gen/acc}$	pięciu$_{nom/acc}$	mężczyzn$_{gen}$

In order to show that the numeral phrases in the NOM (and ACC) row are really *accusative*, we will concentrate on the apparent disparity between the possible determiners in both patterns. The crucial fact in the masculine-human declension pattern is that *tych* can normally choose only between genitive and accusative cases:

(26) a. *tych mężczyzn*
these men$_{gen/acc}$

b. * *tych mężczyźni/mężczyznom/mężczyznami/mężczyznach*
these men$_{nom/dat/ins/loc}$

Note first that the non-masculine-human declension allows any of the two determiners *te*$_{nom/acc}$ (which agrees with the numeral *pięć*$_{nom/acc}$) and *tych*$_{gen}$ (which agrees with the noun *kobiet*$_{gen}$). In fact, *te* cannot be nominative; if it were nominative, then analougous (but masculine) nominative determiner should be also allowed in the masculine-human declension pattern. As the judgement below shows, this is definitely not the case:

(27) * *ci pięciu mężczyzn*
these$_{nom}$ five men

This means that, since *te* cannot be nominative, it has to be accusative. But if it is accusative, the numeral it agrees with also has to be accusative; in Polish determiners always agree (with respect to case) with the phrases they modify. So, the NOM and ACC rows of the non-masculine-human declension pattern finally look as follows:

(28) Non-masculine-human declension (extract):

	these	*five*	*women*
NOM	te$_{acc}$/tych$_{gen}$	pięć$_{acc}$	kobiet$_{gen}$
ACC	te$_{acc}$/tych$_{gen}$	pięć$_{acc}$	kobiet$_{gen}$

But this, in turn, means that the numeral in the masculine-human declension also has to be accusative:

(29) Masculine-human declension (extract):

	these	*five*	*men*
NOM	$\text{tych}_{gen/acc}$	pięciu_{acc}	mężczyzn_{gen}
ACC	$\text{tych}_{gen/acc}$	pięciu_{acc}	mężczyzn_{gen}

Note that the foregoing analysis has been conducted outside any specific linguistic theory and without any assumption as to the real structure of the nominal phrases as the ones above; i.e., we have not commited ourselves to any decision on what constitutes head of such phrases: it might be either a numeral (*pięciu*), or a noun phrase (*mężczyzn*). So far any analysis seems plausible, as far as we postulate a linear precedence (LP) rule stating that in (Polish) numeral phrases determiners precede numerals which, in turn, precede noun phrases. This LP rule, when applied to $pięć\ tych_{gen}\ kobiet_{gen}$ gives $tych_{gen}\ pięć\ kobiet_{gen}$.

However, now we commit ourselves to a particular analysis of numeral phrases: we will analyze them as *true* numeral phrases, i.e., headed by a numeral. This stance is consistent not only with the rows corresponding to lexical cases (GEN, DAT, INS and LOC), but also with the ones for ACC; indeed, if we analyzed such phrases as headed by a noun, we would have to explain why an accusative phrase is headed by a genitive noun. The ensuing sections will provide us with more arguments for such analysis of numeral phrases.[23]

The most important conclusion of the foregoing discussion is this: numeral phrases in sentential positions cannot be nominative. Moreover, since they are headed by a numeral (and we have established that the case of the numeral in NOM row is accusative), such phrases have to bear the accusative case.[24] This conclusion explains a number of facts, e.g., the same form of masculine-human numeral phrases in the NOM and ACC rows: nominative and accusative numeral phrases are the same because... there are no nominative numeral phrases! It is the *accusative* numeral phrases that fill the sentential subject positions. However, the most striking conclusion concerns the agreement pattern between numeral phrases in the subject position and the verb.

3.1.2 Numeral subject-verb agreement

The article Czuba and Przepiórkowski (1995) presents a parochial grammatical principle (called Subject-Verb Agreement Principle) which describes agreement patterns between the verb and its subject in Polish. In short, this principle says that if the subject is a nominative nominal phrase then 'full' agreement of gender, number and person takes place, while in all other instances the verb takes *neuter*, *singular* and *third* as values of the respective grammatical categories. These 'other instances' are for example: lack of subject (i.e., empty SUBJECT list), e.g., in the case of meteorological verbs (*mżyć*, 'drizzle'); sentential subject,[25]

[23] See also Saloni and Świdziński (1985) for other arguments for this position.

[24] It has to be emphasized that this is a very unorthodox result, although it has been signalled a.o. by Zabludowski (1989) and Franks (1994). The traditional grammarians analyze numeral phrases in subject position as headed by a genitive noun (cf. Klemensiewicz (1986), p. 121), while the formal Polish grammar Saloni and Świdziński (1985) analyzes them as headed by a nominative numeral.

[25] See Świdziński (1993) and Przepiórkowski (1994).

as required e.g. by *dziwi*, 'makes one wonder'; and nominal subjects with case values different from nominative (it might be argued that verbs such as *ubywać*, 'wane', subcategorize for such subjects).

Being in the possession of this strongly-motivated principle, we do not have to posit any new mechanisms in order to explain the fact that numeral phrases in the subject position also trigger such a 'reduced' agreement pattern:

(30) a. *Dwóch facetów jadło jabłko.*
Two guys $\text{eat}_{3rd,sing,neut,past}$ apple.
'Two guys were eating an apple.'

b. *Pięć kobiet poszło do kina.*
Five women $\text{go}_{3rd,sing,neut,past}$ to cinema.
'Five women went to the cinema.'

Now, in view of our theory of agreement, the 3rd person singular neuter marking on the verb follows from the fact that the numeral phrase (in the examples above: *dwóch facetów* and *pięć kobiet*) is not nominative, and hence invokes the 3rd person singular neuter values of respective categories of the verb.

3.1.3 What are numerals?

The natural question that arises here is why numeral phrases have such heterogeneous declension patterns. In order to answer this question we will first of all posit that — syntactically speaking — numerals are nouns, i.e., numeral phrases are really noun phrases.[26] This position has two practical advantages over any other analyses of numerals: lexical items do not have to subcategorize separately for noun phrases and numeral phrases, and — more importantly — numeral phrases are in the scope of Case Principle (as far as they are structural). Our modelling of numerals in terms of HPSG *sign*s will be based on the observation that there is no case agreement between the numeral and the NP it subcategorizes for only in NOM and ACC rows (cf. (21) and (22)), i.e., exactly in these cases which have to be structural (i.e., which do not have lexical counterparts; cf. case lattice for Polish (4) on page 205).

This observation leads us to the conclusion that there are two kinds of numerals, or rather that each numeral has to have two entries in the lexicon, one with lexical and one with structural case. This is exemplified below:[27]

[26] As far as morphological (esp. inflectional) properties are concerned numerals constitute a separate class, cf. Saloni and Świdziński (1985).

[27] Notice that the PHON feature is a function of STEM and AGR; we follow here Kathol (1995).

(31) a.

$$\left[\begin{array}{l} word \\ \textsc{phon}\ PF(\boxed{1},\boxed{2}) \\ \textsc{stem} \mid \textsc{phon}\ \boxed{2}\langle pi\c{e}\acute{c}\rangle \\ \\ \textsc{synsem} \mid \textsc{loc} \mid \textsc{cat} \end{array}\begin{bmatrix} cat \\ \textsc{head} \begin{bmatrix} nom \\ \textsc{numeral}\ + \\ \textsc{agr}\ \boxed{1}\left[\textsc{case}\ \boxed{3} lex\right] \end{bmatrix} \\ \textsc{comps}\ \langle NP[\textsc{agr} \mid \textsc{case}\ \boxed{3}]\rangle \end{bmatrix}\right]$$

b.

$$\left[\begin{array}{l} word \\ \textsc{phon}\ PF(\boxed{1},\boxed{2}) \\ \textsc{stem} \mid \textsc{phon}\ \boxed{2}\langle pi\c{e}\acute{c}\rangle \\ \\ \textsc{synsem} \mid \textsc{loc} \mid \textsc{cat} \end{array}\begin{bmatrix} cat \\ \textsc{head} \begin{bmatrix} nom \\ \textsc{numeral}\ + \\ \textsc{agr}\ \boxed{1}\left[\textsc{case}\ str\right] \end{bmatrix} \\ \textsc{comps}\ \langle NP[\textsc{agr} \mid \textsc{case}\ gen]\rangle \end{bmatrix}\right]$$

Note that in our account numerals are simply nouns with their NUMERAL feature (appropriate for type *noun*) set to '+'. (All other nouns are specified as −NUMERAL.) Note also that in fact one lexical entry for each numeral will suffice in the lexicon in the technical sense as a simple lexical rule can be used to derive the other — this way we will be able to express the systematic relation between the two signs in (31). These signs will receive morphological case when combined with other words: lexical numerals will receive their case from the subcategorizing element via Subcat Principle, while structural numerals will get it via Case Principle.[28] Of course, Case Principle cannot remain as it is now or it would require nominative numeral phrases (which do not exist!) in the sentential subject position. Numeral phrases (and *only* numeral phrases) in subject positions are accusative, so the Case Principle has to depend on the feature NUMERAL. The revised version of this principle is shown below:

Case Principle (Penultimate Version)
In a *head-complement-structure* of category

- **verb**: the structural subject has a CASE value of *snom* if −NUMERAL or *sacc* if +NUMERAL,
- **verb[−neg]**: the structural object has a CASE value of *sacc*,
- **verb[+neg]**: the structural object has a CASE value of *sgen*,
- **preposition**: the structural object has a CASE value of *sacc*,
- **noun**: any structural argument has a CASE value of *sgen*.

These are the only saturated or almost saturated *head-complement-structures* with structural arguments.

3.1.4 An Example

The above results will be illustrated here with the analysis of an example sentence (32):

[28] We hope that the reader will not be deceived by the 'transformational' language that we use here to describe 'declarative' constraints.

(32) *Pięciu facetów zarządzało siedmioma firmami.*
Five$_{acc}$ guys$_{gen}$ manage$_{neut,3rd,sing,past}$ seven$_{ins}$ companies$_{ins}$.
'Five guys managed seven companies.'

The head of this phrase is the verb *zarządzało* which subcategorizes for a structural nominal subject and an instrumental object:

(33) zarządzało: ⟨NP[*str*], NP[*lins*]⟩

Since the phrase *pięciu facetów* is structural, its head, *pięciu*, is — according to the Head Feature Principle — also structural. This means that it has to satisfy the description (31b) on page 210 and subcategorize for a genitive NP. This is indeed so: *facetów* bears the genitive case. Moreover, since *pięciu* is the head of the subject, +NUMERAL, and *str*, the subject itself is also +NUMERAL and structural. Now Case Principle comes into action and requires the subject to bear accusative case. This in turn, via Subject-Verb Agreement Principle, triggers the reduced agreement pattern which results in the neuter third singular values of the respective categories of the verb form *zarządzało*.

The second argument of the verb is the instrumental object *siedmioma firmami*. The case value of this numeral phrase is lexical, so its head, *siedmioma*, has to satisfy description (31a) above. But this in turn means that the numeral and the nominal phrase it subcategorizes for have to agree in case. Hence, the instrumental case value of *firmami*.

3.2 More Facts

The previous section describes the behaviour of numerals from *pięć* ('five') onwards (with exceptions, see below). Numerals *dwa* ('two') to *cztery* ('four') behave in Polish in a slightly different way; they adhere to the usual declension patterns and always agree with the noun phrases they govern:

(34) Non-masculine-human declension:

	these	*three*	*women*
NOM	te$_{nom/acc}$	trzy$_{nom/acc}$	kobiety$_{nom/acc}$
GEN	tych$_{gen}$	trzech$_{gen}$	kobiet$_{gen}$
DAT	tym$_{dat}$	trzem$_{dat}$	kobietom$_{dat}$
ACC	te$_{nom/acc}$	trzy$_{nom/acc}$	kobiety$_{nom/acc}$
INS	tymi$_{ins}$	trzema$_{ins}$	kobietami$_{ins}$
LOC	tych$_{loc}$	trzech$_{loc}$	kobietach$_{loc}$

(35) Masculine-human declension:

	these	*three*	*men*
NOM	ci$_{nom}$	trzej$_{nom}$	mężczyzni$_{nom}$
GEN	tych$_{gen/acc}$	trzech$_{gen/acc}$	mężczyzn$_{gen/acc}$
DAT	tym$_{dat}$	trzem$_{dat}$	mężczyznom$_{dat}$
ACC	tych$_{gen/acc}$	trzech$_{gen/acc}$	mężczyzn$_{gen/acc}$
INS	tymi$_{ins}$	trzema$_{ins}$	mężczyznami$_{ins}$
LOC	tych$_{loc}$	trzech$_{loc}$	mężczyznach$_{loc}$

Considerations similar to those above lead us to the conclusion that these numerals behave like nouns as far as declension patterns are concerned. In particular, they can bear the nominative case, and the NOM and ACC rows in masculine-human declension (35) differ. Hence, we will analyze them as 'normal' (i.e., −NUMERAL) nouns:

(36)
$$\left[\begin{array}{l} word \\ \text{PHON}\ PF\langle \boxed{1}, \boxed{2} \rangle \\ \text{STEM} \mid \text{PHON}\ \boxed{2}\langle trzy \rangle \\ \\ \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \end{array} \left[\begin{array}{l} cat \\ \\ \text{HEAD} \left[\begin{array}{l} nom \\ \text{NUMERAL} - \\ \text{AGR}\ \boxed{1} \end{array}\right] \\ \\ \text{COMPS}\ \langle NP[\text{AGR}\ \boxed{1}] \rangle \end{array}\right] \right]$$

In fact (cf. Buttler *et al.* (1971)), there is a tendency in modern Polish to adopt a uniform system of numerals. One of the symptoms of these changes is the behaviour of numerals *dwa* ('two') to *cztery* ('four'). Apart from the declension patterns (34) and (35) shown above, these numerals have an alternative masculine-human declension which parallels that of other numerals (such as *pięć*, cf. (22)):

(37) Masculine-human declension:

	these	*three*	*men*
NOM	$\text{tych}_{gen/acc}$	trzech_{acc}	mężczyzn_{gen}
GEN	tych_{gen}	trzech_{gen}	mężczyzn_{gen}
DAT	tym_{dat}	trzem_{dat}	mężczyznom_{dat}
ACC	$\text{tych}_{gen/acc}$	trzech_{acc}	mężczyzn_{gen}
INS	tymi_{ins}	trzema_{ins}	mężczyznami_{ins}
LOC	tych_{loc}	trzech_{loc}	mężczyznach_{loc}

As Buttler *et al.* (1971) claim (p. 343), there is a rapid shift in the contemporary Polish towards the usage of the latter form of the masculine-human declension patterns. That is, in terms of our *sign* feature structures, there is a rapid expansion of lexical entries such as (31).

One more example of this tendency is given by the numerals such as *tysiąc* ('thousand'), *milion* ('million'), etc. Traditionally, they are analyzed as nouns which always assign genitive case to their NP complements:

(38)

	thousand	*men*	(*women*)
NOM	$\text{tysiąc}_{nom/acc}$	mężczyzn_{gen}	(kobiet_{gen})
GEN	tysiąca_{gen}	mężczyzn_{gen}	(kobiet_{gen})
DAT	tysiącu_{dat}	mężczyzn_{gen}	(kobiet_{gen})
ACC	$\text{tysiąc}_{nom/acc}$	mężczyzn_{gen}	(kobiet_{gen})
INS	tysiącem_{ins}	mężczyzn_{gen}	(kobiet_{gen})
LOC	tysiącu_{loc}	mężczyzn_{gen}	(kobiet_{gen})

The reader will immediately notice that numeral phrases involving *tysiąc*, etc. crucially have to be analyzed as true numeral phrases headed by *tysiąc*: the case of the whole phrase is

the same as the case of the numeral, while the subcategorized NP is always genitive. This provides us with one more argument, an argument of uniformity, for analyzing numerals in *all* numeral phrases as heads.

Notice also that *tysiąc* cannot be analyzed here just as a −NUMERAL noun for the reasons we give presently. In Polish *tysiąc* has the masculine gender. If, when in sentential subject position, it were really just a normal (i.e., −NUMERAL) nominative noun, it would take part in gender agreement with the past tense verb as all other nominative nouns do. This is, however, not the case:

(39) a. *Tysiąc* *mężczyzn* *poszło* *do* *pracy.*
One thousand$_{nom/acc,\mathrm{masc}}$ men went$_{3rd,sing,\mathrm{neut}}$ to work.
'One thousand men went to work.'

b. * *Tysiąc* *mężczyzn* *poszedł* *do* *pracy.*
One thousand$_{nom/acc,\mathrm{masc}}$ men went$_{3rd,sing,\mathrm{masc}}$ to work.
'One thousand men went to work.'

This[29] strongly suggests that the noun (numeral) phrase *tysiąc mężczyzn* is not assigned nominative case. But this fits well in the picture drawn so far: we will posit that *tysiąc*, *milion*, etc. are +NUMERAL nouns which combine with genitive (plural) NPs but which are not themselves specified for case:

(40)
$$\left[\begin{array}{l} word \\ \textsc{phon}\ PF(\boxed{1},\boxed{2}) \\ \textsc{stem} \mid \textsc{phon}\ \boxed{2}\langle tysiąc\rangle \\ \textsc{synsem} \mid \textsc{loc} \mid \textsc{cat} \left[\begin{array}{l} cat \\ \textsc{head} \left[\begin{array}{l} nom \\ \textsc{numeral}\ + \\ \textsc{agr}\ \boxed{1} \end{array}\right] \\ \textsc{comps}\ \langle NP[\textsc{agr} \mid \textsc{case}\ gen]\rangle \end{array}\right] \end{array}\right]$$

Notice, that phrases headed by this sign can appear both in environments requiring lexical case and in those requiring structural case.[30] As usual, in the former instance case values will be assigned by a lexical element, while in the latter — by the Case Principle. In particular, numeral phrases in subject position headed by *tysiąc*, being +NUMERAL and not specified for the CASE value (i.e., potentially structural), will be assigned the structural accusative case and, hence, trigger the 3rd person singular neuter agreement pattern on the verb as predicted by the analysis of agreement in Czuba and Przepiórkowski (1995). This explains the ungrammaticality of (39b) where the verb *poszedł* is marked as masculine.

[29] Some speakers find (39b) also acceptable. This suggests that their lexical entries for *tysiąc* are ambiguous with respect to the NUMERAL feature.

[30] Actually, in this respect *tysąc* behaves like a 'normal' (−NUMERAL) noun.

3.3 Collective Numerals

There is one more kind of numerals that we have said nothing about: the so-called collective numerals.[31] This is a group of numerals used with noun phrases describing people of mixed sex (e.g., *pięcioro studentów*, 'five students (of mixed sex)'), children (*czworo dzieci*, 'four children'), small animals (*troje kurczqt*, 'three chickens'), and with some *plurale tantum* nouns (*pięcioro drzwi*, 'five doors'). However, as these numerals are to some extent redundant and their declension is very ill-behaved (see below), the 'regular' numerals take over in contemporary Polish.[32] Thus, in the examples below, the (b) form, although still not sanctioned by the linguistic norm, becomes more and more widespread:[33]

(41) a. *W klasie było dwadzieścioro dzieci.*
In class were$_{3rd,sing,neut}$ twenty$_{coll,acc}$ children$_{gen}$.
'There were twenty children in the class.'

b. *W klasie było dwadzieścia dzieci.*
In class were$_{3rd,sing,neut}$ twenty$_{reg,acc}$ children$_{gen}$.
'There were twenty children in the class.'

(42) a. *Zapukał do pięciorga drzwi.*
Knocked$_{3rd,sing,masc}$ to five$_{coll,gen}$ doors$_{gen}$.
'He knocked at five doors.'

b. *Zapukał do pięciu drzwi.*
Knocked$_{3rd,sing,masc}$ to five$_{reg,gen}$ doors$_{gen}$.
'He knocked at five doors.'

Simultaneously, apart from this quantitative change, collective numerals undergo another, qualitative change in declension pattern. Consider first the current declension pattern of these numerals:

(43)

	five	*children*
NOM	pięcioro$_{acc}$	dzieci$_{gen}$
GEN	pięciorga$_{gen}$	dzieci$_{gen}$
DAT	pięciorgu$_{dat}$	dzieciom$_{dat}$
ACC	pięcioro$_{acc}$	dzieci$_{gen}$
INS	pięciorgiem$_{ins}$	dzieci$_{gen}$/*dziećmi$_{ins}$
LOC	pięciorgu$_{loc}$	dzieciach$_{loc}$

[31] In this section we will draw heavily on obervations made by Buttler *et al.* (1971).

[32] They have already taken over to a large extent: collective numerals were used with any noun phrases once.

[33] These examples are taken from Buttler *et al.* (1971), p. 30. They also note that the supersession of collective numerals by regular numerals becomes visible in the lingusitic norm which allows both following constructions:

(i) *W przedszkolu jest trzydzieścioro/trzydzieści czworo dzieci.*
In kindergarten is$_{3rd,sing}$ thirty$_{coll/reg,acc}$ four$_{coll,acc}$ children$_{gen}$.
'There are thirty four children in the kindergarten.'

Notice that this pattern differs from that of 'regular' numerals (cf. (21) and (22)) in the INS row: the numeral requires a genitive (rather than instrumental) noun phrase here. Apparently, this awkward behaviour of collective numerals is caused by their transitional status from the 'agreeing' pattern[34] (like that of regular numerals, e.g., *pięć*) to the 'governing' pattern (like that of *tysiąc*). The existence of this process is confirmed by the fact that a steady shift towards the regular governing pattern (especially in locative) can be observed in contemporary Polish.[35]

Of course, these diachronic considerations do not relieve us from the duty of modelling the current state of the language. Thus, on the basis of the foregoing discussion, we will posit the following lexical entries for collective numerals:

(44) a.
$$\begin{bmatrix} word \\ \text{PHON} \;\; PF(\boxed{1},\boxed{2}) \\ \text{STEM} \mid \text{PHON} \;\; \boxed{2}\langle pięcioro\rangle \\ \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \;\; \begin{bmatrix} cat \\ \text{HEAD} \;\; \begin{bmatrix} nom \\ \text{NUMERAL} \;\; + \\ \text{AGR} \;\; \boxed{1}\left[\text{CASE} \;\; \boxed{3} lex \wedge \neg lins\right] \end{bmatrix} \\ \text{COMPS} \;\; \langle NP[\text{AGR}\mid\text{CASE} \;\; \boxed{3}]\rangle \end{bmatrix} \end{bmatrix}$$

b.
$$\begin{bmatrix} word \\ \text{PHON} \;\; PF(\boxed{1},\boxed{2}) \\ \text{STEM} \mid \text{PHON} \;\; \boxed{2}\langle pięcioro\rangle \\ \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \;\; \begin{bmatrix} cat \\ \text{HEAD} \;\; \begin{bmatrix} nom \\ \text{NUMERAL} \;\; + \\ \text{AGR} \;\; \boxed{1}\left[\text{CASE} \;\; str \vee lins\right] \end{bmatrix} \\ \text{COMPS} \;\; \langle NP[\text{AGR} \mid \text{CASE} \;\; gen]\rangle \end{bmatrix} \end{bmatrix}$$

These feature structures differ from those for 'regular' numerals (cf. (31)) minimally; the only difference is slightly more complex values of AGR|CASE feature in (44).

4 Indefinite Numerals

4.1 The Analysis

In this section we will show that the vast majority of the so-called indefinite numerals patterns the numerals described in the previous section. We will also present an intriguing puzzle, the highly idiosyncratic behaviour of an indefinite numeral *dużo* ('a lot of'), and argue that this idiosyncrasy confirms the account of case in Polish given so far in a striking way.

[34] Of course, the numeral 'agrees' with the noun phrase only in lexical cases, while it 'governs' it (i.e., requires genitive case) in structural cases.

[35] The interesting question is what rules — if, indeed, any — govern the changes described above, that is, why the transition from 'agreeing' to 'governing' valency of collective numerals started in instrumental case, why the transition from Genitive of Negation to the lack of it in Russian takes place as described in Timberlake (1986), etc. These matters are, of course, well outwith the scope of this work.

Morphologically, indefinite numerals constitute a very heterogeneous class. They are traditionally (cf. Buttler *et al.* (1971), pp. 341–343) divided into pronominal numerals (*tyle*, *ile*, etc.), adjectival numerals (*dużo*, *wiele*) and nominal numerals (*szereg*, *część*), but this will not concern us here. What is important for us is their valency. From this point of view we can split indefinite numerals into three classes:

CLASS 1 Indefinite numerals which parallel 'ordinary' numerals (such as *pięć*, 'five'). These are mainly pronominal numerals and some adjectival numerals, e.g., *wiele* ('many'), *kilka* ('a few'), *ile* ('how many'), *tyle* ('that many'), *parę* ('a couple'), etc. Their lexical entries will be almost identical with those of 'ordinary' numerals (see declension patterns (21)–(22) and feature structures (31)).

CLASS 2 Indefinite numerals which behave like *tysiąc*, *milion*, etc. (See declension patterns (38) and feature structure (40).) These are mainly nominal numerals such as *mnóstwo*, *mnogość* ('lots of'), *szereg* ('series'), *część* ('part of'), etc.

Due to the common (in Polish) phenomenon of numeralization, CLASS 2 is currently the most actively expanding group of numerals. Numeralization is the process of transforming nouns into numerals. As we noted above, numerals (apart from 1–4) do not bear nominative case; instead the Case Principle assigns structural accusative case to those which function as sentential subjects. This, in turn, triggers the 3rd person singular neutral agreement pattern rather than the usual subject-verb agreement pattern. Hence, in practice, numeralization can be witnessed when what used to be a nominal phrase co-occurs with 3rd person singular neuter verb. Examples of nouns that seem to be undergoing the process currently are (cf. Buttler *et al.* (1971), p. 347) *szereg* ('series'), *moc* ('plenty'), *część* ('part of'). The usage is shifting from patterns such as (45a) towards (45b).

(45) a. *Szereg* *osób* *wiedział* *o* *tym.*
Series$_{nom/acc,sing,\mathrm{masc}}$ people$_{gen}$ knew$_{3rd,sing,\mathrm{masc}}$ about this.
'A series of people knew about this.'

b. *Szereg* *osób* *wiedziało* *o* *tym.*
Series$_{nom/acc,sing,\mathrm{masc}}$ people$_{gen}$ knew$_{3rd,sing,\mathrm{neut}}$ about this.
'A series of people knew about this.'

Of course, in terms of our feature structure numeralization is simply a change of value of NUMERAL from '−' to '+'. We do not have much to say about lexical entries of CLASS 2 indefinite numerals as they closely match that of *tysiąc* (cf. (40)).[36]

The most interesting class of indefinite numerals is, however, CLASS 3:

CLASS 3 Indefinite numerals which are traditionally analyzed as having only nominative and accusative forms (cf. Doroszewski (1980)), e.g., *dużo* ('a lot'), *mało* ('little'), *trochę* ('a little'), *sporo* ('quite a lot'), etc.

[36]It is perhaps worth noting here that the same behaviour is also exhibited by the so-called fractional numerals (*półtora*, 'one and a half', *dwie trzecie*, 'two thirds', etc.) and, to some extent, collective numerals (see section 3.3). Again, these numerals have to be analyzed as heads of the nominal phrases they occur in.

Numerals such as *dużo* ('a lot') do not decline, they always have the same nominative/accusative form and always combine with genitive NPs. Below we present the defective declension pattern for *dużo*:[37]

(46)

	a lot of	*men*	*(women)*
NOM	dużo$_{acc}$	mężczyzn$_{gen}$	(kobiet$_{gen}$)
GEN	—	—	—
DAT	—	—	—
ACC	dużo$_{acc}$	mężczyzn$_{gen}$	(kobiet$_{gen}$)
INS	—	—	—
LOC	—	—	—

The puzzle concerning these numerals is that they are grammatical in some positions which normally require genitive case, but not in others:

(47) a. *Nie mam w domu (zbyt) dużo chleba.*
Not have$_{1st,sing}$ in home (too) a lot of$_{nom/acc}$ bread$_{gen}$.
'I don't have (too) much bread at home.'

b. *Nie mam w domu chleba.*
Not have$_{1st,sing}$ in home bread$_{gen}$.
'I don't have bread at home.'

c. * *Nie mam w domu chleb.*
Not have$_{1st,sing}$ in home bread$_{acc}$.
'I don't have bread at home.'

(48) a. *Nie lubię dużo osób.*
Not like$_{1st,sing}$ a lot of$_{nom/acc}$ people$_{gen}$.
'I don't like a lot of people.'

b. *Nie lubię tych osób.*
Not like$_{1st,sing}$ these$_{gen}$ people$_{gen}$.
'I don't like these people.'

c. * *Nie lubię te osoby.*
Not like$_{1st,sing}$ these$_{acc}$ people$_{gen}$.
'I don't like these people.'

(49) a. * *Boję się dużo osób.*
Fear$_{1st,sing}$ REFL a lot of$_{nom/acc}$ people$_{gen}$.
'I am afraid of a lot of people.'

[37] *Dużo*-phrases, when subjects of sentences, always trigger the 3rd person singular neuter agreement patters. This means, that just as other numeral phrases, they should be analyzed as accusative, rather then nominative, phrases.

b. *Boję się tych osób.*
Fear$_{1st,sing}$ REFL these$_{gen}$ people$_{gen}$.
'I am afraid of these people.'

c. * *Boję się te osoby.*
Fear$_{1st,sing}$ REFL these$_{nom/acc}$ people$_{nom/acc}$.
'I am afraid of these people.'

In the examples above, the (b) and (c) sentences show that a genitive NP is required by *nie mam* ('I don't have'), *nie lubię* ('I don't like') and *boję się* ('I am afraid') (see (b)), and that it cannot be realized by an accusative phrase (see (c)). However, in (47a) and (48a) *dużo*-phrases are allowed, while in (49a) they are not.[38]

The careful reader will have noticed that these examples themselves suggest an answer to the quandary: *dużo*-phrases are allowed under the Genitive of Negation, but not as a genitive complement of a verb.[39] This, and the underlying assumption we made implicitly, namely that verbs requiring genitive complements specify them as *lexical* genitive, suggests that the indefinite numerals of CLASS 3 can only be assigned structural case, never lexical. Thus, the ungrammaticality of (49a) stems from the fact that *boję się* subcategorizes for a lexical genitive phrases (which cannot be realized by *dużo*-phrases), while the grammaticality of (47a) and (48a) is a consequence of the fact, that *mieć* ('have') and *lubić* ('like') require a structural complement (which can be realized by *dużo*-phrases).

These considerations lead us to postulating the following lexical entry for *dużo*:

(50)
$$\left[\begin{array}{l} word \\ \text{PHON } \langle dużo \rangle \\ \\ \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \end{array}\right. \left[\begin{array}{l} cat \\ \\ \text{HEAD} \left[\begin{array}{l} nom \\ \text{NUMERAL} + \\ \text{AGR} \mid \text{CASE } str \end{array}\right] \\ \text{COMPS } \langle NP[\text{AGR} \mid \text{CASE } gen] \rangle \end{array}\right] \left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right]$$

Notice that the interaction of the Case Principle, agreement patterns, and simple lexical entries accounts in a very elegant way for the quirky behaviour of CLASS 3 indefinite numerals. Being +NUMERAL, structural and nominal, these indefinite numerals get (via the Case Principle) accusative case (*sacc*) when in sentential subject position. As they bear a case different from nominative, they trigger the 3rd person singular neuter agreement pattern. This in turn means that the verb has the 3rd person singular neuter agreement features:

(51) *Dużo osób poszło do domu.*
A lot of$_{acc}$ people$_{gen}$ went$_{3rd,sing,neut}$ to home.
'A lot of people went home.'

[38] Actually, some speakers feel uncomfortable with (48a), but they always deem it more grammatical than (49a).

[39] Saloni and Świdziński (1985) seem to simplify things suggesting (p. 83) that *dużo*-phrases are allowed with verbs and disallowed as complements of nouns: examples like (49a) are clearly ungrammatical for all the native speakers we have consulted.

4.2 Some Ramifications

Our analysis of *dużo* supports many of the decisions we have taken in the previous sections. In this subsection we will point out two of them: the analysis of numeral phrases as headed by a numeral (see p. 209), and the analysis of accusative complements of prepositions as structural (see section 2.3, p. 201). We will start with the latter.

The crucial observation that we will employ here is that *dużo*-phrases are specified as bearing an arbitrary structural case, but *only* structural case. This analysis has allowed us to explain the extremely idiosyncratic behaviour of *dużo*-phrases, esp. the fact that these phrases seem to be allowed in some genitive environments, while disallowed in others. In other words, we have found ourselves in possession of a convenient test for checking structurality of any given environment.

This test confirms our analysis of accusative prepositional arguments:

(52) a. *Maria czeka na dużo osób.*
Mary waits on a lot of people.
'Mary is waiting for a lot of people.'

b. *Janek przejeżdżał przez dużo wsi i miasteczek.*
John went through a lot of villages and towns.
'John went through a lot of villages and towns.

On the other hand, if we analyzed prepositions as markers, we would have two options. One would be to assume that all 'marked' nominal phrases are lexical (cf. Heinz and Matiasek (1994)), but this would contradict the judgements above (because *dużo*-phrases are structural). The other would be to allow structural 'marked' phrases, but then we would have to add several clauses for 'marked' nominal phrases to the Case Principle (each corresponding to a different category of *head-complement-structure*: verb, noun, etc.). By contrast, our analysis allows us to add just one clause to the Case Principle (that for prepositions) and to get rid of lexical accusative altogether.

The other point we want to emphasize here is more fundamental. We have already given several arguments for analyzing numeral phrases as *true* numeral phrases, i.e., as phrases headed by a numeral. Our analysis of *dużo*-phrases provides us with one more, essentialy an argument of uniformity (similar to the one mentioned in section 3.2). It is crucial that in phrases such as *dużo osób* ('a lot of people') it is the numeral that heads the phrase. If it were the genitive noun (*osób*), then it would be very difficult to account for the following judgements:

(53) a. * *Boję się dużo osób.*
Fear$_{1st,sing}$ REFL a lot of$_{nom/acc}$ people$_{gen}$.
'I am afraid of a lot of people.'

b. *Boję się wielu osób.*
Fear$_{1st,sing}$ REFL a lot of$_{gen}$ people$_{gen}$.
'I am afraid of a lot of people.'

c. *Boję się tych osób.*
Fear$_{1st,sing}$ REFL these$_{gen}$ people$_{gen}$.
'I am afraid of these people.'

If *osób* were to be the head here, then, in order to explain the ungrammaticality of (53a) and the grammaticality of (53b) and (53c), we would have to postulate that *bać się* subcategorizes for a genitive NP *not* modified by any indefinite numerals of CLASS 3 (cf. (53a)) but possibly modified by some other numeral (cf. (53b)), or not modified at all (cf. (53c)); a highly *ad hoc* explanation to give. On the other hand, upon our account the NP *dużo osób* is headed by the numeral[40] whose case is specified simply as *str*. This means that the case value of the whole NP is *str* (by the Head Feature Principle), and, thus, it cannot fulfill syntactic requirements of the verb *bać się* subcategorizing for a lexical NP (specifically, for NP[CASE *lgen*]). On the other hand, the indefinite numerals *wielu* and *tych* behave like most numerals and can bear either lexical (as in the examples above) or structural case.

4.3 Nominalization Revisited

In section 2.2 we have stated that "Polish parallels German" as far as nominalization is concerned. Now, in view of some foregoing results, we will have to change our view on the matter.

We will again apply the test on structurality of a given environment provided by *dużo*-phrases, this time to investigate complements of nominalized verbs. We illustrate our considerations with the transitive verb *zjeść* ('eat').

(54) *Janek zjadł dużo rodzynek.*
John$_{nom}$ ate a lot of raisins$_{gen}$.
'John has eaten a lot of raisins.'

(55) *Janek* ***nie*** *zjadł dużo rodzynek.*
John$_{nom}$ not ate a lot of raisins$_{gen}$.
'John has not eaten a lot of raisins.'

The above examples show that the object of the verb is structural and — just as predicted by the Case Principle and the lexical entry for *dużo* (cf. (50)) — *dużo rodzynek* is allowed as an object.

However, judgements such as the one below seem to contradict the Case Principle:

(56) * *Zjedzenie dużo rodzynek przez Janka mogło mu zaszkodzić.*
Eating a lot of raisins by John might have he$_{dat}$ harm.
'John's eating many raisins might have harmed him.'

[40] Remember that numerals *are* nouns!

The unacceptability of the above sentence cannot be a matter of semantic restrictions as the sentence below having the same meaning as (56) is perfectly grammatical.

(57) *Zjedzenie wielu rodzynek przez Janka mogło mu zaszkodzić.*
Eating a lot of raisins by John might have he$_{dat}$ harm.
'Eating many raisins might have harmed John.'

Note that the implicit assumption in our (as well as that of Heinz and Matiasek (1994)) rendering of nominalization is that this process, realized as a lexical rule, does not change CASE values of SUBCAT elements. In other words, structural arguments of a verb stay stuctural as arguments of deverbal nouns. Examples like (57) argue against maintaining this assumption. Instead, we will assume that the nominalization lexical rule changes all structural CASE values of SUBCAT elements to *lgen*.

Such an analysis should not seem *ad hoc* as there are independent reasons for the nominalization lexical rule to make changes in SUBCAT. The most conspicuous such change concerns sentential subjects. As the examples below show, they can be realized either by NP[*gen*] (cf. (58)), or by PP[PFORM 'przez' + *acc*] (cf. (59)):

(58) a. *Maria czeka na Janka.*
Mary$_{nom}$ waits on John$_{acc}$.
'Mary is waiting for John.'

b. *czekanie Marii na Janka*
waiting Mary$_{gen}$ on John$_{acc}$
'Mary's waiting for John'

(59) a. *Jan je rodzynki.*
John$_{nom}$ eats raisins$_{acc}$.
'John is eating raisins.'

b. *jedzenie rodzynek przez Jana*
eating raisins$_{gen}$ by John$_{acc}$
'John's eating raisins'

Thus, in the process of nominalization the SUBCAT list changes considerably.

We will not attempt to formally state the nominalization lexical rule here, as its technical characterization could only distort the picture drawn above. Instead, we will give examples of operation of this rule:

(60)
$$\left[\begin{array}{l} \textit{word} \\ \text{PHON } \langle \textit{czekać} \rangle \\ \\ \text{SYNSEM|LOC|CAT} \end{array}\right. \left[\begin{array}{l} \textit{category} \\ \text{HEAD } \textit{verb} \\ \text{SUBCAT } \langle \textit{NP}\boxed{1}\textit{[str]}, \textit{PP}\boxed{2}\textit{[}\text{PFORM } \textit{'na' + str]}\rangle \end{array}\right] \Bigg] \mapsto$$

(61)
$$\begin{bmatrix} word \\ \text{PHON } \langle czekanie \rangle \\ \text{SYNSEM|LOC|CAT} \quad \begin{bmatrix} category \\ \text{HEAD } noun \\ \text{SUBCAT } \langle (NP\boxed{1}[lgen]),\ (PP\boxed{2}[\text{PFORM } `na' + str]) \rangle \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} word \\ \text{PHON } \langle jeść \rangle \\ \text{SYNSEM|LOC|CAT} \quad \begin{bmatrix} category \\ \text{HEAD } verb \\ \text{SUBCAT } \langle NP\boxed{1}[str],\ NP\boxed{2}[str] \rangle \end{bmatrix} \end{bmatrix}$$

$$\mapsto$$

$$\begin{bmatrix} word \\ \text{PHON } \langle jedzenie \rangle \\ \text{SYNSEM|LOC|CAT} \quad \begin{bmatrix} category \\ \text{HEAD } noun \\ \text{SUBCAT } \langle (NP\boxed{2}[lgen]),\ (PP\boxed{1}[\text{PFORM } `przez' + str]) \rangle \end{bmatrix} \end{bmatrix}$$

(62)
$$\begin{bmatrix} word \\ \text{PHON } \langle pomagać \rangle \\ \text{SYNSEM|LOC|CAT} \quad \begin{bmatrix} category \\ \text{HEAD } verb \\ \text{SUBCAT } \langle NP\boxed{1}[str],\ NP\boxed{2}[ldat] \rangle \end{bmatrix} \end{bmatrix}$$

$$\mapsto$$

$$\begin{bmatrix} word \\ \text{PHON } \langle pomaganie \rangle \\ \text{SYNSEM|LOC|CAT} \quad \begin{bmatrix} category \\ \text{HEAD } noun \\ \text{SUBCAT } \langle (NP\boxed{1}[lgen]),\ (NP\boxed{2}[ldat]) \rangle \end{bmatrix} \end{bmatrix}$$

The first two examples ((60) and (61)) correspond to the nominalization examples (58) and (59). In these examples all structural nominal phrases become lexical genitive, all the arguments become optional, and the nominative subject is changed to a prepositional phrase (example (61)). The last example shows that lexical complements (in this case *ldat*) do not change their case in the process of nominalization.

By positing such a lexical rule we have transferred part of the scope of Case Principle to the lexicon. The revised Case Principle will be rid of the noun clause:

Case Principle (Last Version)
In a *head-complement-structure* of category

- **verb**: the structural subject has a CASE value of *snom* if −NUMERAL or *sacc* if +NUMERAL,
- **verb[−neg]**: the structural object has a CASE value of *sacc*,
- **verb[+neg]**: the structural object has a CASE value of *sgen*,
- **preposition**: the structural object has a CASE value of *sacc*.

These are the only saturated or almost saturated *head-complement-structures* with structural arguments.

The reader familiar with Chomsky's GB will note that now the Case Principle is compatible with the independently motivated Case Assignment Principle of GB which states that "an NP receives Case at S-structure if it is governed by and adjacent to [−N]. [−N] elements are INFL[+tense], V and P" (cf. (Cowper, 1992, p. 102)).[41]

5 Passive

In this — very short — section we will show that (unlike in German, cf. Heinz and Matiasek (1994) and Pollard (1994)) passivization in Polish does not seem amenable to an analysis in terms of structural vs. lexical case dichotomy.

First of all, note that there are verbs requiring lexical objects which can be nevertheless passivized.

(63) a. *Jan kieruje fabryką.*
John$_{nom}$ manages factory$_{ins}$.
'John manages a factory.'

b. *Fabryka jest kierowana przez Jana.*
Factory$_{nom}$ is managed by John$_{acc}$.
'A factory is managed by John.'

In this example it is the (lexical) instrumental object that gets passivized. According to the case lattice for Polish (cf. (4) on page 205) instrumental cannot be an instance of structural case. This observation is confirmed by the inability of the object in (63a) to change its case (to genitive) under nominalization or negation (i.e., by its failure to pass the two tests of structural environment):

(64) a. *kierowanie fabryką/*fabryki*
managing factory$_{ins/gen}$
'managing a factory'

b. *Jan nie kieruje fabryką/*fabryki.*
John$_{nom}$ not manages factory$_{ins/gen}$.
'John does not manage a factory.'

Moreover, only some verbs subcategorizing for instrumental complements can be passivized:[42]

[41] See also Chomsky (1986a) p. 36, Franks (1990), Franks (1994) and Netter (1994).

[42] In Polish, unlike in German, the passivized object always receives the nominative case. Thus the sentence:

(i) * *Chorągiewką jest machana przez Jana.*
Banner$_{ins}$ is waved by John$_{acc}$

is clearly ungrammatical.

(65) a. *Jan macha chorągiewką.*
John$_{nom}$ waves banner$_{ins}$.
'John waves a banner.'

b. * *Chorągiewka jest machana przez Jana.*
Banner$_{nom}$ is waved by John$_{acc}$
'A banner is being waved by John.'

In the example above, *macha* ('waves') seems to have the same syntactic subcategorization requirements as *kieruje*, but it cannot passivize.

Contrasts such as (63) vs. (65) suggest that passivization in Polish is independent of the syntactic characterization of the SUBCAT arguments. This conlusion is further supported by the observation that, contrary to the generalization often made, not all verbs which are transitive (in the sense that their SUBCAT value is ⟨NP[str], NP[str]⟩) can be passivized. Some exceptions are given below:

(66) a. *Brzuch boli Jana.*
Stomach$_{nom}$ aches John$_{acc}$.
'John has a stomach ache.'

b. * *Jan jest bolony przez brzuch.*
John$_{nom}$ is ached by stomach$_{acc}$.
'John has a stomach ache.'

(67) a. *Noga swędzi Jana.*
Leg$_{nom}$ itches John$_{acc}$.
'John has an itchy leg.'

b. * *Jan jest swędzony przez nogę.*
John$_{nom}$ is itched by leg$_{acc}$.
'John has an itchy leg.'

Again, this contrasts with the usual behaviour of Polish transitive verbs:

(68) a. *Jan lubi Marię.*
John$_{nom}$ likes Mary$_{acc}$.
'John likes Mary.'

b. *Maria jest lubiona przez Jana.*
Mary$_{nom}$ is liked by John$_{acc}$.
'Mary is liked by John.'

Note that this contrast does not seem to be justified by any difference in case markings of the arguments of *boleć* and *swędzić* on the one hand, and *lubić* on the other.[43]

[43] Actually, verbs like *boleć* or *swędzić* are interesting in one more respect: they cannot be nominalized. This may be caused by the fact that the first argument of these verbs is not an agent. Thus, we could add the condition of agentivity to the prerequisites of nominalization lexical rule and, perhaps, passivization lexical rule. Such an amendment, however, would not explain the passivization behaviour of instrumental objects described above (cf. examples (63)–(65)).

The foregoing observations lead us to the conclusion that in Polish the phenomena of passivization is (to a large extent) independent of the syntactic case values of verb's arguments, and — as such — outwith the scope of this paper.[44]

6 Conclusion

The main thesis of this paper is this: the stuctural vs. lexical case dichotomy known in GB and transferred to HPSG by Heinz and Matiasek (1994) is confirmed by many phenomena in Polish. Of these phenomena, the behaviour of numerals and indefinite numerals provides the most striking such confirmation. We claim that our analysis constitutes an elegant and empirically adequate account of Polish numerals, the class of lexemes infamous for their quirky behaviour.

However, we leave many interesting questions concerning both case assignment and numerals unanswered. For example, we have had nothing to say here about the origin of case in circumstantials or about the syntax of names of numbers.[45] We have also ignored the problem of so-called *distributivus* (the case after distributive preposition *po*, cf. e.g. Gruszczyński (1989)). These phenomena are, as far as we know, still an uncharted region of formal linguistics.

Acknowledgments

I would like to thank Marek Świdziński for reading the first version of this paper, and Bob Borsley, Zelal Güngördü and Johannes Matiasek for helpful comments on the penultimate version. This paper originated as a part of Czuba and Przepiórkowski (1995). Of course, all errors remain my own.

References

Babby, L. H. (1986). The locus of case assignment and the direction of percolation: Case theory and Russian. In R. D. Brecht and J. S. Levine, editors, *Case in Slavic*. Slavica Publishers Inc.

Buttler, D., Kurkowska, H., and Satkiewicz, H. (1971). *Kultura języka polskiego: Zagadnienia poprawności gramatycznej*. Państwowe Wydawnictwo Naukowe, Warszawa.

Carpenter, B. (1992). *The Logic of Typed Feature Structures: With Applications to Unifications Grammars, Logic Programs and Constraint Resolution*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge.

Chomsky, N. (1986a). *Barriers*. MIT Press, Cambridge, Mass.

Chomsky, N. (1986b). *Knowledge of Language: Its Nature, Origin and Use*. Convergence. Praeger, New York.

[44] The reader interested in semantic account of passivization in Polish is referred to Holvoet (1991).

[45] Pullum and Gazdar (1982) suggest that "knowledge of how to construct such names... is knowledge of mathematics rather than of language" (fn. 18).

Comrie, B. (1986). On delimiting cases. In R. D. Brecht and J. S. Levine, editors, *Case in Slavic*, pages 86–106. Slavica Publishers Inc.

Corbett, G. (1978). Numerous squishes and squishy numerals in Slavonic. In B. Comrie, editor, *Classification of Grammatical Categories*, pages 43–73. Linguistic Research Inc., Edmonton.

Cowper, E. A. (1992). *A Concise Introduction to Syntactic Theory: The Government-Binding Approach*. University of Chicago Press, Chicago.

Czuba, K. and Przepiórkowski, A. (1995). Agreement and case assignment in Polish: An attempt at a unified account. ICS PAS Report 783, Polish Academy of Sciences, Warsaw.

Doroszewski, W., editor (1980). *Słownik Poprawnej Polszczyzny*. Państwowe Wydawnictwo Naukowe, Warszawa.

Franks, S. (1986). Case and the structure of NP. In R. D. Brecht and J. S. Levine, editors, *Case in Slavic*, pages 220–243. Slavica Publishers Inc.

Franks, S. (1990). Case, configuration and argumenthood: reflections on second dative. *Russian Linguistics*, **14**, 231–254.

Franks, S. (1994). Parametric properties of numeral phrases in Slavic. *Natural Language and Linguistic Theory*, **12**.

Gruszczyński, W. (1989). *Fleksja rzeczowników pospolitych we współczesnej polszczyźnie pisanej (na materiale „Słownika języka polskiego" PAN pod redakcją W. Doroszewskiego)*, volume 122 of *Prace Językoznawcze*. Ossolineum, Wrocław.

Haegeman, L. (1991). *Introduction to Government and Binding Theory*. Basil Blackwell, Oxford.

Haider, H. (1985). The case of German. In J. Toman, editor, *Studies in German Grammar*. Foris Publications, Dordrecht.

Heinz, W. and Matiasek, J. (1994). Argument structure and case assignment in German. In J. Nerbonne, K. Netter, and C. Pollard, editors, *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes. Center for the Study of Language and Information, Stanford, Ca.

Holvoet, A. (1991). *Transitivity and Clause Structure in Polish. A Study in Case Marking*, volume 95 of *Prace Slawistyczne*. Slawistyczny Ośrodek Wydawniczy PAN, Warszawa.

Kathol, A. (1995). Agreement and the syntax-morphology interface in HPSG. OSU Working Papers in Linguistics 45, Ohio State University.

Klemensiewicz, Z. (1986). *Podstawowe wiadomości z gramatyki języka polskiego*. Państwowe Wydawnictwo Naukowe, Warszawa, xiii edition.

Mel'čuk, I. (1986). Toward a definition of case. In R. D. Brecht and J. S. Levine, editors, *Case in Slavic*, pages 35–85. Slavica Publishers Inc.

Nerbonne, J., Netter, K., and Pollard, C., editors (1994). *German in Head-Driven Phrase Structure Grammar*. Number 46 in CSLI Lecture Notes. Center for the Study of Language and Information, Stanford, Ca.

Netter, K. (1994). Towards a theory of functional heads. In J. Nerbonne, K. Netter, and C. Pollard, editors, *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes. Center for the Study of Language and Information, Stanford, Ca.

Polański, K., editor (1993). *Encyklopedia językoznawstwa ogólnego*. Ossolineum, Wrocław.

Pollard, C. (1994). Toward a unified account of passive in German. In J. Nerbonne, K. Netter, and C. Pollard, editors, *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes. Center for the Study of Language and Information, Stanford, Ca.

Pollard, C. and Sag, I. A. (1987). *Information-Based Syntax and Semantics: Fundamentals*. Number 13 in CSLI Lecture Notes. Center for the Study of Language and Information, Stanford, Ca.

Pollard, C. and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. CSLI and University of Chicago Press, Stanford, Ca. and Chicago, Ill.

Przepiórkowski, A. (1994). Transmisja wymagań składniowych. ICS PAS Report 766, Polish Academy of Sciences, Warsaw.

Pullum, G. K. and Gazdar, G. (1982). Natural languages and context free languages. *Linguistics and Philosophy*, **4**, 471–504.

Saloni, Z. and Świdziński, M. (1985). *Składnia współczesnego języka polskiego*. Państwowe Wydawnictwo Naukowe.

Strutyński, J. (1993). *Zarys gramatyki polskiej: Morfologia*. Uniwersytet Jagielloński, Instytut Filologii Polskiej, Kraków.

Świdziński, M. (1993). Dalsze kłopoty z bezokolicznikiem. In J. Sambor, J. Linde-Usiekniewicz, and R. Huszcza, editors, *Językoznawstwo synchroniczne i diachroniczne*, pages 303–314. Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Timberlake, A. (1986). Hierarchies in the genitive of negation. In R. D. Brecht and J. S. Levine, editors, *Case in Slavic*, pages 338–360. Slavica Publishers Inc.

Willim, E. (1990). On case-marking in Polish. *Papers and Studies in Contrastive Linguistics*, **XXV**, 203–220.

Zabludowski, M. (1989). Quantificational noun phrases in Polish.