THE PUZZLE OF CASE AGREEMENT BETWEEN NUMERAL PHRASES AND PREDICATIVE ADJECTIVES IN POLISH

Adam Przepiórkowski and Agnieszka Patejuk
Institute of Computer Science, Polish Academy of Sciences

Proceedings of the LFG12 Conference
Miriam Butt and Tracy Holloway King (Editors)

2012
CSLI Publications

http://csli-publications.stanford.edu/
Abstract

This paper addresses the optionality of case agreement between a numeral phrase in the subject position and its modifying or predicating adjectives in Polish: such adjectives either agree with the numeral or – apparently – reach into the numeral phrase and agree with the noun phrase within it. While previous analyses of this phenomenon postulated special agreement mechanisms, we account for the troublesome facts by assimilating Polish numeral phrases to coordinate structures.

1 Introduction

The puzzle addressed in this paper concerns the two agreement patterns exhibited in Polish constructions such as (1).

(1) Ostatnie dziesięć lat było fatalne / fatalnych.
    last.ACC ten.ACC years GEN was 3.SG.N terrible.ACC/GEN
    ‘The last ten years were terrible.’

On the assumption, justified below, that ostatnie dziesięć lat is an accusative subject in (1), the accusative form of the predicative adjective fatalne is expected, but the genitive form, fatalnych, is completely surprising – it appears as if the predicative adjective fatalnych looked into the subject numeral phrase to agree in case with the genitive noun lat within it.

To the best of our knowledge, all previous analyses, when they get the facts right, stipulate special agreement machinery to account for such constructions. In this paper, on the other hand, we propose an LFG account which does not assume any new agreement mechanisms. Rather, what is special about constructions exemplified in (1) is the internal structure of Polish numeral phrases, which is essentially the structure assumed in LFG analyses of coordination.

The outline of this paper is as follows. Section 2 justifies various assumptions concerning Polish numeral phrases in the subject position. Then, Section 3 presents the problem at hand in a more detailed way and Section 4 provides an LFG analysis. Finally, Section 5 sketches previous attempts at dealing with this problem and Section 6 concludes the paper.

2 Assumptions

When talking about Polish numerals, we restrict ourselves to a morphosyntactically defined subclass of cardinal numerals, i.e., those which have number (always plural) and inflect for case and gender. So, for example, ordinal numerals, which inflect for number, are excluded from our considerations.

Such cardinal numerals may agree in case with the accompanying noun or they may require the noun to occur in the genitive case. In Polish linguistics, a new category, accommodability, was introduced (Bień and Saloni 1982) to distinguish
between the two classes of numeral forms: the value of this category is $\text{congr}$ in case of agreeing numerals and $\text{rec}$ in case of numerals governing the genitive case. Roughly speaking, governing numeral forms are:

- all forms of non-paucal (5 and above) numerals and some forms of paucal (2–4) numerals
- which occur in – loosely speaking – accusative and apparently nominative positions.\(^1\)

The pattern exemplified in (1) involves governing numerals in subject positions. It is completely uncontroversial that such numeral phrases are in fact subjects: they bind anaphors, participate in control constructions and may be coordinated with nominative subjects.

It is slightly less clear that they are really headed by the numeral. One argument is that only the noun may be elided in such constructions, as shown below:

\[(2)\]
\[
a. \ \text{Pięć kobiet stało.} \\
\text{five.f women.f.gen stood.3.sg.n} \\
\text{‘Five women were standing.’}
\]
\[
b. \ \text{Pięć stało.} \\
\text{five.f stood.3.sg.n} \\
\text{‘Five were standing.’}
\]
\[
c. \ *\text{Kobiet stało.} \\
\text{women.f.gen stood.3.sg.n} \\
\text{‘Women were standing.’} \quad \text{(putative)}
\]

Another argument is that phrases such as $\text{ostatnie dziesięć lat}$ (in (1)) and $\text{pięć kobiet}$ (in (2)a) cannot occur in genitive positions, which would be surprising if the genitive noun were the head. On the other hand, they can occur in accusative positions, which is expected if the accusative numeral is the head.

This brings us to the last assumption to be introduced here, namely, that such numeral subjects are in fact accusative, not nominative.\(^2\) This has been noted repeatedly at least since Małecki 1863,\(^3\) and is defended at length in Przepiórkowski 1999 (within HPSG). One argument comes from examples like (3)–(4), where the sudoku-like puzzle posed by the syncretic case forms may be satisfactorily solved only if the numeral bears the accusative case and agrees with the accusative feminine ($\text{f.acc}$) $\text{te}$ in (3) and the accusative human-masculine ($\text{hm.acc}$) $\text{tych}$ in (4).

---

\(^1\)It is largely a matter of convention whether numeral forms in genitive positions, co-occurring with genitive nouns, should be analysed as agreeing or governing, but see the HPSG analysis of Przepiórkowski 1999, §5.3.1.3, which implies that some such occurrences of genitive numerals should be analysed as governing, and others as agreeing.

\(^2\)We leave unresolved the question of whether the numeral has a nominative form at all. In the XLE implementation of the grammar of Polish, Patejuk and Przepiórkowski 2012, we assume that such governing numerals are defective and do not have a nominative form.

\(^3\)Other (numerous) bibliographical references are provided in Przepiórkowski 2004.
‘These five women were standing.’

‘These five men were standing.’

Tych in (3) is also possible and reflects the agreement with the genitive noun; this implies that (4) is structurally ambiguous, with tych agreeing either with the accusative numeral or with the genitive noun. Crucially, the common assumption that Polish nominal subjects are always in the nominative is contradicted by the ungrammaticality of (4) with the unambiguously nominative ci. Hence, in the two examples above, the possible case values are those marked in bold.

The conclusion that pięć in (3) and pięciu in (4) are accusative is corroborated by the default (non-agreeing) features on the verb, which patterns with non-nominative subjects in other Indo-European languages, e.g., Icelandic, see (5):

‘The boys lack food.’

Also, in the process of numeralisation, it is the accusative form of the noun that becomes a numeral, as in (6), where the nominative noun masa in (6a) enters into the subject-verb agreement, but the accusative masę in (6b) shows the default agreement patterns typical of numeral subjects.4

3 Problem

The problem of dual agreement between a numeral phrase and its modifier, exemplified in (1), is not limited to predicative adjectives. In fact, (3) above shows similar duality with respect to an attributive adjectival form: accusative te and genitive tych.

In case of attributive modifiers the problem does not seem acute, as an explanation for the preposed tych in terms of simple word order rules within an NP or a

4Note that there is no separate part of speech quantifier in the repertoire of grammatical classes assumed here. Semantic quantifiers may be expressed as numerals and nouns, among other parts of speech.
Numeral Phrase is in principle possible. However, as already shown in (1), there are analogous facts involving predicative adjectives that do not seem to be amenable to such an NP-internal discontinuity analysis; see the attested examples below:5

(7) Następne kilkadziesiąt metrów było czyste.
next.acc several tens.acc metres.gen was.3.sg.n clean.acc
‘The next few tens of metres were clean.’

(8) Pięć osób zostało rannych.
five.acc persons.gen became.3.sg.n wounded.gen
‘Five people were wounded.’

(9) Kolejnych jedenaście zarzutów było podobnych.
further.gen eleven.acc charges.gen was.3.sg.n similar.gen
‘Further eleven charges were similar.’

While (7) illustrates the expected agreement between the accusative numeral phrase and the accusative predicative adjective 

czyste, (8)–(9) are totally unexpected: it seems as if the genitive predicative adjectives rannych and podobnych can reach into the subject numeral phrase and agree with the genitive noun within it.

Note that both in (7) and in (9) the attributive adjective and the predicative adjective bear the same case: accusative in (7) (następne, 
czyste) and genitive in (9) (kolejnych, 
podobnych). Examples following these acc/acc and gen/gen patterns may be easily found in the corpus, and the same holds for the acc/gen pattern, illustrated in (10). While rarer, the gen/acc pattern is also attested, cf. (11).6

(10) Kolejne pięćdziesiąt aut zostało uszkodzonych.
further.acc fifty.acc cars.gen became.3.sg.n damaged.gen
‘Further fifty cars became damaged.’

(11) Minionych dwanaście miesięcy było najgorsze w historii.
past.gen twelve.acc months.gen was.3.sg.n worst.acc in history
‘The past twelve months were the worst in history.’

It must be stressed that the possibility of dual agreement concerns only morphosyntactically numeral phrases headed by a governing numeral. The example (1) above, repeated as (12) below, should be contrasted with (13) and (14).

---

5While examples (3)–(4) are constructed (but uncontroversial), the following examples – with the exception of (16) from Kallas 1974 – are based on attested uses found in the National Corpus of Polish (http://nkjp.pl/).

6The query [pos=adj] [pos=num] [pos=adj]* [pos=conj] [pos=conj] [pos=adj] (było | zostało) [pos=adj] (see http://nkjp.pl/poliqarp/help/en.html for the query syntax) on the complete 1.8-billion-segment National Corpus of Polish gives 113 results, including the following numbers for the total of 34 true positives: 23 for the acc/acc pattern, 5 for acc/gen, 4 for gen/gen and 2 for gen/acc. Other examples may be found by changing the word order, considering adjectival participles instead of adjectives, etc.
The last ten years were terrible.

‘A dozen drops were prescribed.’

‘Something like this is needed.’

Example (13) involves *tuzin*, a form of a lexeme which has a numeral meaning but is a morphosyntactic noun here, as evidenced by the singular masculine agreement with the verb. Unlike in (12), only the adjectival form agreeing with *tuzin* is possible. Similarly, in (14), involving a special construction consisting of the indefinite non-human pronoun *coś* and an NP-internal genitive modifier, the external predicative adjective *potrzebne* must agree with the pronoun and, hence, cannot occur in the genitive.

4 Analysis

4.1 Idea

The main pre-theoretical idea of the analysis is that Polish numeral phrases of the kind considered here are somewhere between being single-headed, like typical NPs, and somewhat multi-headed, like coordinate structures. In other words, they seem to be 1.5-headed – single-headed for the purpose of being assigned case (they may occur only in accusative – not genitive – positions), but bi-headed for the purpose of agreement (they may agree with accusative and genitive adjectives).

In terms of LFG, we postulate that such phrases, like coordinate structures, are represented by a hybrid feature structure (Dalrymple and Kaplan 2000), where the numeral and the noun (with any immediate modifiers) are elements of the set encoded by such a structure. However, unlike in the case of coordination, one of the elements of this set, representing the numeral, is at the same time the whole hybrid structure, i.e., the relevant Polish numeral phrases are represented by cyclic structures of the form given schematically in (15).

\[\text{tuzin} \text{kropel} \text{był} \text{przepisywany} / \text{*przepisywanych.} \]

‘A dozen drops were prescribed.’

\[\text{coś} \text{takiego} \text{jest} \text{potrzebne} / \text{*potrzebnego.} \]

‘Something like this is needed.’

In fact, just as many other nouns of this kind, *tuzin* may also be used numeratively, in which case the genitive form of the adjective is possible:

\[(i) \text{tuzin kropel było przepisywanych.} \]

‘A dozen drops were prescribed.’

Note that, due to the nominative/accusative syncretism of many inanimate masculine nouns in Polish, the nominative noun in (13) and the accusative denominal numeral in (i) have the same form: *tuzin*. 
4.2 Empirical consequences

The real advantage of this analysis over alternatives mentioned in Section 5 is that nothing special needs to be said about the accusative / genitive optionality of case agreement with predicative adjectives: the accusative case marking on the predicative adjective, as in (7) or (11), represents agreement with the accusative numeral phrase, while the genitive marking in (8)–(10) is handled by whatever mechanism is responsible for the single conjunct agreement (e.g., Kuhn and Sadler 2007), a phenomenon which independently occurs in Polish (Kallas 1974):

(16) Pachniał wiatr i morze.
    smelled.sg.m wind.sg.m and sea.sg.n
    ‘Wind and sea smelled.’

Also, the analysis naturally extends to agreement with attributive adjectives, without the need to assume discontinuous structures of numeral phrases in the relevant variants of (3)–(4) or in (9) and (11).

One potential problem for this analysis is that, to the best of our knowledge, Polish linguistic literature only reports cases of closest conjunct agreement, as in (16), while some of the examples involving attributive adjectives, namely those just referred to, as well as some examples involving predicative adjectives, as in (17), involve agreement with the furthest element.

(17) Niezbędnych było dobre parę metrów kwadratowych wykładziny.
    indispensable.gen was3.sg.n good.acc couple.acc meters.gen square.gen carpet.gen
    ‘A good couple square meters of carpet were indispensable.’

However, cases of furthest conjunct agreement may be readily found in the National Corpus of Polish (see (18)), so we conclude that Polish syntax makes this option independently available.\footnote{We crucially assume here that case is not a distributive feature; cf. Przepiórkowski and Patejuk 2012.}

\footnote{On the other hand, it should still be explained why this option is much more readily available in case of agreement with numeral phrases than in case of agreement with true coordinate structures; we leave this for future research.}

\n
Given such a structure, any case assignment or case checking mechanisms will target the whole feature structure, whose case is structure-shared with the numeral element, i.e., accusative (in constructions considered here).\footnote{We crucially assume here that case is not a distributive feature; cf. Przepiórkowski and Patejuk 2012.}
A prospective defeat or draw would cost them the leader’s yellow jersey.

4.3 Technical details

4.3.1 Independent assumptions

For the LFG analysis of Polish case assignment (Patejuk and Przepiórkowski 2011), largely carried over from the previous HPSG analysis (Przepiórkowski 1999), we assume the distinction between structural and lexical case assignment. Lexical case assignment happens in the lexicon; for example, the verb pomagać ‘help’ lexically specifies its complement to be dative, and it will remain dative regardless of the structural configuration or the form of this verb. On the other hand, the verb wspierać ‘support’ specifies its object as structurally case assigned; it will normally be accusative, but it will be realised as genitive in the syntactic scope of negation or when the form of the verb is gerundial.

For structurally case assigned (“sc = +”) subjects, we assume a simple statement which says that the case of such subjects depends on their category: normally it is nominative, but when the subject is a numeral phrase headed by a governing numeral (“acm = rec”; cf. the accommodability category introduced in Section 2 above), it is accusative.

Technically, we formalise this statement as two implications:¹⁰

\[\begin{align*}
  &a. \quad (\uparrow \text{subj}) : (\text{sc} = c + \land \text{acm} \neq \text{rec} \rightarrow \text{case nom} = +) \\
  &b. \quad (\uparrow \text{subj}) : (\text{sc} = c + \land \text{acm} = c \text{ rec} \rightarrow \text{case acc} = +)
\end{align*}\]

Note that here and henceforth, unlike in the schematic (15) above, we assume the representation of case proposed in Dalrymple et al. 2009. According to this representation, and given the 7 morphological cases in Polish, an unambiguously nominative noun phrase such as ewentualna porażka in (18) will have the case value as shown in (20) below (instead of the atomic nom):

\[
\begin{bmatrix}
\text{NOM} & + \\
\text{ACC} & - \\
\text{GEN} & - \\
\text{DAT} & - \\
\text{INST} & - \\
\text{LOC} & - \\
\text{VOC} & -
\end{bmatrix}
\]

¹⁰See Przepiórkowski and Patejuk 2012 for details and justification.
More importantly, as independently justified at length in Przepiórkowski and Patejuk 2012 (in these proceedings), we assume a more subtle approach to the distributivity of features, where it is not features that are distributive, but statements.\footnote{By default, all statements are distributive; non-distributive ones are marked explicitly.} In particular, each of the two implications in (19) must hold for each element in a hybrid feature structure separately.

In case of the numeral element, (19a) applies vacuously, because the antecedent is false ("\(\text{ACM} \neq \text{REC}\)" is false, as the numeral is governing), and (19b) applies non-vacuously and assigns the accusative case.

In case of the nominal element, both clauses apply only vacuously, because the noun is not structurally case marked (so "\(\text{sc} = \_+\)" is false) – instead, the genitive is checked by the relevant numeral phrase rule (see (21a) in §4.3.2 below).

Note that, since the numeral element is structure-shared with the whole hybrid feature structure, the numeral phrase as such is unambiguously accusative. But since we assume that case is not a distributive feature (no feature is distributive by itself), this accusative case does not distribute to the nominal element, so there is no feature clash.

### 4.3.2 Structure of numeral phrases

After introducing these independently needed assumptions, the only special part of the analysis is the c-structure rule (21a), which gives rise to cyclic f-structures such as (21b), headed by the governing numeral, containing a genitive NP and occurring in structurally case marked positions.

\[(21) \quad \text{NumP} \rightarrow \text{Num NP} \]

\[\quad \quad (\downarrow \text{ACM}) =_{c} \text{REC} \quad (\downarrow \text{CASE GEN}) =_{c} +
\]

\[\quad \quad (\downarrow \text{sc}) =_{c} + \quad \downarrow \in \uparrow
\]

\[\quad \quad \uparrow = \downarrow
\]

\[\quad \quad \downarrow \in \uparrow
\]

\[b. \quad \begin{bmatrix}
\text{ACM} & \text{REC} \\
\text{SC} & +
\end{bmatrix}
\]

\[\quad \quad \begin{bmatrix}
\text{CASE} & [\text{GEN} +]
\end{bmatrix}
\]

When such a structure is the value of subj, the statement (19b) ensures that the numeral phrase is in the accusative case, as explained in §4.3.1.

Let us note that cyclicity is not a frequent feature of LFG analyses, but a very similar structure is proposed in Fang and Sells 2007, p. 209, to account for the Chinese verb copy construction:\footnote{Fang and Sells 2007, fn. 6, attribute this solution to Ron Kaplan, whom we would like to thank for pointing out to us the similarity between the two analyses.}
While cyclicity is never mentioned in Fang and Sells 2007, rule (22) gives rise to cyclic structures in the same way as (21a) does. Fang and Sells (2007), who assume the usual approach to the distributivity of features, use such cyclic structures to ensure that all non-initial VPs in the construction at hand “inherit” all relevant features from the first VP.

5 Previous attempts

It might be tempting to analyse numeral phrases as bi-headed or as structurally ambiguous, i.e., alternatively headed by the numeral and the noun. Such accounts are considered and rejected on various grounds in Przepiórkowski 2001 and we will not repeat this discussion here.

The alternative analysis proposed there assumes instead that the genitive noun is the subject of the numeral and that subjects are “visible” outside of their phrases.13 Given this assumption, case agreement between a predicative adjective and its subject (structure-shared with the numeral subject of the copula) is formalised via a disjunctive constraint stating that the adjective agrees either with its subject (the accusative numeral phrase) or its subject’s subject (the genitive noun).

Formally, case agreement is invoked in the two principles (23)–(24), and it is defined in (25).

(23) Attributive case agreement:
\[
\begin{array}{c}
\text{head} \\
\text{CASE} \text{[} \text{[} \text{cat|head|case} \text{]} \text{]} \\
\text{MOD|LOC} \text{[} \text{[} \text{cat|head|case} \text{]} \text{]} \\
\end{array}
\rightarrow \text{case-agreement(} \text{[} \text{[} \text{cat|head|case} \text{]} \text{]}, \text{[} \text{[} \text{cat|head|case} \text{]} \text{]} \text{)}
\]

(24) Predicative case agreement:
\[
\begin{array}{c}
\text{category} \\
\text{HEAD} \text{[} \text{[} \text{case} \text{]} \text{]} \\
\text{PRD} + \\
\text{SUBJ} \text{[} \text{loc} \text{[} \text{cat|head|case} \text{]} \text{]} \\
\end{array}
\rightarrow \text{case-agreement(} \text{[} \text{[} \text{cat|head|case} \text{]} \text{]}, \text{[} \text{[} \text{cat|head|case} \text{]} \text{]} \text{)}
\]

(25) Definition of case agreement:
\[
\text{case-agreement(} \text{[} \text{[} \text{cat|head|case} \text{]} \text{]}, \text{[} \text{[} \text{cat|head|case} \text{]} \text{]} \text{)} \leftrightarrow
\left( \text{[} \text{[} \text{cat|head|case} \text{]} \text{]} \lor \left( \text{[} \text{[} \text{cat|head|case} \text{]} \text{]} \land \left( \text{[} \text{[} \text{cat|head|case} \text{]} \text{]} \lor \ldots \right) \right) \right)
\]

13While this assumption is natural in LFG, it is controversial on the strong view of locality held in HPSG, but see Sag 2007 for discussion.
In particular, according to (24)–(25), case agreement between a case-bearing phrase and its predicative head means the structure sharing of case values of the predicative element on one hand and either the phrase’s case or the case of the phrase’s subject (initial element on its ARG-ST list) on the other, the latter taking place only when the phrase and its subject have the same index value (as arguably the numeral and its noun complement do).

Although this HPSG analysis may be carried over to LFG, and it still seems empirically adequate, it is rather unsatisfactory in the sense that, in order to explain a very specific construction involving subject numeral phrases, it proposes a more complicated (and disjunctive) general mechanism of case agreement. The current analysis seems to be theoretically more satisfying, as it provides an equally empirically adequate account in terms of a rather special structure of relevant numeral phrases, and leaves general agreement mechanisms untouched.

6 Conclusion

Although this paper deals with a very special parochial construction of Polish, the analysis is based on a couple of ideas that may be of a broader theoretical interest. First of all, we propose to extend the use of hybrid structures, previously employed in analyses of coordination, to the representation of a class of numeral phrases in Polish. Second, the analysis relies on the possibility of case being non-distributive. Third, we show how relevant case assignment and agreement facts can be dealt with by the assumption that hybrid feature structures representing numeral phrases are cyclic, i.e., that one of the elements of the set represented by the hybrid structure is the structure itself.

It needs to be noted, however, that these three aspects of the analysis have a very different standing. The assimilation of a class of numeral phrases to coordinate structures is crucial for the current analysis, as it makes it possible to recycle the standard agreement mechanisms, including the single conjunct agreement. What is also crucial is the possibility that a hybrid feature structure may bear a case value different from that of one of the elements in the set represented by the feature structure, i.e., we assume that case is not a distributive feature – perhaps it is not features that are distributive, but statements, as argued in Przepiórkowski and Patejuk 2012 (in these proceedings).

On the other hand, cyclicity is not a necessary feature of the current account: what is crucial is that the morphosyntactic features of the numeral, especially its case, be the same as those of the complete numeral phrase, and this can be ensured by equating just the values of the relevant features, without equating the whole feature structures. Nevertheless, we believe that Polish numeral phrases are rather naturally modelled as cyclic feature structures and that the proposed analysis opens the question of the place of such cyclic structures in LFG.
Acknowledgements

This research is supported by the POIG.01.01.02-14-013/09 project which is co-financed by the European Union under the European Regional Development Fund. The publication of the resulting implemented grammar is supported by the CESAR project (European project CIP ICT-PSP 271022, part of META-NET).

The authors are grateful to Tracy Holloway King for valuable comments on the previous version of this paper.

References


Przepiórkowski, Adam. 1999. Case Assignment and the Complement-Adjunct Di-


