Three Kinds of 'Homogeneous' Patterns of Czech Numerals:
A Phrasal Spell Out Account[*]

Pavel Caha
Masarykova Univerzita, Brno

This paper is concerned by the case patterns found with Czech higher numerals. In a well-known paper on cognate Russian numerals, Babby (1985) established and provided theoretical grounding for a split into two distinct case patterns, which he called heterogeneous (in structural cases) and homogeneous (in oblique cases). This bifurcation is based on the case found on the counted noun: it is either a genitive (in the heterogeneous pattern) or a case that co-varies with the case of the whole noun phrase (in the homogeneous pattern). The distinction can be extended to a number of other Slavic languages (see esp. Franks 1994), including Czech.

   Against this background, I show that in Czech, the homogeneous pattern can be further sub-divided into three distinct patterns depending on the morphology of the numeral. The shape of the numeral either reflects the particular case (agreeing pattern), it is invariably nominative (phrasal case pattern), or it is invariably genitive (the genitive-as-oblique pattern).

   I show how these three patterns can be captured in the split-case system of Caha (2009), where individual case features head separate syntactic projections. A crucial part of the proposal is the use of phrasal

spell-out introduced in Starke (2009). I also make use of the possibility that case affixes attach to a whole syntactic phrase, rather than to each word.

1 The Homogeneous/Heterogeneous Pattern in Russian and Czech

In Russian, as well as in most other Slavic languages, higher numerals (≥ 5) have a special behavior. When phrases including such numerals are used in the subject or object position, the counted noun appears in the genitive plural, see (1). This pattern has been called 'the heterogeneous pattern,' because the case of the counted noun ('bottles') does not match the syntactic context (object) or the case of the numeral (accusative).

(1)   pjať   bolšix   butylok   vina (Russian, Babby 1985:2)
       five$_{ACC}$ big$_{GEN.PL}$ bottles$_{GEN.PL}$ wine$_{GEN}$
       'five big bottles of wine'

In oblique cases, like the instrumental, the counted noun drops its genitive and appears in the case appropriate for its syntactic/semantic role (2). This pattern has been called 'the homogeneous pattern' by Babby, because of the match between the case of the noun and the case of the numeral.

(2)   pjaťju bolšimi butylkami   vina   (Russian, Babby 1985:2)
       five$_{INS}$ big$_{INS.PL}$bottles$_{INS.PL}$ wine$_{GEN}$
       'five big bottles of wine'

Czech apparently exhibits the same split between the heterogeneous and the homogeneous patterns. In the object position, the counted noun is in the genitive plural (3a), which is just like (1). In the instrumental case, the counted noun drops its genitive, and appears in the instrumental (3b).

(3) a. pět   velkých flašek   vína
        five$_{ACC}$ big$_{GEN.PL}$ bottles$_{GEN.PL}$ wine$_{GEN}$
    b. pěti   velkýma flaškama   vína
        five$_{???}$ big$_{INS.PL}$ bottles$_{INS.PL}$ wine$_{GEN}$
        both: 'five big bottles of wine'

There is, however, a difference that has to do with the case borne by the numeral. If Czech was like Russian, the gloss in (3b) should have INS under the numeral 'five'; instead, I put there three question marks, because it is actually unclear what the gloss should be. Why is that so?

In both Czech and Russian, the numeral 'five' inflects like a noun of a particular type. For Russian, I show this in the first two columns of the table in (4). Comparing the declension of 'notebook' and 'five,' we see a perfect match in the endings. *Pjaťju* seen in (2) shows no quirk; it looks exactly like an instrumental. The data are taken from Timberlake (2004).

(4) *The declension of 'five' in Russian and Czech*

	notebook	five	bone	five
NOM	*tetrad'*	*pjať*	*kost*	*pět*
ACC	*tetrad'*	*pjať*	*kost*	*pět*
GEN	*tetrad'-i*	*pjať-i*	*kost-i*	*pět-i*
LOC	*tetrad'-i*	*pjať-i*	*kost-i*	*pět-i*
DAT	*tetrad'-i*	*pjať-i*	*kost-i*	*pět-i*
INS	*tetrad'-ju*	*pjať-ju*	*kost-**i***	*pět-**i***

In Czech, however, the comparison shows that the instrumental ending of the numeral is different from the instrumental ending of the corresponding noun. They differ in vowel length, which is distinctive in Czech.

This quirk is not limited to 'five.' The numerals 'six,' 'seven,' 'eight,' 'nine,' 'ten' and 'thousand' all follow the same paradigm as 'five.' Also the partly irregular compound numerals which contain 'ten' as the head (i.e., 'twenty,' 'thirty,' …) show the same peculiarity: there is something odd with the instrumental of Czech numerals in general.

Should we ignore this difference, replace the question marks in (3b) by INS, and treat the unexpected form as an arbitrary morphological noise? Or is the replacement of –í by –i significant beyond pure morphology, suggesting that Czech and Russian differ in the syntax of the 'homogeneous pattern?' In the following sections, I argue that the latter is the case. Specifically, I will be lead to conclude that the –i found on the Czech version of 'five' is actually a genitive marker throughout the oblique paradigm; a fact that only becomes obvious in the instrumental because of the observed contrast with the corresponding noun. In other words, the question marks in (3b) should be replaced by GEN. Why?

2 The Agreeing Pattern and the Phrasal-Case Pattern

The main reason for such a claim is that it can be shown independently of 'five' that in Czech, there is a problem for the INS gloss in (3b). This can be seen when we look at the numeral 'hundred.' This numeral can enter into three distinct patterns of case marking, which are most clearly distinguished in the oblique cases. I will now demonstrate the three different patterns on the example of the dative.

   The first option of expressing a phrase 'to hundred crowns (the Czech currency)' is to say (5a), which I call the nominal pattern. In this pattern, 'hundred' behaves like a noun: it bears dative and its complement is in the genitive. In (5b, c), the counted noun drops the genitive and bears the case appropriate for the syntactic role of the whole phrase; a hallmark of the homogeneous pattern. It is hard to say whether the patterns differ in meaning, and if yes, how.

(5) a. ke   st-u   korun
         up to 100_{DAT} $CZK_{GEN.PL}$
    b. ke   st-u   korun-ám
         up to 100_{DAT} $CZK_{DAT.PL}$
    c. ke   st-o   korun-ám
         up to 100_{NOM} $CZK_{DAT.PL}$

The two last examples differ in the form of the numeral. In (5b), the numeral is dative just like the noun; but in (5c), the numeral looks as if it was in the nominative (*sto korun* is the nominative of '100 CZK').[1]

---

[1] A reviewer points out that numerical phrases sometimes behave differently after a preposition and after a verb, wondering whether all the patterns can be used after a verb. The answer is yes; after a dative selecting verb (e.g., *blížit se* 'approach,' as in 'the price approaches 100 CZK'), all the three patterns are still grammatical.

It can be shown that the numeral in (5c) not only looks like nominative, it also triggers nominative agreement on modifiers. For instance, numerals like 'three' generally agree with the counted noun in Czech. When such a numeral is added to a structure like (5b), we get (6a); 'three' is dative, copying expectedly the case of 'hundreds.' When the numeral 'three' modifies the suspected nominative seen in (5c), the agreeing numeral is also nominative, see (6b).

(6) a. k   (těm)   tř-em   st-ům   korun-ám
         to   those$_{DAT}$ three$_{DAT}$ hundreds$_{DAT}$ CZK$_{DAT.PL}$
     b. k   (těm)   tř-i   st-a   korun-ám
         to   those$_{DAT}$ three$_{NOM}$ hundreds$_{NOM}$ CZK$_{DAT.PL}$
         'to (those) three hundred crowns'

A reviewer asks what happens when a demonstrative is added between the preposition and the numeral. We see that it is dative in both (6a) and (6b). This can be explained under the assumption that the structure of the example is [those [*[three hundred]* CZK]], and nominative is only available for elements inside the projection of 'hundred' (in italics).

To sum up: *sto* 'hundred' enters into three patterns: a nominal pattern and two types of the homogeneous pattern. In one of its versions (in 5b), the numeral has the same case as the counted noun. Therefore, I call it the *agreeing* pattern. In (5c), the numeral is nominative, and the dative case is only expressed once in the phrase consisting of the numeral and the noun. This is reminiscent of languages with so-called phrasal case marking; therefore I call this the *phrasal-case* pattern, and treat it as an instance of the homogeneous pattern, even though there is no homogeneity in case. The reason for using the term 'homogeneous' is to preserve the traditional term and provide further distinctions within the established terminology.

The difference between the agreeing and the phrasal-case pattern is relevant for processes that are known to correlate with that distinction. For instance, in languages with agreement, it is generally possible to elide the noun with its affixes, leaving a modifier in place (see, e.g., Lipták and Saab 2014). This is possible in (6a), yielding (7a). However, in languages where modifiers do not agree, such an ellipsis is impossible. Correlating with that is the observation that ellipsis is impossible in the phrasal case pattern (6b), yielding the ungrammatical (7b). Including or

omitting the demonstrative is orthogonal to this split behavior; I come back to this later on.

(7) *How much did it cost?*
   a. někde   k   (těm)   tř-em   st-ům   ~~korun-ám~~
       somewhere to those_DAT three_DAT hundreds_DAT CZK_DAT.PL
   b. ?*někde   k   (těm)   tř-i   st-a   ~~korun-ám~~
       somewhere to those_DAT three_NOM hundreds_NOM CZK_DAT.PL
       'somewhere around three hundred'

(6) and (7) then show that the issue here is not that the expression of the dative case is variable (*st-o* or *st-u*). Instead, it seems that what varies is the case borne by 'hundred', with at least two different options available within a single grammar.

3 The Agreeing Pattern in the Instrumental

With the three way distinction between the nominal, agreeing and phrasal-case patterns in place, let me turn to the behavior of the three patterns in the instrumental case. The expected forms are below:

(8) a. se   st-em   korun
        with 100_{INS} CZK_GEN.PL
    b. *se   st-em   korun-ama
        with 100_{INS} CZK_INS.PL
    c. se   st-o   korun-ama
        with 100_{NOM} CZK_INS.PL
        all 'with hundred CZK'

(8a) shows the nominal pattern, with instrumental on the numeral and genitive plural on the counted noun. (8c) shows the phrasal-case pattern, with nominative on the numeral, and instrumental on the noun.

   A strange thing happens when we look at the agreeing pattern. From (5), we know that the agreeing pattern arises by combining the cased form of the numeral and the 'homogeneous' form of the noun. When we construct an example with these properties in the instrumental (i.e., 8b), it turns out that it is ungrammatical. The purely descriptive conclusion is that 'hundred' lacks the agreeing pattern in the instrumental case. It is

impossible to have the instrumental case both on the noun and the numeral.

This should be compared to the fact noted in section 1, namely that many numerals lack the expected instrumental form altogether. The point is that this now seems to be just an instance of the same pattern that we observe for 'hundred,' only more general: in Czech, it is impossible to have the instrumental expressed simultaneously on both a higher numeral and the noun. The numeral has to be in a different case (nominative for 'hundred').

4 The Genitive-as-Oblique Pattern

But if *pět-i* in (3b) is not an INS (because Czech in general does not allow INS both on the numeral and the noun), then what is it? 'Hundred' shows up in the nominative form *st-o*; but *pět-i* 'five' cannot be analyzed this way, because it is not identical to the nominative form *pět*. Looking at other quantifiers sheds some light on this. Consider, for instance, *mnoho* 'many,' see (9). This item has a paradigm where *–a* stretches from GEN to INS. This is the same context where we find *–i* with 'five.'

(9) *'town' : 'many' = 'bone' : 'five'*

	city	many	bone	five
NOM	*měst-o*	*mnoh-o*	*kost*	*pět*
ACC	*měst-o*	*mnoh-o*	*kost*	*pět*
GEN	*měst-**a***	*mnoh-**a***	*kost-**i***	*pět-**i***
LOC	*měst-ě*	*mnoh-**a***	*kost-i*	*pět-**i***
DAT	*měst-u*	*mnoh-**a***	*kost-i*	*pět-**i***
INS	*měst-em*	*mnoh-**a***	*kost-í*	*pět-**i***

The difference between 'many' and 'five' is that they belong to different declensions. Whereas 'five' inflects much like 'bone,' 'many' draws its case markers from the declension of the noun 'city' (at least in NOM to GEN). That is relevant, because 'city' (unlike 'bone') has a highly differentiated paradigm – all the oblique cases are distinct. Because of this, we can see that the invariant oblique ending *–a* corresponds to the genitive *–a* of 'city.' This pattern leads to the idea that 'five' (just like 'many') appears in all the oblique case environments bearing an ending of the genitive case. In sum, the claim is that all the boldfaced endings in

the table should be understood as true genitive endings, and not as a surface conflation of distinct feature structures (syncretism). In plain language, the numeral is in the genitive case when it counts a noun in the instrumental.

I will call this pattern the 'genitive-as-oblique pattern,' because the numeral has in its paradigm only two case forms, tracking the division structural/oblique, with the oblique category expressed by the genitive.

I admit that the actual forms of 'five' are in principle open to other interpretations (the genitive form is only in the instrumental, the remaining forms are in the appropriate case; the form in the instrumental is not genitive but dative, etc.), but the overall system of Czech does not lend much support to such alternatives, as far as I am aware. On the other hand, interpreting the pattern of 'five' as a pattern where the genitive form spreads to other cases is independently supported by the pattern of 'many.'

If this way of understanding the facts is on the right track, we are led to the conclusion that in Czech, the homogeneous pattern really sub-divides into three distinct sub-patterns, depending on the case of the numeral. The numeral may either bear the same case as the noun (the agreeing pattern), it may be invariably nominative (the phrasal-case pattern) or it may be invariably genitive (the genitive-as-oblique pattern). The question I ask in the remainder of the paper is how to analyze these patterns, adopting certain independent proposals about case.

5 Case Decomposition

In many frameworks, it is believed that individual cases are not primitive entities, but internally complex objects that can be decomposed into features. I am also going to rely on case features in providing one possible explanation for the existence of the three distinct patterns.

In prototypical case-decomposition theories going back to Jakobson's pioneering work, researchers rely on binary features such as +/- structural, +/- oblique, etc. (see Niedle 1988, Franks 1995, Halle 1994 among many others). However, there are reasons to think that decomposition into binary features is not the best way to represent (minimally Slavic) case. McCreight and Chvany (1991) and Johnston (1997) (see also Baerman et al. 2005) point out that such a system is inappropriate to capture certain restrictions on syncretism. Taking

Russian as their paradigm language, they observe that syncretism is
restricted by a linear contiguity constraint, such that only adjacent cases
in a linear sequence show syncretism. For Russian, they state the
sequence NOM-ACC-GEN-LOC-DAT-INS. Caha (2009:ch.8) argues
that the same restrictions apply in Czech. All the works cited in this
paragraph conclude that such a constraint, if true, cannot be captured by
binary features. As Caha (2013:1027-1028) sums up the observations,
any approach based on binary features either under-generates or
massively over-generates (to the extent that anything goes), depending
on how the details are set.

Various alternatives to binary features have been proposed, but it is
impossible to do the justice to this literature here. Instead, I directly turn
to the alternative explored by Caha (2009), where binary features are
replaced by syntactically arranged privative features, such that individual
cases stand in a containment relationship to other cases, as dictated by
the syncretism sequence observed first by McCreight and Chvany
(1991):

(10) *Case containment*
       a.   NOM   =   [A]
       b.   ACC   =   [B [A]]
       c.   GEN   =   [C [B [A]]]
       d.   LOC   =   [D [C [B [A]]]]
       e.   DAT   =   [F [D [C [B [A]]]]]
       f.   INS   =   [E [F [D [C [B [A]]]]]]

With this proposal in place, the adjacency/contiguity restriction on
syncretism can be easily derived, see Caha (2009, 2013). Details of the
proposal aside, it builds on similarity between cases. For instance, if
GEN and NOM are treated as similar enough for a spell-
out/pronunciation by one marker (they both share A and lack D), the
relevant notion of 'similarity' must also include ACC (which also has A
and lacks D). This way, contiguity is guaranteed for any syncretic
relation between cases.

Note as well that Caha builds on Starke's (2009) proposal that
individual morphemes may correspond to a whole phrase, such that a
single instrumental marker may spell out all the projections in (10f), each
projection containing just a single feature; see Caha's work for details.

The proposal in (10) – intended to capture syncretism patterns – can shed light on the peculiar properties of the three distinct patterns, a proposal I spell out below in the following sections. My goal is to avoid construction specific proposals, or special morphological rules known from Distributed Morphology. For instance, within DM, one can imagine that a structure with full concord is subject to various degrees of Impoverishment, yielding the genitive-as-oblique pattern if Impoverishment deletes all features but A,B,C on the numeral; or the phrasal case pattern, if Impoverishment deletes all case features. Such an approach seems to me overly simplistic in that it loses (*prima faciae*) a handle on the fact that the distinctions in agreement correlate with syntacic distinctions, such as ellipsis. The three tools I will be using instead are (1) ellipsis, (2) variable pied-piping, and (3) whether concord is present or absent in the base-generated structure.

6 Two Cases, One Host: Case Attraction in Numerical Phrases

There are many theories of numerals in the literature, and I cannot do any justice to them here. Instead, I start directly by providing one possible explanation for how the homogeneous pattern arises.

I start from the observation that in a number of languages, we can find constructions where something looking like case competition takes place. For instance, there are languages where in free relative clauses (*He ate what John cooked*), the relative pronoun (*what*) can in principle bear either the case assigned by the matrix verb (*He **ate** what*), or by the embedded verb (*what John **cooked***). But despite there being two potential candidates for assigning case, the pronoun only appears with one case. Which case that turns out to be (the matrix or the embedded one) is governed by a set of rules, which partly differ from language to language.

What is important is that in many cases, the set of rules governing such a competition incorporates something looking like a case hierarchy. In Gothic, for instance, Harbert (1983) observes that if any of the two verbs assigns DAT, the realization of this case takes precedence over all other cases (GEN, ACC, NOM). This phenomenon is called case attraction, and we say that DAT attracts GEN/ACC/NOM. Looking at the competition in other cases, Harbert reports that in the hierarchy NOM-

ACC-GEN-DAT, it is always the case on the right that attracts cases on the left.

The standard term for such an effect is 'markedness.' The most marked case is realized, the less marked case is left out. The decomposition in (10) allows us to understand this 'markedness' effect as a regular instance of ellipsis. When two sets of features appear on one host, say both GEN and DAT, but only one set may be realized, the other set must be eliminated by ellipsis. Ellipsis (here as elsewhere) relies on recoverability: the content of the elided material must be recoverable on the basis of an antecedent. Notice that (10) provides an independently motivated proposal that determines which case may be recovered: NOM (A) may be recovered on the basis of ACC (A,B), which in turn may be recovered on the basis of GEN (A,B,C), and so on. Hence, when GEN and DAT meet, GEN is eliminated and DAT is expressed. The case with more features survives.

Caha (2015) extends this logic to Czech numerical phrases. He proposes that in numerical phrases, the numeral is a nominal-like head (as previously proposed by Ionin and Matushansky 2006), and it always assigns the genitive to the counted noun. In order to reflect the proposal, I will be calling numerals also numerical nouns. Caha (2015) further proposes that in addition to the genitive, the counted noun is also marked by the same case as the numeral has (due to concord). This proposal is depicted below, where the counted noun has two cases, one assigned by the nominal numeral, and the other copied by concord. It is thus a type of an agreeing genitive construction, or a *Suffixaufnahme* structure, a term used in Plank (1995). This is important, since I want to avoid any construction specific proposal, and derive the numerical pattern through an interplay of processes and structures that are independently attested.

(11)   *The numeral construction:*
       [*Numeral*-CASE$_i$ [*Counted-Noun*-GEN-CASE$_i$]]

As in free relatives, only one case survives on the host, and the other case is eliminated. The decomposition in (10) ensures that it is always the case with more features that survives, and the other case is elided. In other words, in the sequence NOM-ACC-GEN-LOC-DAT-INS, cases to the right of GEN license its ellipsis, while cases to the left of GEN (NOM, ACC) are elided, and GEN preserved. There is thus no

construction specific process involved in the numerical construction; any theory of grammar has to provide for agreeing genitives and for ellipsis.

This theory provides an account of the split between the heterogeneous and homogeneous patterns. In the heterogeneous pattern, the concord case is elided and the genitive preserved. In the homogeneous pattern, the genitive is elided and the concord marker preserved. However, as it stands, the account only provides an explanation for how the agreeing pattern arises, with both the numeral and the counted noun marked for the particular case. The following sections explore how to capture the additional patterns. What I will suggest is that in order to capture these patterns, the theory expressed in (10) and (11) must be further enriched by two additional (and independently needed) parameters of variation.

7 The Host of Case Markers: the Agreeing Pattern vs. the Rest

The first parameter concerns the *host* for case marking. For example, in Turkish, case is expressed only once in the whole noun phrase. I will thus be saying that it is the whole extended NP that serves as the host. In Czech, case is marked on almost every member of the noun phrase; we see it on Ns, As, DEMs, NUMs and a subset of possessors. For Czech, I will thus be saying that (nearly) every extended projection in the NP serves as the host. Theoretically, I encode this in a way that the case marker combines with the particular constituent that correponds to the host: [K [*host*]].

I will then assume that for each language, a set of hosts for case marking will be determined. There are various technical ways of doing this. In works inspired by recent Chomsky's work (Chomsky 2001), this aspect of variation is handled by placing uninterpretable features at the appropriate places in the structure (or deriving this from the theory of categories, as in Baker 2008). In a DM inspired tradition (Bobaljik 2008), case is determined after the syntactic computation, and concord is handled by special rules at Morphological Structure (see, e.g., Norris 2014).

A third option, which I adopt here, is to assume that depending on the language, case features may be base-generated at one or multiple places inside the noun phrase. In concrete terms, I assume that in Czech, case features are base-generated inside the extended projection of the

adjective, yielding concord. Turkish or English lack case inside the extended projection of the adjective and no concord is observed.

In languages where such features are base-generated in the adjectival projection, they need to be checked Spec-Head (Chomsky, 1995). This is achived by treating adjectives as a type of a relative clause, such that a copy of the head noun is present inside the relative clause, and it is this copy which does the checking. I refer the reader in particular to Leu (2008) for a detailed proposal along these lines. Technicalities aside, what is crucial is that there is indisputable variation in what the host of case marking is. I will call this 'the host parameter.'

An important observation is that such a variation can also be present within a single language. In the interest of space, I have to skip examples, but one relevant languages to point at is East Slovak Romani. The important message is that the variation in what the host is, will be used here as one of the factors involved in the multiplicity of numerical patterns. In particular, when *both* the numerical noun and the genitive serve as hosts, the agreeing pattern arises. When only the whole phrase [Num N] serves as the host, the remaining patterns arise. We will need movement to distinguish further among the possibilities, but it is relatively easy to see that when the (boldfaced) case on the dependent genitive is left out from (12a), the nominal pattern arises in (12b):

(12)   *The numerical construction: the host parameter*
       a.   AGR pattern: [*Num*-CASE$_i$ [*Counted-Noun*-~~GEN~~-**CASE$_i$**]]
       b.   The nominal pattern: [*Num*-CASE$_i$ [*Counted-Noun*-GEN]]

It is worth noting that the two patterns in (12) are also implicated in the alternation between a plain genitive and an agreeing genitive construction (see Corbett 1995 for a discussion of such an alternation in Slavic), which relates to my goal that the analysis should not be construction specific.

Let me also make a note here concerning the strange fact that in Czech, the agreeing pattern (12a) is not available for numerals in the instrumental case, while it is available in DAT, LOC and GEN. A partial understanding of this fact comes from the observation that case is not a uniform category, and that languages tend to switch from concordial marking to phrasal marking somewhere along the hierarchy of cases in (10). For instance, in Icelandic, NOM, ACC, GEN and DAT are

expressed on every member of the NP (there is concord for these cases), while the meaning of an instrument is expressed only once (there is a preposition marking this meaning). The pattern of case marking in Czech – that is when it comes to the phrase [Num N] – apparently has a similar system, where GEN and DAT are expressed both on the noun and the numeral, while the meaning of an instrument is marked only once per phrase.

Finally, let me address a question raised by a reviewer concerning the pattern in (5). The pattern is repeated below. In this pattern, (5a-c) present three logical options of placing a dative marker in the string, either on the numeral, on the counted noun, or on both. (5d) corresponds to the fourth logical option of not placing it on either of them. A potential line of thinking is: we know that *sto* is compatible with dative environment (5c), and so is *korun* (5a); so we just combine these two and see what happens.

(5) a. ke   st-u   korun   c. ke   st-o   korun-ám
     up to 100_{DAT} $CZK_{GEN.PL}$ up to 100_{NOM} $CZK_{DAT.PL}$
   b. ke   st-u   korun-ám   d. *ke st-o   korun
     up to 100_{DAT} $CZK_{DAT.PL}$ up to 100_{NOM} $CZK_{GEN.PL}$

What happens is that the example is ungrammatical – and the question is why. The answer is that the logic which only looks at single words in the pattern is too simplex: we have to look at the phrase as a whole. If we do that, (5a,c) are just two different ways of expressing the dative case inside a complex phrase, an intuition I spell out in detail in what follows. (5d) is different: there is no dative at all here, and that is why the example is out as a phrase in an environment that requires dative (whatever its host).

8 Locating Case with respect to Host: the Nominal vs. Phrasal Pattern

I will now turn to the second parameter of variation, namely movement, and show how it may be used to further differentiate among the various patterns. Involving movement in the picture relates to a descriptive variation along the following lines: for a given host, how is case marking

realized on that host? Does it precede the host, does it follow the host, or is the marking super-imposed on the host (suppletion, ablaut, etc.)?

In Czech, the instrumental meaning is expressed by a marker that follows the host (where the set of hosts include the noun, the adjective and so on); we get *t-ím velk-ým kladiv-em* 'that-INS big-INS hammer-INS.' In Bulgarian, on the other hand, the host of case marking is the whole noun phrase (since the meaning of an instrument is expressed only once in the whole phrase), and the phrasal marker precedes the host; we get *s tozi guljam čuk* 'INS that big hammer.'

There are reasons to think that case (labelled K) is a separate funcional head, which takes its host as the complement (Bittner and Hale 1996). If that is so, the ordering between case and its host (whatever that host is) can be understood as a consequence of regular principles governing the ordering of heads and complements in the syntax.

The theory of ordering I adopt here falls in the camp of LCA based theories (Kayne, 1994, see also Cinque 2005). According to such theories, any functional head precedes its complement in the base structure. If nothing happens, the functional head will keep preceding the host, and we get a language like Bulgarian (13a). However, the complement of K may move to its left, placing the host to the left of the case marker, which in Czech happens with both nouns (13b) and adjectives (13c). The noun and the adjective are assembled together only after each of them crosses its K.

(13)   *The ordering of the host and the case marker*
       a. Bulgarian: [K [HOST]]   [s [tozi guljam čuk]]
       b. Czech: [HOST [K [~~HOST~~]]]   [kladiv [em [~~kladiv~~]]]
       c. Czech: [HOST [K [~~HOST~~]]]   [velk [ým [~~velk~~]]]

If this view is correct, we must update our structure for Czech numerical phrases. This will have little effect when case marking is placed on every member of the noun phrase; we will just have to move each host to the left of its case marker before assembling them together. However, in cases where the whole numerical phrase only has one case marker, different patterns arise depending on how movement proceeds. To show that, consider first the base structure (14a). Here we have the numerical noun (*Num*) taking a complement (*N*) in the genitive, and on top of this

constituent, we base-generate a single case marker (K) signalling the semantic/syntactic role of the phrase including the numeral and the noun.

(14)   a.   K [*Num* [*N-GEN*]] base structure
       b.   *Num* [K [~~*Num*~~ [*N-GEN*]]] the nominal pattern
       c. [*Num* [*N-GEN*]] [K ~~[*Num* [*N-GEN*]]~~]   the phrasal-case pattern

In Czech, case is a suffix, so something moves to the left of K. In Czech, there are two options. Either the head of the complement moves, and we get the structure (14b). Here the case affix ends up on the numerical noun (we have the sequence *Num-K*), and the complement of the numerical noun is in genitive (*N-GEN*). This corresponds to the nominal pattern.

   Alternatively, the whole complement of K moves, and we get the phrasal case pattern (14c). Here, the numeral is uninflected for case (*Num* is directly followed by N), and the case marker follows linearly the counted noun. The sequence we have on the surface is such that after the counted noun in the genitive, an additional case marker should appear. In the surface strings (8), however, we never see phrasal case marking superimposed on the genitive marking. The reason for that has already been explored in section 6: ellipsis. The inner GEN is elided because its features are contained inside any oblique K that follows it.

   To sum up, the analysis says that the phrasal case pattern should be analyzed as an instance of phrasal affixation, where a case affix attaches to the whole phrase containing the numeral and the noun. In addition, the sequence of two case affixes is simplified by eliding the genitive (because its features can be recovered from the other oblique affix).

   This proposal singles out the agreeing pattern against the nominal and the phrasal-case patterns by proposing two distinct base-generated structures. The agreeing pattern corresponds to (12a) with two case markers in the base structure (in addition to GEN); the nominal pattern and the phrasal-case pattern only have one case marker (in addition to GEN), see (14a). This split seems to be empirically justified by the fact, seen in (8), that 'hundred' in the instrumental has both the nominal pattern and the phrasal-case pattern (corresponding to the base structure in 12b), but lacks the agreeing pattern (corresponding to 12a).

   Let me finaly turn to the question of how ellipsis operates in these examples; recall from (7) that ellipsis cannot eliminate the counted noun

in the phrasal case pattern (7b); this can now be understood as a consequence of the fact that the elided material (boldfaced in (15a)) does not form a constituent. That is because the numeral is included in the smallest constituent containing the counted noun and the case marker, so ellipsis cannot apply to the relevant string.

   The demonstrative (if present) has its own case marker internally in its projection, and it would be added on top of the whole structure (15a). Therefore, it would not interact with the ellipsis (which it doesn't, as we saw in (7)).

(15)   a. [*Num* [*N-GEN*]] [K ~~[*Num* [*N-GEN*]~~]]
         b. [[Num [K ~~Num~~]] **[N-GEN [K ~~N-GEN~~]]**   ]

The structure assumed for the greeing pattern is in (15b). Both the genitive and the numeral each have their own case marker. Each of them moves across it, and then they are assembled. In this case, the counted noun and the case marker form a constituent (in bold), which can be elided.

   What remains is to provide an account for the genitive-as-oblique pattern, a task which I turn to now. The analysis I provide below is rather tentative. In simple terms, it builds on the idea that the numeral has a defective paradigm: it only has the nominative, accusative and genitive – but it does not have the locative, dative or instrumental case (see Bošković 2001 for a discussion of Serbian examples of this sort). In the theory outlined here, this means that the numeral has restricted movement options – it can only move as high as the feature C, but no higher. However, in order to express the instrumental meaning, the remaining features must be somehow expressed. The idea I am going to encode formally below is that they are spelled out on the complement noun.

9 Case Decomposition and Pied-Piping

As the first step of the (tentative) explanation for the genitive-as-oblique pattern, let me first specify in more detail the derivation of the nominal pattern. What I do first is to enrich the simplified representation in (14b) by the proposal in (10), namely that case decomposes into a number of features. When we do that, replacing the single K for a sequence of

projections, we get the base structure (16a), and, after moving the numerical noun as in (14b), the derived structure (16b). The problem with (16b) is that the features of the instrumental A—F do not for a constituent (there is the genitive noun inside the constituent as well). Hence, if we adopt the idea of phrasal spell-out, these features cannot be pronounced by a single morpheme in (16b). A solution is provided by the proposal – put forth in Cinque's (2005) theory of ordering in the noun phrase – that genitive complements always move high up to the left, as shown in (16c). After this movement, the case features A—F do form a constituent (ignoring traces), and they can be spelled out by the instrumental marker as indicated below the line (16c). Finally, a step of remnant movement takes the numerical noun back to the left of the noun (again as in Cinque's 2005 proposal). The only non-standard aspect of the proposal is that I treat the numeral as a noun.

(16)   a.   [F [E … [A [*Num* [*N-gen*]]]]]
       b.   *Num* [F [E … [A [~~*Num*~~ [*N-gen*]]]]]
       c.   *N-gen* [*Num* [F [E … [A [~~*Num*~~ [~~*N-gen*~~]]]]]]
                          Num ------ *ins* ------
       d. *Num-ins N-gen*

Let me now combine this more detailed and accurate derivation with the idea that some numerical nouns are defective, and only move as high as C in the structure, but never as high as F.

   The derivation is shown in (17). We first merge features A—C and build the genitive case (17a), which is as high as the numerical noun may move (17b). What we do then is add the rest of the case features that make up the instrumental. If the numeral were not defective (like 'hundred'), we could move it to the left of the feature F as we did in (16b); but remember that 'five' and its kin are defective and may not move higher than C. Therefore, we skip moving the numeral, and move the counted noun in a way similar to (16c), thereby creating a constituent containing the features A—C, which can now be spelled out as a genitive following the numeral. The numerical noun now undergoes remnant movement as in (16d), and ends up to the left of the counted noun. However, unlike in the nominal pattern (16d), the numeral is marked genitive. The final observation is that the features D—F now follow the counted noun, which itself has the features A—C. The final point of the

analysis is that the features of the counted noun (A—C) and the features that actually belong to the whole phrase (D—F) now form a complete set of features (A—F) that correspond to the instrumental (see the boldfaced part). Since spell out is post-syntactic, it does not care about the derivational origin of the features, and spells them out as one marker. This way, the genitive-as-oblique pattern emerges.

(17) a.   [C [B [A [*Num* [*N-gen*]…
      b.   *Num* [C [B [A [~~*Num*~~ [*N-gen*]…
      c.   [F [E [D [*Num* [C [B [A [~~*Num*~~ [*N-gen*]…
      d.   *N-gen* [F [E [D [*Num* [C [B [A [~~*Num*~~ [~~*N-gen*~~]…
                                            Num   --- *gen* ---
      e. *Num-gen* **N-gen [F [E [D** ~~[*Num* [C [B [A [*Num* [N-gen*~~]…

The way the features belonging to two different nouns merge in a single marker is reminiscent of clitic-cluster effects. For instance, in Pazar Laz (Blix 2012), the following pattern is found. If the subject or object is 3rd plural, the suffix *–an* appears on the verb. Further, if the subject or object is 3rd singular, -*s* appears, and if the subject or object is plural (but not 3rd person) the suffix *–t* appears. In this system, when a 3rd singular subject and a second plural object appear as arguments of the verb, we would expect the combination of suffixes *–s-t*, where *–s* marks a third singular argument (the subject), and *–t* marks a plural argument (the object). But what in fact happens is that we get a single marker: the 3rd plural *–an*. The features of subject (3rd) and object (Pl) merge together. Similarly in (17e), features with different sources (some belong to the counted noun, and some belong to the whole numerical phrase) merge in a single morpheme.

10 Conclusions

The current paper makes two contributions. The first one is to enrich our typology of numerical constructions. Specifically, what has been called the homogeneous pattern divides in Czech into three distinct sub-patterns: the agreeing pattern, the phrasal-case pattern, and the genitive-as-oblique pattern. The second goal was to explain the syntax of these patterns using a particular analysis of the homogeneous pattern as an instance of case attraction. Assuming this analysis, the existence of the

various sub-patterns was explained as arising from varying two parameters: what the host of case marking is, and how movement proceeds.

References

Babby, Leonard H. 1985. Noun Phrase Internal Agreement in Russian. Russian Linguistics 9, 1-15.

Baerman, Matthew, Brown, Dunstan and Corbett, Greville G. 2005. *The Syntax-Morphology Interface. A Study of Syncretism*. Cambridge: Cambridge University Press.

Baker, Mark C. 2008. *A Theory of Agreement and Concord*. Cambridge: Cambridge University Press.

Bittner, Maria and Hale, Ken. 1996. The Structural Determination of Case and Agreement. *Linguistic Inquiry* 27, 1-68.

Blix, Hagen. 2012. *Kartvelian and the pieces of inflection.* Ms., Leipzig.

Bobaljik, Jonathan D. 2008. Where's Phi? Agreement as a post-syntactic operation. In *Phi-Theory: Phi features across interfaces and modules*, ed. David Adger, Susanna Béjar and D. Harbour, 295-328, Oxford: Oxford University Press.

Caha, Pavel. 2009. *The Nanosyntax of Case*. PhD thesis, Tromsø.

Caha, Pavel. 2010. The Parameters of Case Marking and Spell Out Driven Movement. *Linguistic Variation Yearbook* 10: 32-77.

Caha, Pavel. 2013. Explaining the Structure of Case Paradigms by the Mechanisms of Nanosyntax. *Natural Language and Linguistic Theory* 31: 1015-1066.

Caha, Pavel. 2015. Czech numerals and No Bundling. In *Beyond the Functional Sequence*, ed. Ur Shlonsky, 173-196. Oxford: Oxford University Press.

Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 1-52. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 2005. Explaining Greenberg's Universal 20 and Its Exceptions. *Linguistic Inquiry* 36: 315-332.

Corbett, Greville. 1995. Slavonic's Closest Approach to Suffixaufnahme: The Possessive Adjective. In *Double Case*, ed, Frans Plank, 265-282. New York, Oxford: Oxford University Press.

Franks, Steven. 1994. Parametric properties of numeral phrases in Slavic. *Natural Language and Linguistic Theory* 31: 1015-1066.

Franks, Steven. 1995. Parameters of Slavic Morphosyntax. Oxford: Oxford University Press.

Halle, Morris. 1994. The Russian Declension: An Illustration of the Theory of Distributed Morphology. In *Perspectives in Phonology*, J. Cole & C. Kisseberth, eds., 29-60. Stanford: CSLI Publications.

Harbert, Wayne. 1983. On the nature of the matching parameter. *The Linguistic Review* 2: 237-284.

Ionin, Tania and Ora Matushansky. 2006. The Composition of Complex Cardinals. *Journal of Semantics* 23: 315–360.

Johnston, Jason. 1996. *Systematic Homonymy and the Structure of Morphological Categories*. PhD thesis, University of Sydney.

Kayne, Richard. 1994. *The Antisymmetry of Syntax.* Cambridge, MA: MIT Press.

Leu, Thomas. 2008. *The Internal Syntax of Determiners*. PhD thesis, NYU.

McCreight, Katherine and Chvany, Catherine V. 1991. Geometric representation of paradigms in a modular theory of grammar. In *Paradigms: The Economy of Inflection*, ed. Frans Plank, 91- 112. Berlin, New York: Mouton de Gruyter.

Neidle, Carol. 1988. *The Role of Case in Russian Syntax.* Dordrecht: Kluwer: Dordrecht.

Norris, Mark. 2014. *A Theory of Nominal Concord*. PhD thesis, UCSC.

Plank, Frans. 1995. Re-introducing Suffixaufnahme. In *Double Case*, ed, Frans Plank, 1-99. New York, Oxford: Oxford University Press.

Saab, Andrés and Anikó Lipták. 2015. Movement and deletion after syntax. *Studia Linguistica* (early view).

Starke, Michal. 2009. Nanosyntax: A Short Primer to a New Approach to Language. In *Nordlyd*, ed. Svenoius, Peter et al., 1-9. Tromsø: Universitet i Tromsø.

Timberlake, Alan. 2004. *A Reference Grammar of Russian.* Cambridge: Cambridge University Press.

pcaha@mail.muni.cz