

Czech numerals in a phrasal spell out model

The goal of the paper is to provide a Nanosyntactic analysis of a peculiar case marking pattern attested with Czech/Slavic higher numerals. Nanosyntax (Starke 2009, Caha 2009) is a framework whose central idea is that the mental representations attributed to speakers (i.e., syntactic structures) are very fine grained, with each feature harboured by a separate terminal node. Morphemes which correspond to more than one feature are treated as portmanteau markers that spell out a non-trivial syntactic constituent (a phrase):

(1) $[_{AP} A B] \Leftrightarrow$ morpheme

There are two consequences of this idea that – as I argue -- are useful in our attempts to understand the morphology and syntax of numerical phrases in Slavic (see previous work by Babby, Franks, Leko, Matushansky, Pesetsky, Rutkowski (in alphabetical order), among many others).

A. Case attraction

The most peculiar behaviour of Czech numerals (which holds across Slavic more generally) concerns the case marking of the counted noun. When the numerical phrase is either the subject or the object, the counted noun is marked genitive (see 2a). When the numeral is dative or instrumental, the genitive case disappears, and it is replaced by the appropriate case, see (2b,c).

(2a) 5-nom/acc cars-gen

(2b) 5-dat cars-dat

(2c) 5-ins cars-ins

The shared intuition has always been to understand the difference in terms of the structural/inherent distinction, giving rise to many sophisticated technical proposals. In a broad perspective, the structural/inherent distinction has been (and still is) prototypically understood as a derivational distinction in some sense; in earlier times, inherent case was assigned before structural cases at D-structure. In Pesetsky's modern (2013) analysis, timing is also crucial although it goes the other way round (obliques are assigned late and override the earlier genitive).

Breaking with the tradition, I develop a new representational perspective. I use a view on the structural/inherent distinction that originates the work by Bayer et. al. (2001), and has been elaborated on in Nanosyntax, specifically in Caha's (2009) proposal, where inherent cases are structurally more complex than structural cases, see (3):

3 $[ins E [dat D [gen C [acc B [nom A xNP]]]]]$

According to this proposal, cases correspond to collections of features (nom=A, acc=A,B, gen=A,B,C, ...) and the features are added on top of the (extended) projection of the noun as syntactic heads. As a consequence, the traditional *derivational* distinction between structural and inherent case is understood as a *representational* difference: inherent case has more structure than structural case.

My proposal for numerals takes the structure in (3) as given and independently motivated by syncretism and other effects discussed in Caha's work. The way I use it to explain the behavior in numerical patterns falls in line with approaches such as Pesetsky (2013) or McCreight's earlier (1988) work, who rely on the concept of multiple case marking. Specifically, I propose that the counted noun ALWAYS bears two cases, namely the genitive assigned by the numeral and in addition an agreement like case appropriate for the syntactic environment. So underlyingly, the data in 2 look as shown in the column SYNTAX.

	SYNTAX		PRONUNCIATION
(4a)	5-nom/acc $[[cars-gen] nom/acc]$	---	cars-gen+nom/acc
(4b)	5-dat $[[cars-gen] dat]$	---	cars- gen +dat
(4c)	5-ins $[[cars-gen] ins]$	---	cars- gen +ins

The idea is that a regular ellipsis process applies to such structures, eliminating one of the cases. The case which the ellipsis deletes is the one which is recoverable on the basis of an antecedent (the

other case). In (4a), nom/acc are recoverable on the basis of gen, because gen contains them (see (3)). In (4b,c), gen must be eliminated, because dat/ins contain it per (3). If correct, the peculiar case marking pattern warrants nothing special to be said. It is reduced to the interaction of two independent operations. The first operation is multiple case marking, also known as suffixaufnahme (in work by Plank 1995) or case stacking (in work by Richards 2013). The second concept is ellipsis under recoverability. Such a reduction of the peculiar case pattern is a welcome property of the 'representational' approach to the structural/oblique distinction.

The question that remains is how to understand the fact that in Czech/Slavic, only numerals (and some quantifiers) enter the peculiar suffixaufnahme structures (with two cases). The answer I suggest draws on the traditional insight that the numerals stand somewhere in between lexical nouns (assign genitive, decline like nouns), and functional categories in the NP (enter into concord). I spell out the idea in the next section.

B. Syntactic categories

In most conceptions of grammar, lexical items belong to exactly one category: A or B. In effect, this means that a lexical item like the Czech *deset* 'ten' or *sto* 'hundred' is either a numeral (for Czech the common stand) or a noun (see Matushansky and Ionin 2006), but it can never be both at the same time. However, the consequence of phrasal spell out is that this widely shared assumption is too simplistic. The morpheme which spells out the tree in (1) contains both the categories A and B at the same time, and to the extent that B has an effect on the outer environment of AP, the morpheme will exhibit a 'mixed' categorial behavior.

The phrasal spell out model can thus be used to model the mixed behavior of higher numerals in Czech and elsewhere. In particular, the fact that they assign genitive is something they share with nouns. At the same time, they differ from nouns in oblique contexts, and this difference can be shown to follow from the proposal that numerals spell out not only the lexical N head, but also a functional Num head in addition. Such a lexical entry leads to a particular structure, where the genitive dependent of the embedded noun is forced to move out of the constituent spelled out by the numeral, and acquires concord as a result of its special landing site.

The second fact that the proposal captures is that some numerals are ambiguous between a numeral use and a noun use, a fact similar to Kayne's (2006) English examples *hundred mistakes* and *hundreds of mistakes*. In the nominal use, the numerals combine with diminutive morphology, plural morphology, and always require the genitive on the counted noun, etc. I show how the ambiguity can be easily accounted for relying on the so called Superset Principle (Starke 2009). The principle says that the morpheme in (1) can spell out either the whole phrase, or a sub-phrase. The nominal use of the numerals then corresponds to a predicted scenario, where the lexical entry in (1) is used in order to insert the "numeral" solely under the syntactic N node.

Most importantly, the proposal forces any complements of the noun to be located outside of NumP. This is because otherwise Num and N could not spell out together. (i.e., Spell out targets constituents, the complement of N in a low position would be inside the constituent where Num and N are, and block their spell out.) There are reasons to believe that moving or being generated high in the functional structure of the NP correlates with agreement, hence the presence of the second case on the genitive.

CONCLUSIONS

In sum, the point is to show that phrasal spell out is a tool that provides new and interesting analytical options in the domain where syntax and morphology interact in intriguing ways. It allows for a very fine grained case decomposition (as in 3), which neatly captures case competition in numerical phrases. Secondly, the numerals themselves receive a treatment that both explains their affinity with nouns (they spell out N), and explains the differences (they spell out Num in addition).