

# Czech numerals and No Bundling

Pavel Caha  
CASTL, Tromsø

January 25, 2013

## Abstract

One of the hypotheses investigated in Cartography (Cinque and Rizzi 2010) and Nanosyntax (Starke 2009) is that each feature corresponds to a separate node in syntax (No Bundling). This paper confronts the hypothesis with some apparently problematic data in the domain of Czech numerals. I argue that No Bundling requires that numerals are nouns, and that consequently, numerical phrases are bi-nominal structures. I support this conclusion by drawing parallels between the Czech numerical constructions and uncontroversial bi-nominal structures.

## 1 Introduction

This paper looks at Czech higher numerals (five to ten, hundred and thousand).<sup>1</sup> An interesting property they share is the following. When the whole phrase including the numeral and the counted noun is in the subject or object function, an ‘unexpected’ genitive case shows up on the counted noun. I give an example in (1-a); (1-b) shows that it is impossible for the counted object to appear in the nominative or accusative form.

- (1) a. pět chlap-ů  
five.NOM/ACC guys-GEN.PL  
‘five guys’
- b. \*pět chlap-i/chlap-y  
five.NOM/ACC guys.NOM.PL/guys.ACC.PL  
‘five guys’

However, in the oblique cases (i.e., LOC, DAT, INS), the counted noun in numerical phrases is not marked genitive, see (2-a), but it bears the relevant oblique case instead (2-b). (This holds for the numerals ‘five’ to ‘ten.’ ‘Hundred and ‘thousand’ allow for both of the patterns in (2). I come back to this later.)

- (2) a. \*Dal to pět-i chlap-ů  
he gave it five-DAT guys-GEN  
‘He gave it to five guys.’
- b. Dal to pět-i chlap-ům  
he gave it five-DAT guys-DAT  
‘He gave it to five guys.’

This is in sharp contrast with other nouns, including ‘group’ nouns (denoting various collections of individuals). The subject form of such complex phrases is shown in (2-a), and it has an ‘expected’ genitive case on the complement of the noun ‘group.’ The genitive is preserved in oblique cases, as (3-b) demonstrates. It is impossible to use the numerical pattern where the dependent noun is marked by the relevant oblique case, see (3-c).

- (3) a. skupin-a chlap-ů  
group-NOM guys-GEN.PL  
‘a group of guys’
- b. Dal to skupin-ě chlap-ů  
he gave it group-DAT guys-GEN
- c. \*Dal to skupin-ě chlap-ům  
he gave it group-DAT guys-DAT  
‘He gave it to a group of guys.’  
(both (b) and (c))

These facts taken together suggest that it is impossible to analyze numerals as ordinary nouns, because they do not assign GEN in (2) (but see Ionin and Matushansky 2006). Similarly, they cannot be analyzed as ordinary adjectives (because they do assign genitive in (1), and Czech adjectives never do). Consequently (and correctly, I believe), numerals are treated as objects of a special type in a number of works. Most prominently, numerals are considered to be a category *sui generis*, generated either as the head of a dedicated functional projection, or as a phrase in its Spec (or both) (Babby

---

<sup>1</sup>Other higher numerals are composite. For these, I basically assume the account given in Ionin and Matushansky (2006).

1987, Franks 1994, Pereltsveig 2007, Rutkowski 2006, Brattico 2011, Danon 2012). In what follows, I will refer to this as the standard view.

This common analysis reflects the state of the art in theoretical research into categorization of expressions. Specifically, it is assumed that an item cannot simultaneously belong to two categories. Applied to our specific case, belonging to the special class of numerals is incompatible with belonging at the same time to the class of nouns. I will refer to this idea as the ‘single category’ view.

Such a stand plausibly reflects the idea that lexical items may only occur as terminals of the syntactic tree. With such a view in mind, the only way to express the fact that an element differs in some grammatical aspect from a run-of-the-mill noun is to say that it is not a noun, but the exponent of an altogether distinct grammatical category (Num).

At the same time, there has been a constant tension between the ‘single category’ view and the observation that many items stand somewhere in between prototypical lexical and prototypical functional categories. The Czech numerals I focus on here, for instance, require (in some environments) a case on the counted noun that is typical for nominal dependents, and keep (to a certain extent) nominal declension (the dative *-i* in (2-b) is homophonous with a DAT marker that appears with a class of nouns, and it is distinct from any adjectival agreement marker).

Such items have been sometimes called ‘semi-lexical’ categories, and their investigation has received some focus in the literature. Van Riemsdijk (1998), for example, proposes that all functional categories share a categorial feature matrix with the lexical head, making them functional and lexical (nominal, in our case) at the same time. In a different line of research inspired by Kayne’s work (see, e.g., Kayne 2005), Zweig (2006) proposes that some numerals – even if not nouns themselves – modify a silent noun NUMBER. These approaches share the conviction that insertion targets only terminals, and try to come up with solutions that would make justice to the existence of ‘semi-lexical’ items.

However, the ‘items as terminals’ view has alternatives. In a theory like Nanosyntax (Starke 2009, Caha 2009), lexical items may spell out a whole set of projections, provided these projections form a constituent. If that is so, numerals (and semi-lexical categories in general) may be understood as items whose lexical specification overlaps with nouns, but includes additional functional projections. Specifically, numerals may be conceived of as a special noun that spells out a non-trivial syntactic phrase: a NumP. This is depicted in (4).<sup>2</sup>

$$(4) \quad /numeral/ \Leftrightarrow [ \text{Num} [ \text{N} ] ] \Leftrightarrow \text{Value}$$

To be more explicit: (4) is an entry for a numeral, pronounced /numeral/, which spells out a phrasal node composed of nominal bottom (something like Zweig’s silent NUMBER) and a projection where precise numerical value is encoded in the functional sequence (Num). The exact value is not relevant for syntax (seven and eight have the same syntax in Czech), and I will be assuming that the differences in precise numerical value are specified as that part of the lexical entry which is accessed by the conceptual system (represented by ‘Value’).

Possessing the tool of phrasal spell-out thus enriches our expressive power: we are able to directly encode the fact that numerals are noun-like in some aspects, but that they are special in other respects. In this paper, I set out to explore this view, formally expressed in (4).

## 2 No Bundling

Let me begin by invoking some theoretical considerations that support (4). The main point is this: if we take numerals to be altogether unrelated to nouns (not include an N in their specification), we create a problem for the so called ‘No Bundling’ hypothesis. The ‘No Bundling’ hypothesis says that each morphosyntactic feature is a head in the tree (see Kayne 2005, Starke 2009, Cinque and Rizzi 2010 for suggestions along these lines).

To see why the standard view on numerals is incompatible with ‘No Bundling,’ consider what structure we arrive at assuming No Bundling and the standard theory at the same time. The starting point is that sometimes, there are two distinct cases in numerical constructions: one on the numeral, and a distinct case on the counted noun, see (1-a). As a consequence of No Bundling, we need one case projection above the numeral (it has an ambiguous NOM/ACC form), and another one above the noun (it bears GEN). Combining this with the proposal that Numerals sit in the Num head (or its Spec) leads to the structure given in (5).

$$(5) \quad [ \text{K} [ \text{Num} [ \text{K} [ \text{NP} ] ] ] ]$$

---

<sup>2</sup>Other alternative solutions have been proposed. Corbett (1978) suggests that the numerals have a hybrid category he notates as ADJ/N. Similarly, Giusti and Leko (2005) claim that such higher numerals are sometimes merged as nouns (when they assign genitive), and sometimes as adjectives (when they don’t).

But (5) is incompatible with what we know independently about the nature of the functional sequence. Specifically, the functional sequence is an irreflexive ordering of elements, such that A may never dominate A. Hence, the traditional view on numerals is incompatible with No Bundling. As a consequence, most approaches to numerals in Slavic give up the No Bundling hypothesis, and consider case a feature of the noun, which is never granted the capacity to project. When the feature is hidden from the main projecting line, no issue arises for the irreflexive nature of the functional sequence.

On the other hand, the proposal in (4) is compatible with ‘No Bundling.’ Specifically, if numerals are in fact a special class of phrasal nouns, then there are actually two nouns in the structure, and each of the case projections in (5) is associated to its own nominal element. I show this in (6); the subscripts highlight the fact that we are dealing with two extended functional sequences (each subscript for the whole extended sequence).<sup>3</sup>

(6) [ K<sub>2</sub> [ Numerical Noun<sub>2</sub> [ K<sub>1</sub> [ Counted Noun<sub>1</sub> ] ] ]

Consequently, the current proposal is easily compatible with No Bundling. To the extent that the hypothesis is right, (4) has a point in its favor.

The ‘phrasal numeral’ hypothesis also holds the promise of explaining the peculiar case distribution in the numerical phrases. In order to show that, the next sections establish some general background concerning the syntax of bi-nominal expressions. In these sections, I abstract completely from the Czech specific situation, and look at various phenomena found in bi-nominal constructions in the languages of the world. The idea behind this move is to gain a general understanding of the syntax of bi-nominal expressions without making case specific stipulations about the Czech numerical construction. In the following section, I argue that once the general mechanics of bi-nominal syntax are combined with the ‘phrasal numeral’ hypothesis, the combination is going to yield the empirical facts as an automatic consequence.

### 3 Case attraction: the data

As highlighted, the main empirical reason for analyzing Czech (and more broadly Slavic) higher numerals as something else than nouns is the fact that they contrast with nouns in oblique contexts, recall (2) and (3). As Rutkowski (2006) summarizes the argument, the “claim that numerals are nouns makes it impossible to distinguish between the two case patterns.” In this section, I want to turn the argument on its head, and claim that the pattern in fact provides evidence that numerals are nouns. In order to see that, consider a phenomenon known as ‘case attraction,’ which is illustrated in (7). The data are taken from Classical Armenian, and they show that Classical Armenian has two ways of marking possessors: the possessor is either marked genitive (7-a), or it is ‘attracted’ (7-b). Attraction consists in replacing the genitive by the case borne by the head.

(7)	Classical Armenian (Plank 1995)	b.	i	knoj-ê	t’agawor-ê-n
	a.		i	knoj-ê	t’agawor-i-n
				by wife-ABL.SG	king-ABL.SG-DEF
				by wife-ABL.SG	king-GEN.SG-DEF
					‘by the wife of the king’

The important point is that the distribution of case markers in attraction structures (7-b) is indistinguishable from the pattern we have seen in (2-b) (the oblique case replaces the expected genitive). And because (7-b) is uncontroversially a bi-nominal structure, it is not possible to use an identical distribution of case markers in (2-b) to back up the conclusion that items which exhibit the pattern are not nouns.

In fact, we reach a conclusion that is quite opposite to Rutkowski’s: since there must be a way to generate the case pattern in (7-b) with *knoj-* ‘wife’ a noun, there must also be a way to generate the apparently peculiar pattern (2-b) in the same way (i.e., with ‘five’ a type of a noun). And since that is so, it is superfluous to generate exactly the same pattern in a different way (Occam’s Razor).

The important parallel between Czech numerical phrases and case attraction is strengthened by the observation that the Classical Armenian attraction pattern (7-b) is unattested in NOM and ACC (Plank 1995, p.43); in these cases, genitive marking is the only option:<sup>4</sup>

<sup>3</sup>I will further adopt the hypothesis that the counted noun is in fact generated in a Spec of the numerical noun, but that is orthogonal to the main point.

<sup>4</sup>As Classical Armenian is a dead language, the star in front of the example means that such examples are simply unattested. Their ungrammaticality can only be extrapolated on the basis of the known facts.

- (8) No attraction in ACC (or NOM)
- |    |                 |                 |    |                 |                           |
|----|-----------------|-----------------|----|-----------------|---------------------------|
| a. | nšanagir-s      | alp'abet-ac'    | b. | *nšanagir-s     | alp'abet-s                |
|    | alphabet-ACC.PL | alphabet-GEN.PL |    | alphabet-ACC.PL | alphabet-ACC.PL           |
|    |                 |                 |    |                 | 'letters of the alphabet' |

This should be considered in parallel to the fact that NOM/ACC environments lead to the emergence of the genitive case in the relevant type of Czech numerical phrases, recall (1). In other words: if the Czech numerical pattern is an instance of case attraction (a bi-nominal construction), then the distinction between structural and oblique cases is just an expected consequence of this classification.

Let me also make explicit a distinction between Armenian and Czech which has appeared in the data. In particular, on the basis of the example in (7), attraction seems 'optional' (at the first glance). However, the Czech numeral 'five' enters only the attraction pattern, see (2). And conversely, ordinary nouns cannot exhibit case attraction in Czech. In sum, while the two Armenian bi-nominal patterns each find an analogue in Czech, each type appears to be dedicated to a particular class of items.

However, there are at least two items in Czech that appear in both patterns: 'hundred' and 'thousand'. These allow for their complement to appear either in the genitive, or in the relevant oblique case, see (9).<sup>5</sup>

- (9) a. Dal to st-u chlap-ũ  
 he gave it hundred-DAT guys-GEN  
 b. Dal to st-u chlap-ũm  
 he gave it hundred-DAT guys-DAT  
 'He gave it to hundred guys.'

In order to understand these issues, we need to have an analysis of the mechanics underlying case attraction. In the next section, I explore case attraction and related phenomena in order to gain insight into the syntax of the Czech numerical phrases.

## 4 Case attraction: an analysis

The analysis of case attraction has two logically independent parts. One is to understand the process that targets specifically the dependent noun (prototypically the possessor), and replaces its genitive marking by an agreement-like marker. The other part concerns the external conditions which control whether a given dependent noun is going to be subject to the process (attraction sometimes must and sometimes mustn't occur). I take these up in turn.

### 4.1 The internal mechanics of attraction

Caha (in press) argues that attraction should be analyzed as a combination of two processes: (i) possessor agreement (ii) followed by an ellipsis of the regular genitive marking. The ellipsis is licensed by the agreement marker, which remains present in the string. Let me now highlight the two steps in turn.

According to the first proposal, case attraction is seen as related to a construction that is sometimes referred to as Suffixaufnahme (Plank 1995). An example is in (10). What is relevant is that the possessor carries two case markers: the possessive genitive (*-ndamunu*) and an additional agreement marker (*-du*).

- (10) Dicki-ndamun-du kaya-ngka  
 Dick-GEN-ERG dog-ERG  
 'Dick's dog' (Guugu Jalaŋji, Plank 1995)

Such Suffixaufnahme structures are the input to step (ii), where the genitive is obligatorily deleted:

- (11) Case attraction as agreement plus ellipsis  
 head noun-case<sub>i</sub> [[ dependent noun-~~GEN~~] AGR=case<sub>i</sub> ]

Crucially, the ellipsis of the genitive case must be recoverable. This allows us to explain the fact that case attraction applies in oblique cases, but it does not happen in NOM and ACC. In particular,

---

<sup>5</sup>They still permit only GEN in NOM/ACC environments.

- (i) st-o { chlap-ũ / \*chlap-i }  
 hundred-NOM.SG guys-GEN.PL guys-NOM.PL  
 'hundred guys'

the idea is that in terms of feature content, all oblique cases contain the genitive case at their core (licensing the ellipsis), while NOM and ACC do not (and fail to license it). The proposal concerning containment relations among various cases, and the specific place of the genitive among them, is independently motivated in detail in Caha (2009), with Czech specific section in ch.8. In the interest of space, I do not reconstruct the discussion at this place.

Assuming an account along these lines, a question to be tackled is what happens in structural cases. Here, the genitive marking of the counted noun cannot be elided, because NOM/ACC have fewer features than GEN. As a consequence, there is no antecedent for the deletion, and the genitive has to surface. However, unlike in suffixaufnahme structures, see (10), the genitive is never followed by an overt agreement marker. The question arises how to analyze its apparent absence, and the following options come to mind. (i) The structure of the structural cases is different: the genitive marking is a reflex of an ordinary bi-nominal structure, and no additional agreement is ever present on the genitive; (ii) the syntax of the structural cases is the same as in the oblique cases: there is an agreement marker in the syntax, but it is either phonologically null, or elided. I suggest here that one of the options given in (ii) is correct.

The first reason to explore such a path is the uniformity of the analysis: we know that in oblique cases, numerals like ‘five’ undergo attraction (and hence agreement) obligatorily. It seems theoretically complicated to make sure that agreement does not apply in structural cases, while it has to apply in oblique cases.

The second reason is empirical, and it has to do with demonstrative agreement. The following observation is relevant: in Czech numerical phrases without attraction, demonstratives have to agree with the numeral, and not the counted noun. I illustrate this below on the numeral ‘hundred.’ Recall from (9) that with this particular numeral, the counted noun may be marked GEN also in oblique cases. When that happens, the demonstrative has to be SG, just like the numeral (and not PL like the counted noun).

- (12) a. Dal to **t-omu** **st-u** chlap-ů  
           he gave it that-DAT.SG hundred-DAT.SG guys-GEN.PL  
       b. \*Dal to **t-ěch** st-u **chlap-ů**  
           he gave it that-GEN.PL hundred-DAT.SG guys-GEN.PL  
           ‘He gave it to those hundred guys.’

Now in structural cases, the pattern is the following. For numerals like ‘hundred,’ which allow both attraction and non-attraction (see (9)), the demonstrative may agree either with the numeral, see (13-a), or the counted noun, see (13-b):

- (13) a. **t-o** **st-o** chlap-ů  
           that-NOM.SG hundred-NOM.SG guys-GEN.PL  
       b. **t-ěch** st-o **chlap-ů**  
           that-GEN.PL hundred-NOM.SG guys-GEN.PL  
           ‘those hundred guys’

Starting from (13-a), it seems reasonable to conclude that it represents the counterpart to (12-a): it is an ordinary bi-nominal structure, and the demonstrative agrees with the numeral. However, the fact that (13-b) is possible (in fact, preferred) may come as a surprise: recall that (12-b) was out. The only way out of the puzzle seems to be admitting that (13-b) is not an ordinary bi-nominal structure: it is a ‘case-attraction’ structure.<sup>6</sup>

In sum, the claim of this sub-section is that case attraction is a process very similar to possessor agreement, and differs from it only in that it adds an additional process of genitive ellipsis (under recoverability). As a consequence, the difference between structures with attraction and simple bi-nominal structures (with a regular genitive) is a difference between an ‘agreeing possessor’ construction (attraction) and a ‘non-agreeing possessor’ construction (plain genitive).

In the next section, I set out to explore the external conditions which allow/ban attraction structures to be generated. In doing so, I take the parallel between attraction and possessor agreement seriously. Specifically, I look at the conditions governing possessor agreement in Old Georgian, with the intention to extrapolate the results for attraction structures.<sup>7</sup>

## 4.2 External conditions on attraction

There are languages where agreeing and non-agreeing genitives differ in their syntactic position. Before I get into the fine-grained details, let me say that I am going to assume a view according

<sup>6</sup>I put case attraction in scare quotes to signal the fact that no attraction (genitive ellipsis) actually takes place.

<sup>7</sup>I do this because I am not aware of any study that looks at the difference between attraction/non-attraction in Armenian, or other languages with case attraction.

to which dependent nouns with various interpretations sit in a specifier of the head noun; or, more precisely, in the Spec of one of its functional projections. It may be that they are base-generated there (see Adger 2012) or move there (Kayne 2004, Cinque 2005), a question that is not crucial for my current concerns. What is crucial, on the other hand, is that the particular position of the dependent noun seems to vary depending on whether it agrees with the head or not. A language where this may be well observed is Old Georgian. In this language, as the evidence suggests, non-agreeing possessors have to be low, while agreeing possessors are located high.

In order to understand the relevant details of the language, let me start by giving relevant background. In Old Georgian, adjectives, numerals, demonstratives and articles agree with the head noun, see (14) for an example. For these items, agreement is obligatory.

- (14) or-**n-i**          brma-**n-i**  
two-PL-NOM men-PL-NOM  
'two men' (Boeder 1995, 155)

In bi-nominal constructions, agreement may target also genitive constituents, see (15-a). As (15-b) shows, such agreement marking is present only once in the whole genitive phrase: the adjective 'holy,' modifying the head of the genitive phrase, does not show double case marking.

- (15) Old Georgian Boeder 1995                      b. šecev**n-ita** [cmid-isa sameb-isa]-**jta**  
a. šroša-**n-i**    vel-isa-**n-i**                      help-INS    holy-GEN trinity-GEN-INS  
lily-PL-NOM field-GEN-PL-NOM                with the help of the holy trinity (p. 159)  
'the lilies of the field' (p.159)

Under certain conditions, agreement marking of the genitives is absent. The specific conditions are what interests me here. A basic contrast is that while post-nominal possessors always have to agree, pre-nominal ones do not have to do so. This is shown in the following pair of examples (16). In both of them, the head noun 'womb' receives the instrumental case from the postposition *gan* 'from'. In the first example, see (16-a), the genitive phrase 'his mother' follows the head, and receives an additional instrumental marking, occurring at the right edge of the whole genitive constituent. In (16-b), the genitive phrase precedes the head, and receives no agreement.

- (16) Old Georgian Boeder 1995  
a. mucl-**it-gan**    [ded-isa    tws    -isa]-**jt**  
womb-INS-from mother-GEN her own -GEN-INS  
'from the womb of his [sic] mother' (p.179)  
b. ded-is          mucl-it-gan  
mother-GEN womb-INS-from  
'from the mother's womb' (p.157)

Interestingly, the precise conditions that determine whether a particular genitive phrase is or is not marked by agreement are more intricate than the simple pre- vs. post-nominal distinction. That is shown by the following example, where a pre-nominal genitive agrees with the head:

- (17) Iesu-is-**i**          xilva-**j**  
jesus-GEN-NOM seeing-NOM  
[we want to] see jesus (lit. [we want] seeing of Jesus) (Boeder 1995, 163)

Focussing now on the class of pre-nominal genitives, an interesting contrast emerges between agreeing and non-agreeing items. Specifically, as Boeder (1995) argues, "genitives without Suffixaufnahme appear never to be separated from their heads by any modifiers" (p.164). As highlighted at the outset, I believe that this fact tells us that while non-agreeing genitives sit low in the extended NP, agreeing genitives are located higher up.

The most obvious piece of evidence for the claim would be such that when additional modifiers (adjectives, numerals, demonstratives) are present in the extended NP, non-agreeing genitives follow them, while agreeing ones precede them. As much as this seems to be true, the textual evidence provides little material in terms of minimal pairs. The closest one can get is to juxtapose examples such as (18-a,b):

- (18) a. xul-ta    ma-t            [k**rtil-isa**] pur-ta-gan  
five-PL.OBL art-PL.OBL barley-GEN bread-PL.OBL-from  
'of the five barley loaves' p.164  
b. [m**qec-ta-j**    m**a-t**] u3yeb-**i**            igi          m3wnvareba-**j**  
beasts-GEN-NOM ART-PL insatiable-NOM ART.NOM raging-NOM  
'the insatiable raging of the beasts'

In the example (18-a), we have a non-agreeing genitive ‘of barley’ (boldfaced). It is located in between the initial numeral ‘five’, and a final head noun ‘bread.’ This indicates its relatively low position in the extended NP (lower than numerals). In the example, we also see a clitic article *ma-t* (belonging with the head noun ‘bread’), on which I comment in more detail below.

In (19-b), we see a complex agreeing genitive phrase ‘of the beasts’ (boldfaced), which includes a definite article in the position following its head (i.e., beasts). The whole agreeing genitive phrase is separated from the head by an adjective (insatiable), suggesting the genitive phrase is located higher than the adjectival modifier. It is also separated from the head by the clitic definite article *igi*, to which I turn below.

The examples (18-a,b) illustrate the asserted contrast such that agreeing genitives may – and non-agreeing genitives may not – be separated from the head. However, since one of the examples involves an adjective and the other one a numeral, they are not a minimal pair, and strictly speaking, tell us little about the precise height of the possessor.

However, agreeing and non-agreeing pre-nominal genitives behave differently also with respect to the placement of their head’s clitic definite article. In order to be able to interpret the facts and their implications, let me first provide a brief background on the location of the article in phrases that do not include any genitives. Put briefly, the article is a second position clitic. This means that in phrases where there is just the noun, the article comes after the noun; see (19-a). If a modifier precedes the noun, the article attaches after the first of these modifiers (19-b).

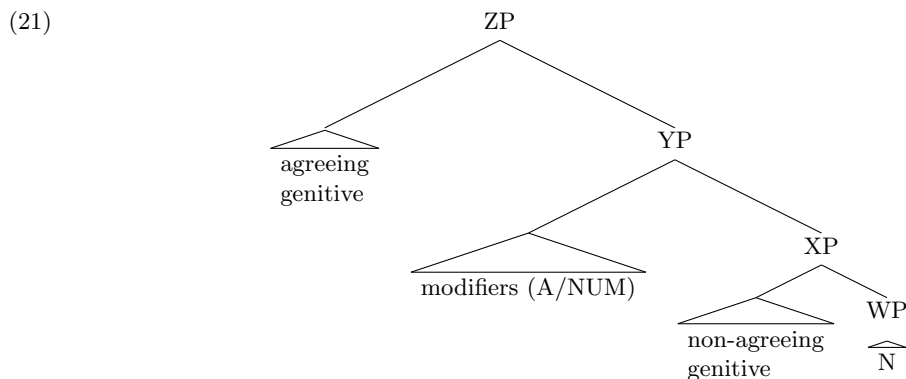
- (19) Boeder (1995)
- a. mter-i **igi**  
 enemy ART.NOM  
 ‘the enemy’ 154
  - b. cmida-j **igi** mcire-j eklesia-j  
 holy-NOM ART.NOM little-NOM church-NOM  
 ‘the holy little church’

From this perspective, agreeing pre-nominal genitives present no puzzle. The boldfaced clitic article follows the bracketed pre-nominal genitive (with the agreement marker factored out), see (20-a). On the other hand, as (20-b) shows, the location of the (boldfaced) article with a non-agreeing pre-nominal genitive (in brackets) is slightly more surprising. The article treats the non-agreeing genitive and its head as a single constituent, and appears in a position following the head, see (20-b):

- (20) Boeder (1995)
- a. [xuro-jsa]-j **igi** ʒe-j  
 carpenter-GEN-NOM ART.NOM son-NOM  
 ‘the son of the carpenter’ 163
  - b. [cmid-isa m-is] ʒma-n-i **igi**  
 holy-GEN ART-GEN brother-PL-NOM ART.NOM  
 ‘the brothers of the saint’

Correctly to my mind, Boeder (1995) interprets this in a way such that the non-agreeing genitival modifier differs from all other agreeing modifiers (including agreeing genitives) and forms a low level tight-knit constituent with the noun (perhaps a sort of a compound), whose inside is inaccessible for the clitic (just like the inside of a compound would be).<sup>8</sup>

What then emerges from the discussion can be structurally represented in the following way:



<sup>8</sup>Hypothetically, it could also be so that the non-agreeing modifier is actually located very high – above the base-generated site of the clitic, and the clitic simply goes to a second position within a domain that includes only the head noun. However, this ‘high’ interpretation of non-agreeing genitives does not square well with the facts of (18-a), where the non-agreeing genitive is clearly located inside the domain within which the second position is determined.

In words: non-agreeing genitives are low, agreeing genitives are high.<sup>9</sup> Equipped with this conclusion, I turn:

## 5 Back to Czech

Let me briefly sum up the relevant points made above. I started to explore the view that Czech numerals are a special type of a lexical item that spells out a phrasal constituent that includes a nominal projection at the bottom as well as a relatively high functional projection (Num). This leads to a bi-nominal analysis of Czech numerical phrases, an analysis which (unlike many alternatives) turns out to be compatible with No-Bundling. An empirical support for this analytical decision is provided by the fact that numerical phrases strongly resemble case attraction structures, a clear instance of a bi-nominal structure.

I have further sketched an analysis of attraction in terms of agreement marking of the dependent genitive plus a subsequent ellipsis of the genitive marking. From this perspective, the difference between attraction and non-attraction is whether the genitive noun has an additional agreement marker or not. Then, in order to see what conditions may govern such variation, I have turned to Old Georgian. In this language, we find relatively good evidence that non-agreeing genitives form a tight knit constituent with the noun, whereas agreeing genitives are located higher up.

There are reasons to believe that similar principles are at work in Czech, still independently of the particular numerical construction. To see that, consider the fact that in Czech, there are two basic ways of realizing additional nominal arguments in an extended NP. A non-agreeing genitive, and an agreeing genitive-like phrase.<sup>10</sup> With a class of nouns, the two can be combined, see (22). The interpretation makes it clear that the non-agreeing genitive (interpreted as the person depicted) must be lower than the agreeing nominal (interpreted as the possessor or author). The fact that the agreeing ‘genitive’ may serve as an antecedent for the non-agreeing one likewise points in the same direction.

- (22) [Petr-*ov-y* [obrazy seb-*e*] ]  
 Petr-poss-agr pictures self-GEN  
 ‘Petr’s pictures of himself’

There are other contrast that show the same thing, but I skip these for the interest of space. Instead, I turn directly to the numerical construction and apply the general observations to our specific case. Let me start by an analysis of numerals like ‘five,’ where the counted noun undergoes attraction obligatorily (see (2)). If attraction in these structures requires an agreeing genitive construction as its input, this means that at some level of representation, the counted noun must be an agreeing genitive (and cannot be a non-agreeing genitive). This in turn means that the counted noun must be located high (and not low) in the functional spine of the numeral, which acts as the head noun. Why should this be so?

The answer, I believe, emanates from the initial analysis of numerals in (4), repeated in (23-a), according to which numerals spell out a relatively large phrasal constituent.

- (23) a. /numeral/  $\Leftrightarrow$  [ Num [ N ] ]  
 b. [GEN] [Num N]

If we now further adopt the proposal (Neeleman and Szendrői 2007, Starke 2009, Radkevich 2009, Caha in press) that spell-out is restricted to constituents, then we derive the effect depicted in (23-b): the dependent genitive must be located no lower than Num. If it were lower, inside NumP, it would be impossible for the numeral to spell out NumP.

Now, under the proposal in (21), non-agreeing genitives are local to the noun (lower than Num, recall (18-a)). But if genitive dependents of numerals must be high (in order not to break the needed constituency, cf. Starke 2011, Caha 2011, Pantcheva 2011), then we rightly expect them to be of the agreeing sort.

On the other hand, ordinary nouns spell out only a relatively low NP node. Their genitive dependents may therefore be located lower than Num:

- (24) [Num [ [GEN] [ NP – the locus of spell out for ordinary nouns ] ] ]

<sup>9</sup>The post-nominal position of the agreeing genitives comes about as a result of moving a large projection of the noun across the agreeing genitive. Apparently, this movement is unavailable for the low NP node, and hence, non-agreeing genitives never end up after the noun.

<sup>10</sup>The agreeing nominal is not strictly speaking an agreeing genitive, even though it shares a number of traits with genitives (see Corbett 1987, 1995).



As a result of their low position, they will be of the non-agreeing (and consequently, non-attracting) type.

This basic outline of the analysis faces a challenge in the form of the numerals like ‘hundred’ and ‘thousand’ which allow both patterns (recall (9), repeated below):

- (25) a. Dal to st-u chlap-ů  
 he gave it hundred-DAT guys-GEN  
 b. Dal to st-u chlap-ům  
 he gave it hundred-DAT guys-DAT  
 ‘He gave it to hundred guys.’

In this aspect, ‘hundred’ and ‘thousand’ diverge from ordinary numerals. Interestingly, it is not the only place where they diverge from the rest of the numerals. First of all, ‘hundred’ and ‘thousand’ are the only numerals that may be counted without undergoing some sort of morphological modification (cf. Kayne 2006):<sup>11</sup>

- (26) a. \*dvě pět two five  
 b. \*dvě deset two ten  
 c. dvě st-a two hundred-PL  
 ‘two hundred’  
 d. dva tisíc-e two thousand-PL  
 ‘two thousand’

This is particularly striking in the case of ‘ten’. As shown in (26-b), the numeral cannot be counted. Correlating with it is the fact that its complement undergoes obligatory attraction:

- (27) a. \*Dal to deset-i chlap-ů  
 he gave it ten-DAT guys-GEN  
 b. Dal to deset-i chlap-ům  
 he gave it ten-DAT guys-DAT  
 ‘He gave it to hundred guys.’

This is so despite the fact that the number 10 is used in Czech to form complex numerals like ‘twenty’. However, these formations involve a different lexical item (just like English has *ten* vs. *-ty*):

- (28) dva-cet  
 two-ty  
 ‘twenty’

Thus, there are reasons to believe in the existence of a non-trivial correlation between the ability of the numeral (a lexical item) to be counted, and the ability of its counted noun to appear in the genitive in the oblique cases.

Similarly, the numerals ‘hundred’ and ‘thousand’ may take plural morphology in a type of approximative reading (cf. Kayne 2006); I give an example with *sto* ‘hundred’ in (29-a). *Deset* ‘ten’ cannot occur in this environment, as shown in (29-b). Thus, we have a correlation between the ability of a numeral to take plural morphology, and its ability to take a non-agreeing genitive dependent.

- (29) a. Ve článku byly st-a a st-a chyb.GEN  
 in article were hundred-s and hundred-s mistakes  
 ‘There were hundreds and hundreds of mistakes in the article’  
 b. \*Ve článku byly deset-y a deset-y chyb  
 in article were ten-s and ten-s mistakes.GEN  
 ‘lit. There were tens and tens of mistakes in the article’

Such type of data are usually taken to indicate that ‘hundred’ and ‘thousand’ are ambiguous between nouns (when they take a non-attracting genitive dependent, plural morphology, or when they are counted), and numerals (when their complement undergoes case attraction).

Such a situation is typical for semi-lexical items; to illustrate the point, consider some examples from English. *Front* has both a prepositional and a nominal use; *head* has a classifier use (as in *three head cattle*) as well as a purely nominal use; English *hundred* has a numerical use (as in *three hundred mistakes* – note the absence of plural and *of*) as well as a nominal use (*hundreds of mistakes* where both plural and *of* are there (Kayne 2006)). Such an ambiguity between a functional and a purely nominal use is common also for a number of ‘semi-lexical’ items in Czech.

<sup>11</sup>The variation between *dva* vs. *dvě* is conditioned by the gender of the head noun (*-a* for the masculine gender, *-ě* for the rest). Both forms are out in the ungrammatical examples.

From that perspective, ‘hundred’ and ‘thousand’ are actually well behaved ‘semi-lexical’ items. In the Nanosyntactic framework, their behavior may be accounted for straightforwardly by appeal to the so called Superset Principle:

- (30) *The Superset Principle (Starke 2009)*: A lexically stored tree matches a syntactic node iff the lexically stored tree contains the syntactic node.

In general, this means that an entry may spell out structures that correspond to its full specification, or a subset thereof. In the case of semi-lexical categories as understood here, (30) entails that such an item may act as a functional element (using its full specification), or an ordinary noun (using a subset of it). In our particular case, a numeral like (23-a) may use its full specification (and act as a numeral), or use only its lower part, and act as an ordinary noun. Thus, the ambiguous behavior of ‘hundred’ and ‘thousand’ is in fact predicted by the theory: they may ‘shrink’ down to the N node, and act as regular nouns. When they shrink, they may combine with non-attracting genitive dependents, because these no longer intervene in the constituent to be lexicalized.

From that perspective, what is rather problematic is the behavior of *pět* ‘five’ and its kin; according to The Superset Principle, these numerals too should be ‘shrinkable’ – but they are not. A possible (rather tentative) explanation for that follows.

Items that cannot shrink have been reported various places in the literature (Starke 2010, Dékány 2011, 131-3). A prototypical instance of such items are idioms. To see this on an example, consider the expression *kick the bucket* in the interpretation of ‘die.’ Because the meaning is not compositional, we need to state it in the lexicon. One way to achieve that is to rely on a large phrasal lexical entry which says that a constituent composed of the particular lexical items ([KICK [THE BUCKET]]) is interpreted as ‘die.’ However, no part of the idiom means ‘die’: the entry is unshrinkable.

Theoretically (drawing here on Starke’s unpublished work), unshrinkability has been attributed to the fact that idioms are phrasal lexical entries that make reference to other entries (e.g., the independent entry of *bucket*). Such a reference is encoded by a special device called ‘pointer’ (see Pantcheva and Caha 2012 for a more detailed presentation). By the definition of a pointer, an entry that has it may only be inserted if the particular entry pointed to has been inserted first in the structure. As a consequence, the entry for *kick the bucket* cannot insert the meaning of die for *kick*, since it may only be used if all the other entries pointed to have been inserted first. Having a pointer in the entry may then be the reason why some numerals cannot shrink.

The following entry shows one possible way to achieve this. It preserves the basic idea of this paper (numerals are phrasal), and adds beyond this the proposal that the entry for ‘unshrinkable’ numerals includes a pointer ( $\rightarrow$ ) to the lexical entry for ‘number,’ recalling Zweig’s (2006) proposal.

- (31)  $/pět/ \Leftrightarrow [\text{Num} \rightarrow [\text{NUMBER}]] \Leftrightarrow 5$

As a consequence of such an entry, the numeral cannot shrink down to the low NP node; this node would be spelled out as the noun *number* (as in *a number of examples*). It is only when such structure is augmented by Num, that the numeral *pět* may be inserted.

## 6 Ordering

According to the current proposal, numerals are phrasal nouns, and the whole construction that includes the numeral and the counted noun is a bi-nominal construction. There is a certain tension between this proposal and the observation that numerals figure as an integral part of ordering statements that generalize over a *single* extended NP (thinking in particular of Cinque 2005). This section provides some general considerations, as well as relevant Czech specific discussion.

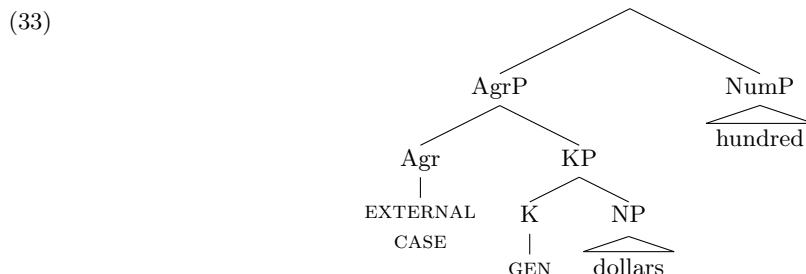
Starting with general remarks, let me mention that the current proposal does not – on its own – prevent numerals from participating in the type of structures that Cinque (2005) proposes. Specifically, it seems possible that these items are merged in Spec,NumP (under Spec-head agreement, since they are NumPs themselves), and simply participate in an extended projection with the counted noun as the ‘head’ (the counted noun receives no genitive marking):<sup>12</sup>

- (32)
- 
- ```

graph TD
    NumP1[NumP] --- NumP2[NumP]
    NumP1 --- Num[Num]
    NumP1 --- NP[NP]
    NumP2 --- hundred[hundred]
    NP --- dollars[dollars]
  
```

<sup>12</sup>Adopting the ideas presented in Starke (2004) would allow for a direct merger of the complex NumP with a noun phrase complement (without the empty head Num).

This is still compatible with their nominal status (due to the presence of N at the bottom of the constituent spelled out by the numeral). It is not clear to me why Czech higher numerals cannot make use of this option.<sup>13</sup> But whatever the reasons, the facts suggest that these numerals act as the head of the whole bi-nominal complex, and require that the counted noun is generated in their Spec, with an accompanying agreement marker that tracks the case of the whole bi-nominal phrase. This is depicted in (33).



There are two issues to take into consideration concerning this structure. Both of them revolve around the fact that the proposal (33) seems to allow the generation of the sequence Num-Dem-A-N. However, Cinque (2005) observes that such a sequence does not seem to be attested, and hence, something should be said about this.

The first issue starts from the fact that the actual surface ordering between the genitive and the numeral is Num>GEN. This is not reflected in the structure (33) (which yields GEN>Num). The simplest solution is to assume in addition that the (phrasal projection of the) numeral moves across the genitive. The general worry about this move is the following: if we allow numerals to move, we derive unattested orders. To see that, suppose that we add a demonstrative on top of (33); then, allowing numerals to move across Dem would yield Num-Dem-[A-N] or Num-Dem-[N-A], both of which are cross-linguistically unattested orders (Cinque 2005).

The second issue concerns the projection of the counted noun. It seems that the size of this projection has to be restricted not to include high projections – the projection where demonstratives are introduced among them. Once again, it seems that allowing a demonstrative to be generated inside the projection of the counted noun in (33), and moving the numeral across the genitive, would lead to the order Num-[Dem-A-N], which is cross-linguistically unattested.

One way to approach the problem would be to look for ways to (i) restrict the movement of the numeral in some way, and (ii) stipulate a restriction on the size of the counted noun. However, I follow a different track: I am going to suggest that these derivations are legal, but they yield a different meaning. If that is so, then the current approach actually correctly predicts a range of examples that would require a special mechanism under the standard account.

I start from the observation that there are in fact sequences in Czech that have the shape Num-Dem-A-N, an example of which is below:

- (34) pět            těch        chlap-ů  
 five.NOM/ACC those.GEN guys-GEN.PL  
 ‘five of those guys’

However, as the translation makes it clear, these sequences have a partitive reading. For this reason, they are not considered in the typological literature (including Cinque’s contribution) as an instance of a basic ordering between Dem, Num and N. Note that examples of this type are subject to attraction, see (35), and hence, there are reasons to believe that they have a similar base structure as the non-partitive examples (with the counted noun an agreeing genitive).

- (35) pěti        těm        chlap-ům  
 five.DAT those.DAT guys-DAT.PL  
 ‘to five of those guys’

Given the existence of such examples, it seems advantageous to consider the option that at least one of the apparently problematic derivations is in fact needed (perhaps both), and yields a partitive effect. Given the page limit, I cannot elaborate on this in detail, but there are reasons to think that the movement derivation might be on the right track. Just to indicate briefly: (i) examples such as

<sup>13</sup>Note that Czech lower numerals may do so; they require no genitive:

- (i) dva chlap-i  
 two guy-NOM.PL  
 ‘two guys’

*six those-gen five-gen men-gen* ‘six out of these five men’ are unattested; that is compatible with the idea that we are moving the numeral high across a small counted noun; (ii) with low numerals, one finds also *three-nom those-nom men-nom* ‘three of those men,’ which is once again compatible with moving the numeral in a structure like (33).

## 7 Conclusions

This paper argued that if Czech numerals are analyzed as phrasal lexical items, with a noun at the bottom, we go a large way towards understanding their peculiar behavior when it comes to the case marking of their counted noun. Similarly, we gain a straightforward understanding of why some numerals (hundred and thousand) are ambiguous between nouns and numerals. Finally, taking numerals to be nominal increases their movement options (in the system of Cinque 2005). It turns out that this is a welcome option that allows us to capture the existence of certain non-canonical orders.

The larger ambition of the paper is to contribute to the understanding of ‘semi-lexical’ categories. The basic idea is that their properties (both lexical and functional) may be captured under a phrasal spell out approach.

## References

- Adger, David. 2012. A syntax of substance. Ms., Queen Mary, London.
- Babby, Leonard H. 1987. Case, pre-quantifiers, and discontinuous agreement in Russian. *Natural Language and Linguistic Theory* 5: 91–138.
- Boeder, Winfried. 1995. Suffixaufnahme in Kartvelian. In *Double Case: Agreement by Suffixaufnahme*, edited by Frans Plank, pp. 151–215. OUP, Oxford.
- Brattico, Pauli. 2011. Case assignment, case concord, and the quantificational case construction. *Lingua* 121.6: 10421066.
- Caha, Pavel. 2009. *The Nanosyntax of Case*. Ph.D. thesis, CASTL, University of Tromsø.
- Caha, Pavel. 2011. The parameters of case marking and spell out driven movement. *Linguistic variation yearbook 2010* 10: 33–78.
- Caha, Pavel. in press. Explaining the structure of case paradigms through the mechanisms of Nanosyntax. *Natural Language and Linguistic Theory*.
- Cinque, Guglielmo. 2005. Deriving Greenberg’s universal 20 and its exceptions. *Linguistic Inquiry* 36: 315 – 332.
- Cinque, Guglielmo and Luigi Rizzi. 2010. The cartography of syntactic structures. In *The Oxford handbook of linguistic analysis*, edited by Bernd Heine and Heiko Narrog, pp. 51–65. Oxford University Press, Oxford.
- Corbett, Greville G. 1978. Universals in the syntax of cardinal numerals. *Lingua* 46 4: 355– 368.
- Corbett, Greville G. 1987. The morphology-syntax interface. *Language* 63: 299–345.
- Corbett, Greville G. 1995. Slavonic’s closest approach to suffix copying: the possessive adjective. In *Double Case: Agreement by Suffixaufnahme*, edited by Frans Plank, pp. 265–282. OUP, Oxford.
- Danon, Gabi. 2012. Two structures for numeral-noun constructions. *Lingua* 122 12: 1282–1307.
- Dékány, Éva. 2011. *A profile of the Hungarian DP. The interaction of lexicalization, agreement and linearization with the functional sequence*. Ph.D. thesis, CASTL, University of Tromsø.
- Franks, Steven. 1994. Parametric properties of numeral phrases in Slavic. *Natural Language and Linguistic Theory* 12: 597–674.
- Giusti, Giuliana and Nedžad Leko. 2005. The categorial status of quantity expressions. *Lingvistički vidici* 34: 121–183.
- Ionin, Tania and Ora Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics* 23: 315–360.

- Kayne, Richard S. 2004. Prepositions as probes. In *Structures and Beyond: The Cartography of Syntactic Structures, vol. 3*, edited by Adriana Belletti, pp. 192–212. Oxford, New York.
- Kayne, Richard S. 2005. Some notes on comparative syntax, with special reference to English and French. In *The Oxford Handbook of Comparative Syntax*, edited by Guglielmo Cinque and Richard S. Kayne, pp. 3–69. Oxford University Press, Oxford.
- Kayne, Richard S. 2006. A note on the syntax of numerical bases. In *In Search of the Essence of Language Science: Festschrift for Professor Heizo Nakajima on the Occasion of his Sixtieth Birthday*, edited by Y. Suzuki, pp. 21–41. Hituzi Syobo, Tokyo.
- Neeleman, Ad and Kriszta Szendrői. 2007. Radical pro-drop and the morphology of pronouns. *Linguistic Inquiry* 38: 671–714.
- Pantcheva, Marina. 2011. *Decomposing Path. The nanosyntax of directional expressions*. Ph.D. thesis, CASTL, Tromsø.
- Pantcheva, Marina and Pavel Caha. 2012. Contiguity beyond linearity. Talk at Worse, University of Tromsø, June.
- Pereltsveig, Asya. 2007. On the universality of DP: A view from Russian. *Studia Linguistica* 61.1: 59–94.
- Plank, Frans. 1995. (Re-)introducing suffixaufnahme. In *Double Case: Agreement by Suffixaufnahme*, edited by Frans Plank, pp. 3–112. OUP, Oxford.
- Radkevich, Nina. 2009. Vocabulary insertion and the geometry of local cases. Ms., UConn, downloadable at: <http://ling.auf.net/lingBuzz/000958>.
- van Riemsdijk, Henk. 1998. Categorial feature magnetism: The endocentricity and distribution of projections. *The Journal of Comparative Germanic Linguistics* 5: 1–48.
- Rutkowski, Pavel. 2006. Why Polish numerals should not be analyzed as nouns. In *Minimalist Views on Language Design: Proceedings of the 8th Seoul International Conference on Generative Grammar*, edited by Steven Franks, Frank Gladney, and Mila Tasseva-Kurktchieva, pp. 249–263. Hankook/Korean Generative Grammar Circle, Seoul.
- Starke, Michal. 2004. On the inexistence of specifiers and the nature of heads. In *Structures and Beyond: The Cartography of Syntactic Structures, vol. 3*, edited by Adriana Belletti, pp. 252–268. Oxford University Press, New York.
- Starke, Michal. 2009. Nanosyntax. A short primer to a new approach to language. In *Nordlyd 36: Special issue on Nanosyntax*, edited by Peter Svenonius, Gillian Ramchand, Michal Starke, and Tarald Taraldsen, pp. 1–6. University of Tromsø, Tromsø. Available at [www.ub.uit.no/munin/nordlyd/](http://www.ub.uit.no/munin/nordlyd/).
- Starke, Michal. 2010. Universal grammar vs. lexically driven derivations, once more. Talk at the SmaSh workshop, University of South Bohemia, České Budějovice, June 23.
- Starke, Michal. 2011. Towards elegant parameters: Language variation reduces to the size of lexically stored trees. Transcript from a talk at Barcelona Workshop on Linguistic Variation in the Minimalist Framework. Available at <http://ling.auf.net/lingBuzz/001183>.
- Zweig, Eytan. 2006. Nouns and Adjectives in Numeral NPs. In *NELS 35: Proceedings of the thirty-fifth annual meeting of the North East Linguistic Society*, edited by L. Bateman and C. Ussery, pp. 663–675.