

# COUNTING AND THE GRAMMAR: CASE AND NUMERALS IN INARI SAMI

Diane Nelson & Ida Toivonen

## Abstract

This paper examines new data from Inari Sami which pose a challenge for two basic principles within generative grammar, namely that the lexicon consists of an unordered set, and that rules of the grammar cannot “count”. Drawing on evidence from the system of case and numerals in Inari Sami, we first discuss the data in light of Case Theory, in order to identify the structural relations and mechanisms involved in Inari Sami. We then examine several theories of numerals from a cross-linguistic perspective, and conclude that none of the existing theories can account for the Inari Sami data while still maintaining the assumptions mentioned above. We claim that the best explanation for the data lies in a structured, ordered lexicon, in which certain lexical items are acquired and stored in a particular numerical order, to which morphosyntactic rules and processes may be sensitive.

## 1 Introduction

One fundamental assumption underlying most syntactic theories is that rules of the grammar cannot involve “counting”<sup>1</sup>. This notion originates in the earliest work in generative grammar and derives from the requirement that syntactic operations must be structure-dependent (ie sensitive to category labels) rather than sensitive to the linear order of elements in an input string. For instance, Chomsky (1965:55) argues that structure-dependent syntactic transformations are not necessarily the simplest computationally:

... grammatical transformations are necessarily “structure-dependent” in that they manipulate substrings only in terms of their assignment to categories... It is impossible, however, to formulate as a transformation such a simple operation as reflection of an arbitrary string (that is, replacement of any string  $a_1 \dots a_n$ , where each  $a_i$  is a single symbol, by  $a_n \dots a_1$ ) or interchange of the  $(2n-1)^{\text{th}}$  word with the  $2n^{\text{th}}$  word throughout a string of arbitrary length, or insertion of a symbol in the middle of a string of even length.

This prohibition on rules that “count” is taken to be a fundamental constraint on the type of syntactic rule that can be acquired. For example, it rules out the acquisition of a rule which forms a yes-no question by moving the fourth word of the input string to the front of the sentence. The human cognitive system may be capable of counting members of a set, but according to standard assumptions, UG disallows linguistic

---

<sup>1</sup> This is a slightly revised version of the paper that appeared in the original (electronic) version of *Leeds Working Papers* 8 (2000). The authors would like to thank A. Kuuva, S. Kuuva, K. Mattus and I. Mattus for invaluable informant help. The data in this paper was collected during fieldwork during the summer of 2000, and is based on information from Sammallahti & Morottaja (1993) and Sammallahti (1998). Thanks are also due to Satu Manninen, Pekka Sammallahti, Greville Corbett, Jim Hurford, and Frederick Newmeyer for valuable discussion. All errors are the responsibility of the authors. This work was supported by British Academy research grant number SG-31040.

rules which manipulate elements based on assigning a numerical value to members of an input string<sup>2</sup>.

Another, related assumption in the generative literature is that the lexicon consists of an unordered list of items, characterized by idiosyncrasy and unpredictability (Chomsky 1965:84, Williams 1994). There are differences across frameworks as to how structured the lexicon is assumed to be,<sup>3</sup> but there is general agreement that lexical entries are not organized in any kind of numerical order or linear string. If the contents of the lexicon are indeed unordered, this again excludes the possibility of “counting” for any rule that applies there.

The case system of Inari Sami seems to provide a counterexample to these generalisations. In Inari Sami, numerals 2 to 6 have different case-assigning properties to numerals 7 and above, suggesting that in this language either a syntactic counting rule is invoked or that at least a part of the lexicon is stored as an ordered string. This paper discusses the case and numeral systems of Inari Sami and other languages. The data, it turns out, raise challenging questions for the “no counting” learnability principles as well as for the distinction between lexical and structural case, and the organisation of the lexicon. Section 2 presents the relevant data from Inari Sami. Section 3 discusses these data in light of Case Theory, and proposes an analysis using the notion of direct case. Section 4 looks at numerals in the grammar from a cross-linguistic perspective. We argue that the facts from Inari Sami provide evidence for a lexicon more structured than previously assumed.

## 2 Case and numerals in Inari Sami: The data

In Inari Sami, one of the Sami languages spoken in the North of Finland around Lake Inari, there is a split between the number six and the number seven with respect to the case marking on the following numeral (Sammallahti 1998:70).<sup>4</sup> Numerals with values 2-6 assign accusative case to the nouns they quantify, while those with values 7 and higher assign partitive case:

- (1) a. kyehti/kulmâ/nelji /vittâ/kuttâ poccuu  
two /three /four /five /six reindeer.acc.sg
- b. kyehti/kulmâ/nelji /vittâ/kuttâ pääрни  
two /three /four /five /six child-acc.sg
- (2) a. čiččâm /kávci /ovce /love /ohtnubáloh /kyehtnubáloh /čyeti ... poccud<sup>5</sup>  
seven /eight /nine /ten /eleven /twelve /100... reindeer.part.sg
- b. čiččâm /kávci /ovce /love /ohtnubáloh /kyehtnubáloh /čyeti ... päρνid  
seven /eight /nine /ten /eleven /twelve /100... child-part.sg

<sup>2</sup> We will ignore features of the grammar which seem to involve “counting” from one to two, eg Verb Second in Germanic, the distribution of Wackernagel’s clitics and the role of Bounding nodes in extraction. Many analyses of these phenomena have been proposed which derive these effects from other principles, see e.g., Halpern 1995, and the works in Halpern and Zwicky 1996.

<sup>3</sup> For example, in the theoretical framework of HPSG, the lexicon is often organized in hierarchies (see Pollard and Sag 1987).

<sup>4</sup> Inari Sami has this in common with the other Eastern Sami languages, Skolt, Akkala, Kildin, and Ter Sami. This system differs from that of Finnish and other Finnic languages, where partitive case is assigned to numerals 2 and higher.

<sup>5</sup> There is an alternative partitive singular form in use for this noun: *puásujid*. This is a dialectal variant. What is important is that all speakers distinguish between accusative and partitive.

Let us first look at the numerals 2-6. When the NP is an object, like in (3-4), it is unclear whether the numeral demands an accusative noun, or whether the noun is in accusative case because accusative is the normal object case marking:

(3) Mun jiem oomâst kulmâ poccuu.  
 I.nom not.1sg own three reindeer.acc.sg  
 ‘I don’t own three reindeer.’

(4) Mun uáinám kulmâ kiedâ.  
 I.nom see.1sg three hand.acc.sg  
 ‘I see three hands.’

However, when the NP containing a numeral is a subject, the noun still gets accusative case marking, although the normal subject case is nominative:

(5) a. Kulmâ kiedâ išeđii muu.  
 three hand.acc.sg helped.3pl me.acc  
 ‘Three hands helped me.’

b. \*Kulmâ kietâ išeđii muu.  
 three hand.nom.sg helped.3pl me.acc

(6) a. Nelji poccuu ruottii máttâl.  
 four reindeer.acc.sg ran.3pl away  
 ‘Four reindeer ran away.’

b. \*Nelji puásui ruottii máttâl.  
 four reindeer.nom.sg ran.3pl away

The examples in (3-6) show that nominals following the numerals two through six are assigned accusative case, regardless of the syntactic position of the NP containing them.

Similarly, nominals following numerals of value 7 and higher are partitive<sup>6</sup>, whether or not the NP is a syntactic subject or object:

(7) a. Čiččâm poccud láá tobbeen.  
 seven reindeer.part.sg are.3pl there  
 ‘Seven reindeer are there.’

b. \*Čiččâm puásui láá tobbeen.  
 Seven reindeer.nom.sg are.3pl there

---

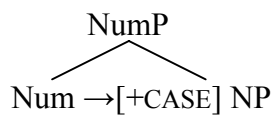
<sup>6</sup> Unlike in Finnish, partitive case is generally unproductive in the Sami languages and does not encode aspectual information. In Inari Sami, partitive is used after numerals, and by some speakers also in comparative expressions:

i. Sun lii muđe puárrsup.  
 he is I.part older  
 ‘He is older than me.’

- (8) a. Tun oinih kyehtnubáloh poccud.  
 You.nom saw.2sg twelve reindeer.part.sg  
 ‘You(sg) saw twelve reindeer.’
- b. \*Tun oinih kyehtnubáloh poccuu.  
 You.nom saw.2sg twelve reindeer.acc.sg

The noun *poccud* in (7) must bear partitive case, not nominative, although it is the subject. Likewise, in (8), the object is necessarily partitive, not accusative.

Another interesting fact illustrated by the examples above is that singular nouns in accusative and partitive case trigger plural subject agreement on the verb (5a, 6a, 7a). This suggests that the numeral, not the noun, is the syntactic head of the quantified noun phrase in Inari Sami and encodes the morphosyntactic feature NUMBER.



In contrast, in Finnish, numerals 2 and higher assign partitive singular case to nouns they quantify, but verb agreement is singular:

- (9) a. Yhdeksän omena-a puto-si maa-han.  
 nine.nom apple.part.sg fall.past.3s earth.ill  
 ‘Nine apples fell to earth.’
- b. \*Yhdeksän omena-a puto-sivat maa-han.  
 nine.nom apple.part.sg fall.past.3p earth.ill

In the examples in (9), the partitive singular noun seems to trigger singular verbal agreement as the phrasal head. This syntactic difference in headedness is associated with a morphological difference between the two languages: Finnish numerals inflect for the full range of possible case and number feature combinations, whereas numerals over 2 in Inari Sami are indeclinable when they precede the nominals they modify.

The numeral one in Inari Sami, *ohtâ*, patterns differently from the other numerals. Firstly, nominals quantified by it are transparent to normal structural case marking to the phrase as a whole:

- (10) Ohtâ kietâ lii tobbeen.  
 one.nom hand.nom.sg is.3sg there  
 ‘One hand is there.’
- (11) Mun uáinám ovdâ kieðâ.  
 I.nom see.1sg one.acc hand.acc.sg  
 ‘I see one hand.’

In (10-11), the noun is case marked according to grammatical function; the subject is nominative and the object is accusative. Secondly, the numeral one, *ohtâ*, inflects for structural case to agree with the noun. Because *ohtâ* patterns so differently from the

other numerals, the remainder of the discussion in this paper will focus on the higher value numerals.

All the examples we have seen until now concern grammatically case marked noun phrases. That is, we have only looked at canonical subjects and objects, which are normally assigned nominative and accusative case, respectively. Let us now turn to noun phrases that receive other types of case marking, for example locative. There are two possibilities: either this kind of case marking overrides the accusative case assigned by the numerals 2-6 and the partitive assigned by the numerals 7 and higher, or the case assigned by the numerals overrides the lexical and semantic case marking. We find that the first hypothesis is correct. Consider examples (12-13):

- (12) a. **Muste** laa kuttâ poccuu.  
I.loc are.3pl six reindeer.acc.sg  
'I have three reindeer.'
- b. Kulmâ **poccust** lii ennuu purrâmâš.  
Three reindeer.loc.sg is.3sg much food.nom  
'Three reindeer have a lot of food.'
- c. Čiččâm **poccust** lii ennuu purrâmâš  
seven reindeer.loc.sg is.3sg much food.nom  
'Seven reindeer have a lot of food.'
- (13) a. Mun adelim kiirjijd **almaj/almajd**.  
I.nom gave book.acc.pl man.ill.sg/man.ill.pl  
'I gave books to the man/the men.'
- b. Mun adelim kiirjijd viiðâ **almaj**.  
I.nom gave book.acc.pl five man.ill.sg  
'I gave books to five men.'
- c. Mun adelim kiirjijd kâvci **almaj**.  
I.nom gave book.acc.pl eight man.ill.sg  
'I gave books to eight men.'

The first argument of the possessive construction in (12) (literally 'on X is Y', where X is the possessor and Y is the item possessed) takes locative case, as we see in (12a). The head noun of the first NP also bears locative case, although it is preceded by a numeral, which normally assigns singular accusative (12b) or partitive (12c). The indirect object of the verb *addeliđ* 'to give' in Inari Sami normally bears illative case, as illustrated by (13a). As we can see, this is not overridden by the numeral case assignments (13b-c).

Table 1 summarizes the generalizations that emerge from the data presented in this section:

**Table 1.**

	Grammatically case-marked subject	Grammatically case-marked object	Lexically or semantically case marked NP
Ns not modified by	nominative	accusative	lexical/semantic

numerals; Ns modified by the number one			case-marking wins
Ns modified by the numbers 2-6	accusative	accusative	“
Ns modified by the numbers 7 and higher	partitive	partitive	“

Before we discuss why the numeral split between 6 and 7 is problematic for theories of grammar and the lexicon, we need to identify the type of case (structural, semantic or lexical) being assigned by the numeral to the noun it governs. The mechanisms underlying case assignment in Inari Sami are analysed in the next section.

### 3 How Numerals in Inari Sami Assign Case

We have seen in the previous section that numerals 2-6 in Inari Sami assign accusative case to the nouns they govern, while numerals 7 and above assign partitive, regardless of the structural position of the NumP. However, when the entire NumP gets locative case marking from the verb, the case assigned by the numeral gets overridden. In order to clarify the lexical properties of the numerals in question, it is necessary to work out which type of case they assign to the nouns they quantify. According to standard Case Theory, there are at least four main types:

- i. semantic
- ii. lexical (inherent/thematic)
- iii. lexical (idiosyncratic)
- iv. structural (syntactic)

Semantic case is associated with semantically transparent non-arguments (typically adverbials). Since we can assume that the numerals and nouns are in some type of head-complement relationship, this option can be ruled out.

Yip, Maling & Jackendoff (1987) distinguish between two kinds of lexical case. The first type is associated with thematic arguments of verbs, linked with particular structural positions. In addition, they argue, cases may be assigned lexically within NPs to, for example, possessors (though not in Icelandic, the main language discussed in their paper). The problem with analysing numerically-assigned accusative and partitive as lexical case is that it is unclear what sort of *thematic* role a numeral might assign to a noun it quantifies. The numeral-noun relationship within a noun phrase is distinct from more clear-cut environments for lexical case, for example possessor-noun, in which the casemarked element is linked to a well-defined thematic argument. Supposing we did want to analyse the accusative and partitive as thematic lexical cases, another problem soon arises. Why would the relationship between numerals 2-6 have a different *thematic* relationship to the nouns they quantify from numerals 7 and above, which results in the realisation of a different morphological case? It is clear then, that the partitive-accusative alternation under numerals in Inari Sami is in some sense due to special lexical properties of either numerals 2-6 or 7 and above, but this does not entail that these numerals assign thematic lexical case.

The other type of lexical case in Yip, Maling and Jackendoff's (1987) model is idiosyncratic in nature, and cannot be predicted by thematic relations. This type of case is associated with, for example, accusative and genitive “quirky” subjects in Icelandic. Since there is no evidence in Inari Sami for semantically- or thematically-based case assignment by numerals, the simplest analysis for the patterns of data

presented above is that one case assigned by numerals is a structural default, and the other is an idiosyncratic lexical case.<sup>7</sup> Since numerals 7 and above form an infinite set, it makes sense to posit the special lexical case feature as being associated with the smaller subset of numerals 2-6, which assign accusative case. Numerals 7 and above, then, assign partitive case structurally as a default.<sup>8</sup>

- i. numerals 2-6: assign accusative case (lexical case)
- ii. numerals 7 and above: assign partitive case (structural default)

If this analysis is correct, the question remains as to what happens to these cases when quantified NPs are embedded under lexical case assigning verbs. Yip, Maling and Jackendoff's (1987:240) analysis makes the following prediction, again to capture patterns of case in Icelandic:

Lexical case on an NP always blocks/inhibits assignment of syntactic [structural] case to that NP.

Work on "case conflict" in Free Relatives (FRs), where a single argument has the potential to get case both from the matrix verb and from the subordinate predicate, reaches similar conclusions. For instance, Vogel (2000) notes that cross-linguistically, it is structural cases that get suppressed in free relatives, not obliques. Vogel argues that case-matching facts from German and other Indo-European languages show that cross-linguistically, where two cases may potentially be assigned to a single NP, it is the structural cases (nominative and accusative) which will be overridden by oblique cases (dative etc) rather than vice-versa. If a matrix verb imposes nominative or accusative case on its relative clause complement, this case will always be suppressed in favour of the case assigned by the verb inside the free relative (14). If the matrix verb imposes an oblique case on the relative pronoun, then it can only appear in free relatives where the relative verb also requires an oblique case (15) (German, from Vogel 2000:3):

- ( 14) a. Weil uns besucht, wer Maria mag. (FR verb requires nom)  
because us-acc visits, who-nom Maria-acc likes  
"...because whoever likes Maria is visiting us."
- b. Weil uns besucht, wen Maria mag.  
because us-acc visits, who-acc Maria-nom likes (FR verb requires acc)  
"...because whoever Maria likes is visiting us."
- c. Weil uns besucht, wem Maria vertraut.  
because us-acc visits, who-dat Maria-nom trusts (FR verb requires dat)  
"...because whoever Maria trusts is visiting us."
- ( 15) a. Ich folge wem immer ich **vertraue**  
I follow who-dat ever I trust  
'I follow whomever I trust'

---

<sup>7</sup> See Ojeda (1997) for a similar analysis of Classical Arabic.

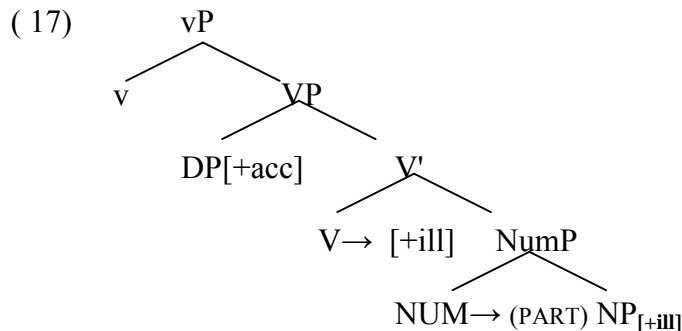
<sup>8</sup> See, for example, Vainikka & Maling (1996) and Nelson (1998) for arguments that partitive is a structural case in Finnish.

- b. \*Ich folge wen immer ich **bewundere**  
 I follow who-acc ever I adore  
*folgen* ‘follow’ assigns dative, *bewundern* ‘adore’ assigns accusative

In (14), the matrix verb requires a structural case-marked complement in nominative case. However, the pronoun always receives case from the verb inside the free relative clause, whether it is nominative (14a), accusative (14b) or dative (14c). The structural case assigned by the matrix is therefore suppressed in favour of the FR verb’s case. On the other hand, if the matrix verb requires an oblique dative complement (15), then this case cannot be overridden by the structural case assigned by the lower verb as in (14), as shown in (15b). Instead, the cases assigned by both verbs must match (15a). Similar patterns emerge in Modern Greek, Spanish and Romanian. This suggests that cross-linguistically, if a single argument can get case from two lexical heads, it is the structural case which gets suppressed in favour of the oblique case.

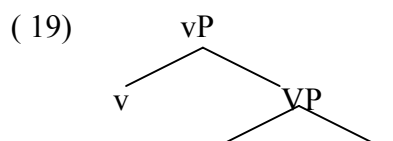
Given this prediction, embedded NPs headed by numerals 7 and above pattern as expected in Inari Sami: lexical locative case assigned by the verb suppresses the structural case (partitive) assigned by the numeral. This is shown in example (13c) with the ditransitive verb *addeliđ*, ‘to give’, repeated here as (16). (Ditransitive verbs are assumed to license a case-assigning VP-shell<sup>9</sup> (Larson 1988)):

- (16) Mun adelim kiirjijd kávcı **almı**.  
 I.nom gave book.acc.pl eight man.ill.sg  
 ‘I gave books to eight men.’



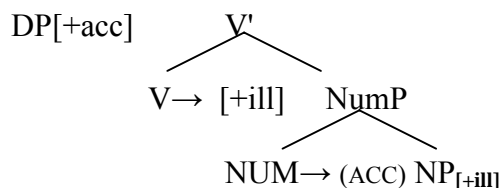
For NumPs quantified by numerals 2-6, however, a problem arises. Recall from section 2 that evidence from agreement suggests that it is the numeral, not the noun, which heads the number phrase in Inari Sami. If the head numerals 2-6 assign lexical rather than structural case to the nouns they quantify, then why does the lexical case assigned by the verb appear to suppress the more local lexical case assigned by the numeral? This is the pattern seen in (13b), repeated here as (18):

- (18) Mun adelim kiirjijd viiđâ **almı**.  
 I.nom gave book.acc.pl five man.ill.sg  
 ‘I gave books to five men.’



<sup>9</sup> It is not relevant for the current discussion whether case is assigned under sisterhood or checked under spec-head agreement via functional projections.





If the accusative case assigned by the numeral is genuinely lexical, then basic morphological principles dictate that case realized on the numeral should reflect the case assigned by the more local (governing) specific assigning head, ie accusative assigned by the numeral. Instead, accusative case is overridden by locative case assigned by the distal head.

What we need, then, is an approach in which lexical accusative has different properties from other lexical cases. Woolford (2000) looks at the distribution of nominative objects in a variety of languages in light of Burzio’s Generalization. She notes that nominative objects in languages like Icelandic are disallowed when the subject is (lexical) accusative (\*ACC-NOM case frame), but allowed when the subject shows lexical dative (yielding the case frame DAT-NOM). Certain languages including Icelandic also allow ACC-ACC case frames. For this reason, she conflates morphological accusative, both structural and lexical, along with nominative into one case type, *direct case*, which encodes distinct morphosyntactic features that must be checked according to general markedness principles. Direct case captures the fact that lexical accusative case shares mixed properties of both lexical oblique and structural accusative case. If Inari Sami numerals assign direct case, then there is no direct conflict between two lexical cases, and the more “structural” of the two is expected to be suppressed<sup>10</sup>.

In this section, the morphosyntactic properties of the cases assigned by numerals were analysed in detail. The accusative case feature assigned by Inari Sami numerals 2-6 was argued to be distinct from other types of lexical cases, such as locative. The case assigning properties of Inari Sami numerals can thus be revised as follows:

- i. numerals 2-6: assign accusative case (direct case)
- ii. numerals 7 and above: assign partitive case (structural default)

The question remains, however, as to how a rule like (i) gets acquired if the lexicon consists of an unordered set. This issue is addressed in the next section.

<sup>10</sup> This analysis also captures similar facts from Finnish, where numerically-assigned partitive case may be overridden by genitive case, which has been argued to be assigned as a structural default to nominal elements in specifier positions (Vainikka 1989):

- i. Näi-n kaksi talo-a  
saw-1s two.nom house-part  
‘I saw two houses’
- ii. Näin kahde-n talo-n palavan  
saw-1s two-gen house-gen burning  
‘I saw two houses burning’

Thanks to Satu Manninen for pointing this out and providing this example.

#### 4 Numerals in the grammar and lexicon

In the previous section, we provided an analysis of case morphology within Numeral Phrases in Inari Sami. We argued that partitive case is assigned by numerals 7 and higher as a structural default, and that numerals 2-6 are specified for a lexical feature that assigns direct accusative case to the quantified noun. This means that numerals 2-6 must form a special lexical subclass in Inari Sami. But if this is the case several important questions arise about the structure of the lexicon and the grammar. If the lexicon does indeed consist of an unordered set, and if no rule of the grammar can be sensitive to the linear order of elements in a sequence or string, then:

- (a) How is this subset acquired? Does a child acquiring Inari Sami have to use a rule in the lexicon or grammar that counts to acquire this special subset?
- (b) How is information relating to lexical classes and subclasses organised in the lexicon?

The evidence from Inari Sami suggests that core numerals in the lexicon are not listed in an unordered, idiosyncratic fashion. Instead, it appears that they are listed in an ordered sequence, and lexically specified morphosyntactic properties may be acquired and stored in a way that is sensitive to this sequence.

Because numerals make up an infinite set in every language, they represent a unique class of lexical item. Hurford (1975, 1987) argues for a set of (language-specific) syntactic rules underlying the generation of higher-value *complex numerals* which draws on a core set of *lexical numerals*. Once the set of rules for forming numerals has been acquired, a speaker of a language is able to compute and generate an unlimited number of numerals; in this respect higher-level numerals behave more like phrasal categories than lexical heads (words). However, the generation of complex numerals is taken to preserve the category label, and therefore the syntactic distribution, of the input. For example, the English complex numeral *thirty-four thousand, six hundred and eighty-eight* (34,688) has the same syntactic properties as the numeral *thirty-four thousand, six hundred and eighty-seven* (34,687) because the operation that derives the new numeral has no effect on the category label of the input. Thus for every rule R that generates a new lexical item deriving the complex numeral  $N_2$  from  $N_1$ , we can say that rule R does not affect the set of morphosyntactic properties of the input lexeme. If  $N_1$  encodes for example the features [+N, +feminine] in a given language,  $N_2$  should possess an identical set of features in its lexical entry. In other words, the arithmetic value of the complex numeral is predicted to have no bearing on its syntactic distribution or its morphosyntactic properties.

While this generalisation may apply to higher-level complex numerals formed by syntactic processes, the picture is less clear with the core lexical numerals of lower value which form the input to the syntactic processes mentioned above, such as 1-10, 100, and so on. Hurford (forthcoming) notes that lexical numerals are not homogeneous with respect to morphosyntactic properties; cross-linguistically, these numerals may fall into special subclasses. For example, the numerals 1-4 in Icelandic inflect for case and gender, while 5 and above are indeclinable<sup>11</sup>:

##### (1) Icelandic accusative case-marked numerals

---

<sup>11</sup> In Icelandic and some other languages, syntactic properties of the output numeral may be determined by the final digit in the string, eg in Icelandic the numerals 1-4 and all complex numerals ending in the digits 1-4 decline, while all other numerals are indeclinable.

	1	2	3	4	5	6	...
masc	einn	tvo	þrjá	fjóra	fimm	sex	
fem	eina	tvaer	þrjár	fjórar	fimm	sex	
neut	eitt	tvö	þrjú	fjögur	fimm	sex	

In contrast to the higher-level complex numerals, then, the morphosyntactic properties of lexical numerals may vary according to their numeric value.

The simplest explanation for this type of variation among lower value numerals is that they are subject to a lexical override specification, which causes them to be exempt from a more general rule of the grammar. The default rule of the grammar of Icelandic is that numerals do not inflect for case and gender with the head noun. The lexical entries for certain numerals in Icelandic might contain an override like the following:

override: lexical entries for {1, 2, 3, 4}: inflect for case, gender

Other examples of similar lexical subsets cross-linguistically include (Hurford forthcoming, Muravyova 1998:530):

Russian: 2-4 assign genitive singular case to noun, 5-10 assign genitive plural

Albanian: 1-4 separate numerals for definite/indefinite

Maltese: distinct ordinals only up to 4th

Chukchee: lexical numerals 1-5 only; 6+ formed as complex compounds.

All of these patterns can be explained by appealing to the presence of lexically specified override rules. However, if we make the crucial assumption that the lexicon consists of an *unordered* list, we would expect some languages to contain special subsets of lexical numerals that are not contiguous, for example:

(4) 2, 5, 9 decline, all others indeclinable

(5) 3, 6 and 8 assign genitive case, all others assign nominative case

In fact, as the examples above suggest, rules like (4) and (5) are unattested (Greville Corbett pers comm); special subsets of lexical numerals invariably form a sequence, and always at the low value end of the numeral set<sup>12</sup>. In other words, it is not a coincidence that numerals 1-4 only inflect in Icelandic, but not 3, 6, and 8; children do not seem to be able to acquire this type of rule for numerals. But if this is the case, how do they acquire these sets of numerals without “counting”? How do these special subclasses arise in a lexicon without reference to the linear order of elements in a numerical sequence?

Several explanations have been put forth to account for this cross-linguistic pattern. It is interesting to note that, numerals up to and including four seem to be of special significance. As mentioned above, Icelandic, Albanian and Russian all have special sets of numerals 1-4; languages grammaticalise number up to trial or perhaps quadral (four), but no attested language grammaticalises numerals above this (Corbett to appear). In an overwhelming majority of cases within the European languages, any special sequence of numerals in a given language will not exceed four (Hurford forthcoming). In acquisition, children learn numeral words for the lowest values first

<sup>12</sup> With the possible exception of Modern Greek, in which numerals 1, 3 and 4, but not 2, are morphologically marked for gender (Hurford forthcoming:31).

(1, 2, and 3) (Bloom 2000). Bloom's and Hurford's explanation of the special status of the lower numerals involves frequency effects: numerals up to three or four occur with much higher frequency than other numerals, so there is pressure on languages to grammaticalise them as a separate lexical subclass. This explanation does not challenge the basic principle of "no counting" in the grammar or lexicon: it is higher frequency, rather than the numeric value of the numeral per se, which allows these numerals to display morphological or syntactic properties distinct from other numerals. Bloom (2000) also argues that it is possible for very young children to possess basic number awareness for very small sets of objects without knowing how to count. For higher numerals, however, data from language corpora challenge the frequency explanation. Although it is true that numerals 1-3 do seem to be statistically more frequent than other numerals, there is not an inverse correlation between the value of a numeral and its frequency among those of higher value. A Spanish frequency dictionary Juillard & Chang-Rodriguez (1964) lists the numeral *cinco* (five) as more frequent than *cuatro* (four), and *ocho* (eight) as more frequent than *siete* (seven). Frequency effects, therefore, may explain how children acquire special subsets of numerals to three or four, but do not provide a robust explanation for the unusual behavior of Inari Sami numerals up to the value six.

For languages which contain special subsets for 1-5, another explanation is available. In Chukchee, for example, lexical numerals only go up to five, with higher numerals being formed as complex compounds. This system may be based on finger-counting (Muravyova 1998:530), which provides an anatomical reason for a special subset of numerals 1-5 in certain languages. Again, however, this explanation cannot extend to Inari Sami, since very few people have six fingers on each hand.

We would like to argue that the direct case assignment feature for numerals 2-6 in Inari Sami is acquired and implemented in a way that is sensitive to the numerical value of the numeral in a sequence. In other words, the special subset of numerals in Inari Sami, and perhaps in some of the other languages mentioned, must be listed in the lexicon in a particular order. A child acquiring Inari Sami, then, must be able to count to six and seven in order to acquire and use the case assigning rule.

This hypothesis is supported by data from acquisition. Bloom (2000, chapter 9) notes that very young babies seem to have a prelinguistic awareness of numerosity: in experiments, seven-month-olds are sensitive to the difference between two and three objects, most two-year-olds can count to three. At the single-word stage, experiments show that English-speaking children are aware of what number words are (ie that they are sensitive to the syntactic context of quantification), and that they refer to numerals greater than one. But it takes much longer, up to a year, for children to develop the ability to count using number words, and therefore to understand that number words refer to sets of objects of a particular numerical value. The fact that children go through a stage where they understand the concept of number, and they understand that words may refer to numbers, but they cannot attach number words to numerical values suggests that the ability to count is not an innate *linguistic* ability. However, Bloom notes that once they have developed the relevant cognitive skills to link number words up to three to particular sets of objects, children rapidly acquire higher numerals four, five, six, seven and so on, and are able to count objects using these words. This shows that higher value lexical numerals such as six and seven are acquired by mapping words onto numerical sequences; children learn these words *as they learn to count*. If this is the case, then children must be acquiring and storing lexical items in a particular linear order, along with morphosyntactic features which are sensitive to this linear order. For the acquisition of numerals, then, counting is

both possible and necessary. The data from Inari Sami provides evidence that “counting” in the lexicon during acquisition makes it possible to acquire “counting” in the grammar.

## 5 Conclusion

Numerals are interesting for linguistic theory in many ways. For one thing, they seem to vary in syntactic category status cross-linguistically: in some languages they are determiners, in others adjectives, and in others they appear to form a category of their own. Another interesting characteristic is that they differ from other non-predicates in that they are able to assign case to the nominal they modify. Furthermore, the formation of numerals seems to be governed by syntactic principles (Hurford 1975). Given the fact that numerals have so many interesting properties, it is surprising that they have been largely ignored in the linguistics literature.

In this paper, we have discussed the interaction of numerals and nouns in Inari Sami. New data was presented which showed that Inari Sami has a unique system of numerals, where a “split” occurs between numerals 6 and 7 with respect to the type of case they assign to the nouns they quantify. As we have seen, this raises important questions about the difference between lexical and structural case. The case marking system of Inari Sami numerals can be understood if we assume that the numerals 2-6 assign a special kind of accusative case, direct case, which accounts for the patterns of interaction between numerals, nouns, and locative cases assigned by a governing verb.

One further issue discussed in this paper was the fact that the split between 6 and 7 makes it seem as if though there is indeed counting in grammars, at least in some sense. We assume that nominals following numbers 7 and higher are assigned partitive by default, and the numbers 2-6 override this case marking lexically. However, this hypothesis leaves unexplained why the numerals with lexically specified case marking form a sequence of consecutive numbers. We argue that this can be understood if we take acquisition into account. According to recent research on language acquisition, numbers are acquired in an ordered sequence. This does not force us to assume that the numbers are *stored* in a particular order, but the mere fact that they are *acquired* in a particular order helps explain the peculiar facts about case marking in Inari Sami.

## 6 References

- Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge: MIT Press.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Corbett, G. (to appear). Number. In Martin Haspelmath, Ekkehard König, Wulf Oestereicher and Wolfgang Raible (eds) *Language Typology and Linguistic Universals: An International Handbook*. Berlin: de Gruyter.
- Halpern, A. (1995). *On the Placement and Morphology of Clitics*. Stanford, CA: CSLI Publications.
- Halpern, A. and A. Zwicky (1996). *Approaching Second: Second Position Clitics and Related Phenomena*. Stanford, CA: CSLI Publications.
- Hurford, J. (1975). *The Linguistic Theory of Numerals*. Cambridge: CUP.
- Hurford, J. (1987). *Language and Number*. Oxford: Basil Blackwell.

- Hurford, J. (forthcoming). The interaction between numerals and nouns. In F. Planck (ed) *Noun Phrase Structure in the Languages of Europe*. European Science Foundation, EUROTYPE project.
- Juillard A. G. & E. Chang-Rodriguez (1964). *Frequency Dictionary of Spanish Words*. The Hague: Mouton.
- Larson, R. (1988). On the double object construction. *Linguistic Inquiry* **19**. 335-392.
- Muravyova, I. A. (1998). Chukchee. In A. Spencer & A. Zwicky (eds) *The Handbook of Morphology*, Oxford: Blackwell, pp. 521-538.
- Nelson, D. (1998). *Grammatical Case Assignment in Finnish*. New York: Garland.
- Ojeda, Almerindo (1997). How to define number categories: A lesson from Semitic. In Kora Singer, Randall Eggert, and Gregory Anderson (eds.), *Proceedings of CLS* 33.
- Pollard, C. and I. Sag (1987). *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. CSLI Lecture Notes no. 13. Stanford, CA: CSLI Publications.
- Sammallahti, P. (1998) *The Saami Languages: An Introduction*. Karasjok, Norway: Davvi Girji OS.
- Sammallahti, P. & Morottaja, M. (1993) *Säämi-Suomâ Sänikirje / Inarinsaamelais-suomalainen sanakirja* [Inari Sami-Finnish Dictionary]. Ohcejohka: Girjegiisá Oy.
- Williams, E. (1994) Remarks on lexical knowledge. In L. Gleitman & B. Landau (eds) *The Acquisition of the Lexicon*. MIT Press, pp. 7-34.
- Vainikka, A. (1989). Deriving Syntactic Representations in Finnish. PhD Dissertation, UMass Amherst.
- Vainikka, Anne and Joan Maling (1996) Is partitive case inherent or structural? In J. Hoeksma (ed) *Partitives. Studies on the Syntax and Semantics of Partitive and Related Constructions*. Berlin/New York: Mouton de Gruyter, pp. 179-208.
- Woolford, Ellen (2000) Deriving (the effects of) Burzio's Generalization. Paper presented at the Case Theory Workshop, DGfS, Marburg March 1-3 2000.
- Vogel, Ralf (2000) Surface matters: the typology of case conflicts in free relative constructions and its implications for Case Theory. Paper presented at the Case Theory Workshop, DGfS, Marburg March 1-3 2000.
- Yip, M., Maling, J. & Jackendoff, R. (1987). Case in tiers. *Language* **63**. 217-50.

Diane Nelson  
Dept of Linguistics & Phonetics  
University of Leeds  
Leeds LS2 9JT  
GREAT BRITAIN

d.c.nelson@leeds.ac.uk

Ida Toivonen  
Dept of Linguistics  
Stanford University  
Stanford, CA 94305-2150  
USA

toivonen@csli.stanford.edu

