

# Natural language and set-theoretic conception of natural number

**Akira Watanabe**

University of Tokyo  
akirawat@l.u-tokyo.ac.jp

**Abstract:** This article presents a new theory of simplex numerals that incorporates a slight revision of Chomsky's (2008) set-theoretic conception of natural number, which assumes that the notion of natural number is innate. The new theory makes it possible to account for the behavior of numerals in counting as well as the developmental stages that children go through in learning numerals. The key idea is that set-theoretic objects corresponding to natural number notions are subject to operations that apply when a syntactic object is converted to phonological form. These operations provide a crucial link that connects the meaning of a numeral with the count list consisting of numerals. A notable feature of the proposed analysis is that the Cardinal Principle is derived by recruiting linguistic computation and therefore is no longer stipulated as such.

**Keywords:** counting; language development; linearization; natural number; numerals

## 1. Introduction

If we take seriously Kronecker's dictum, quoted in Bell (1937), that God made the integers, the notion of natural number must be innate. The dominant view in developmental psychology as reviewed by Rips et al. (2008), however, is that the natural number system is constructed out of other number-related concepts in the course of children's cognitive development (see Carey 2009; Feigenson et al. 2004; Le Corre et al. 2006; Piantadosi et al. 2012; Sarnecka to appear; and Wynn 1992, among many others). The major difficulty for the nativist position is the fact that acquisition of number words (or numerals) by children takes a long time, completed only at the age of three and a half to four years. Experimental results documenting this protracted learning process have been replicated since Wynn's (1992) longitudinal study (Barner et al. 2009; Le Corre & Carey 2007; Le Corre et al. 2006; Sarnecka et al. 2007). Gelman and Gallistel (1978) propose in their seminal work on children's numerical abilities that various innate principles underlie our ability to count. Particularly important among these is the Cardinal Principle, which states that the last nu-

meral used in counting represents the cardinality of the set counted. It has been shown that children's successful application of the Cardinal Principle is also delayed, more or less coinciding with the learning of the meaning of numerals in the count list (Fuson 1988; Le Corre et al. 2006; Wynn 1992). In addition to developmental evidence, it has also been reported that there are languages that do not express natural number concepts via numerals (Frank et al. 2008).

At the same time, the past literature has not gone beyond the talk of the cardinality of a set, leaving unexplored the precise relation between experimental findings and the set-theoretic representation of natural number, which is standard in mathematics. In a recent review by Sarnecka (to appear), for example, no mention is made of the characterization of natural number in terms of sets. This is a serious gap if the official theory of natural number is stated in terms of sets. In this article, I would like to bridge this gap by approaching the problems of developmental delay and typologically uneven distribution from a completely new perspective of set-theoretic conception of natural number. It is well known that Zermelo and von Neumann associate each natural number with a particular set, defining the successor function in a set-theoretic fashion. Benacerraf (1965) claims that there is nothing to choose between Zermelo's version and von Neumann's, but I will show that the way natural numbers are associated with particular sets has nontrivial consequences for helping us understand the relation between natural number concepts and their linguistic expression. In fact, it will turn out that neither Zermelo's system nor von Neumann's is of use for this purpose. A recent proposal by Chomsky (2008), on the other hand, can provide refreshing insights into mundane properties of numerals that are simply stipulated in the past, if properly modified. Significant aspects of numeral acquisition by children will also be shown to fall into place under a modified version of Chomsky's proposal.

In the following section, a revision of Chomsky's proposal will be presented and compared with Zermelo's and von Neumann's systems. This section will also propose a theory of simplex numerals, arguing that the count list, the Cardinal Principle of Gelman & Gallistel (1978), and the numerical notation for writing can be grounded on more fundamental elements of natural language. Section 3 will then take up the question of how children proceed in learning numerals, which poses the biggest obstacle to the nativist view on natural number.

## 2. A new proposal

In order to elucidate the relation between number and natural language, I would like to put forward a novel theory concerning the lexical entry of numerals, which encode individual natural number concepts. Set-theoretic objects that correspond to these notions occupy a central place in such a theory and are subject to operations in linguistic computation to derive phonological representations of numerals in a non-trivial fashion.

### 2.1. The connection between natural language and natural number

Chomsky (2008) suggests that discrete infinity of the natural number system can be given by the syntactic operation of Merge, which puts together two items and forms a two-member set, as in (1).

- (1) Merge  $(X, Y) = \{X, Y\}$

More concretely, he proposes that application of Merge to a single lexical item will give the following series, taking the lexical item in question to be *one*:

- (2) Chomsky's (2008) version of natural numbers  
 $1 = \text{one}, 2 = \{\text{one}\}, 3 = \{\text{one}, \{\text{one}\}\}, 4 = \{\text{one}, \{\text{one}, \{\text{one}\}\}\}, \text{etc.}$

The systematic one-to-one association of set-theoretic syntactic objects with natural numbers is thus established. Note that Merge (one, one) results in  $\{\text{one}\}$ , given the axiom of extensionality. The successor function takes the form of  $n + 1 = \{\text{one}, n\}$ .

My proposal is to shift the series by one, with the following association:

- (3) A revised natural number series  
 $1 = \{\text{one}\}, 2 = \{\text{one}, \{\text{one}\}\}, 3 = \{\text{one}, \{\text{one}, \{\text{one}\}\}\}, \text{etc.}$

Since the single lexical item does not have to be *one*, a more general format, which will be called the UG version, is the following:

- (4) UG version  
 $1 = \{i\}, 2 = \{i, \{i\}\}, 3 = \{i, \{i, \{i\}\}\}, \text{etc.}$

Here,  $i$  is used as a variable that can be identified with any arbitrary lexical item.<sup>1</sup>

An immediate advantage of this revision is that we can accommodate zero, if we replace  $i$  with 0 and take the series to start with 0 (though 0 is not a lexical item but a number concept, or an empty set, technically speaking).<sup>2</sup> Notice that Chomsky's original formulation, which starts with a lexical item (*one*), has no room for zero. The set-forming successor function cannot map anything to the lexical item in question. At the same time, the notion of zero is something to be discovered, as can be seen from the history of mathematics (Menninger 1969). Thus, incorporation of zero is a non-trivial task, depending on the correct choice of the value of  $i$ . See Nieder (2016) for a review of recent research on the cognition of zero.

A further advantage of the new formulation over Chomsky's original proposal is that the set-theoretic representation of  $n$  contains  $n$  instances of  $i$  in the new formulation. This point becomes highly significant when we consider numerals. Since numerals are lexical items, they are expected to carry semantic and phonological information. For their meaning, set-theoretic objects constructed by Merge should suffice.<sup>3</sup> Their phonological information, on the other hand, needs to be treated carefully, for there are at least two distinct ways in which numerals are used, combined in an act of counting. Consider Gelman and Gallistel's (1978) Cardinal Principle in this connection.

- (5) Cardinal Principle (Gelman and Gallistel 1978, 80)

The tag applied to the final item in the set represents the number of items in the set.

When we count, we say things like:

<sup>1</sup> It is extremely interesting to note that Gelman and Gallistel (1978, 131–133) reported that one two-year-old child in their studies used in counting a string of a single numeral, which happens to be *three*, instead of the conventional count sequence, as in “three, three, three”.

<sup>2</sup> See below on the use of the empty set in Zermelo's and von Neumann's systems. The version of (4) with  $i$  replaced by the empty set will be compared with Zermelo's and von Neumann's systems. It may be worth mentioning here that the empty set plays a fundamental role in axiomatic set theory. See George & Velleman (2002) for discussion.

Throughout this article, Arabic numerals are used to denote number concepts. Keep in mind that they are not natural language expressions.

<sup>3</sup> I assume that numerals receive an exact interpretation. For supporting arguments, see Huang et al. (2013); Kennedy (2013); Lipton & Spelke (2006). See also É. Kiss (2010) for evidence from Hungarian that the type of focus imposing exhaustivity helps bring out the exact reading of numerals.

- (6) One, two, three, four, five, five apples.

The Cardinal Principle is intended to capture the fact that the last numeral recited in counting expresses the cardinality of the set counted. I would like to suggest that the relation between cardinality and the final tag in counting is actually encoded as part of the phonological information in lexical entries. Furthermore, this phonological information is linked with the semantic representation in terms of particular sets via the process of linearization, which applies to syntactic objects created by Merge in the mapping from narrow syntax to PF in the case of ordinary linguistic expressions. Let me lay out the entirety of lexical entries for numerals first, using English for illustration,<sup>4</sup> and proceed step by step.

- (7) Lexical information encoded in numerals
- a. set-theoretic conception (meaning)  
 $1 = \{i\}, 2 = \{i, \{i\}\}, 3 = \{i, \{i, \{i\}\}\}$
  - b. actual counting sequence (phonology 1)  
 $1 = \text{one}, 2 = \text{one, two}, 3 = \text{one, two, three}$
  - c. actual phonological form (phonology 2)  
 $1 = \text{one}, 2 = \text{two}, 3 = \text{three}$

(7a) and (7c) need no further comments, except that orthographic spellings are used in place of actual phonological representations in (7c). What is novel is inclusion of (7b), also shown with orthographic spellings. The idea is that a numeral is provided with two distinct, but related phonological representations, namely, an initial segment of the count list and the last item of that segment. (7b) is used to express the result of counting without mentioning a noun that indicates what kind of objects are counted, while (7c) is the form chosen when a numeral is combined with a noun. It is quite unusual for a single lexical entry to have two pieces of phonological information. Take an ordinary noun *cat*. It is associated with just one phonological representation /kæt/. The plural suffix -s, which can be attached to it, is due to the number feature [–singular] independent of individual nouns. The phonological complexity of numerals is a crucial element in the account developed below.

A key step that connects the natural number concept to the phonological behavior of numerals is linearization of (7a), which gives (8).

<sup>4</sup> There may be additional syntactic information that needs to be listed, but that detail does not play a role in the discussion below.

## (8) Linearization of set-theoretic objects

$$1 = i, 2 = i\ i, 3 = i\ i\ i$$

Linearization is the process that applies to the syntactic object {from, him}, for example, to give *from him*. Since there is no precedence relation between members of a set, that relation must be added during the mapping to the articulatory-perceptual interface (so-called PF). One of the unusual properties of numerals is that linearization is applied to the semantic representation. The set-theoretic objects in (7a) will then be turned into sequences of *i*'s as in (8). Once (8) is given by linearization, a crucial next step is matching of (8) and (7b). The latter is available from the count list, which must be memorized. Cutting out a correct initial segment of the count list for each numeral is made possible by one-to-one correspondence with the relevant representation of (8).<sup>5</sup> (7b) is thus added to each lexical entry as phonological information in the course of lexical learning, a topic that will be discussed in detail in the next section.

The representations in (8) are not actual phonological forms, given that *i* is nothing other than a place-holder variable. One might wonder whether they are real. If we turn our eyes to number representations in modalities other than language, however, there are obvious connections. First, the sequence of *i*'s in (8) can be taken to be the basis for recording the number of things counted. Once *i* is replaced with a notch on a tally stick, we have a wide-spread primitive system for representing numbers (Menninger 1969; Overmann et al. 2011). It is quite likely that the first three Roman numerals developed out of the tallying system (Chrisomalis 2010; Dehaene 1997).<sup>6</sup> The forms in (8) may also be mapped to finger configurations that are used to indicate number (e.g., raising the index and middle fingers to mean 'two'). Gunderson et al. (2015) present evidence that children's grasp of finger gestures for number precedes learning of numerals. Their result can be taken to point to the role that the

<sup>5</sup> Since additive and multiplicative numerals are phrases (Hurford 1975), the connection to the count list must take a different route. See Hiraiwa (2016); Kayne (2010, ch. 3), and Watanabe (2010) for recent discussion of these complex numerals.

<sup>6</sup> The Roman numerals for 4 and above do not follow this pattern. The reason is probably related to a well-known property of precise representations of distinct individuals (Carey 2009; Feigenson et al. 2004), the numerical upper limit of which is 3 or 4. Even at a very intuitive level, images like IIII are not visually perspicuous, though IIII is a borderline case. See Chrisomalis (2010) for a recent comprehensive discussion of various numerical notation systems in the world. See also Menninger (1969) for the original suggestion that symbols like V and X are due to tally stick notches used for a grouping, based on evidence such as the Kharosthi writings of ca. 200 B.C.

representation in (8) plays in connecting a number concept to the corresponding count sequence in (7b). Computationally, (8) must be obtained before (7b). Hence the temporal order in acquisition. Let me also mention Hilbert's (1983) proposal that natural numbers should be understood as strings of vertical strokes along the lines of (8), in which case the successor function is concatenation of a single stroke. The Merge-based approach to natural number proposed in this article can be regarded as incorporating Hilbert's idea by means of linearization.

Continuing with explication of unusually complex lexical entries in (7), notice that the Cardinal Principle says that the form in (7c) comes at the very end of the sequence in (7b). I would like to propose that from 2 onwards, the relation between (7b) and (7c) should be captured in terms of the deletion operation, as in (9).

- (9) a. ~~one~~, two  
b. ~~one, two~~, three  
c. ~~one, two, three~~, four

The Cardinal Principle, therefore, should not be stipulated as such, because it is reducible to application of deletion to (7b).<sup>7</sup> The nature of this deletion operation will play a significant role in explaining a striking aspect of children's acquisition of numerals. I will come back to more details of it in section 3.

At this juncture, it should also be pointed out that there are cases in which the Cardinal Principle does not hold for paucal numerals. Hurford (2001) notes that some languages display discrepancies between numerals used together with a noun (called attributive numeral below) and numerals in the count list (called counting numeral), as in (10).

- (10) Distinct counting numerals (Hurford 2001)

	Number	Attributive numeral	Counting numeral
Chinese	2	liang	erh
German	1	ein(e) etc.	eins
German	2	zwei	zwo

This fact indicates that it should not be assumed that blind application of deletion to the sequence in (7b) derives the corresponding form in (7c),

<sup>7</sup> Kayne (2016) pursues a different approach to numerals, but does not address the nature of counting or the Cardinal Principle.

adding another reason why the Cardinal Principle should not be simply stipulated.

To summarize, numerals are special in being endowed with unusually rich phonological information as part of their lexical entries and in having this phonological content linked with their number concept in non-trivial ways. The connection between phonological content and semantics is made possible by a particular set-theoretic conception of natural number. Significantly, what is innate under the new proposal is not the Cardinal Principle but the system of a series of sets generated by the Merge operation. The Cardinal Principle itself is demoted to a relation holding between two pieces of phonological information encoded in the lexical entry of simplex numerals (up to the first numerical base), which is captured by means of deletion. In fact, it should be realized that the Cardinal Principle is nothing more than a descriptive statement about what people do when counting things on the basis of a sortal criterion. Taking the Cardinal Principle for granted, the past literature forgot to ask why counting actually works in the first place. We all know that the Cardinal Principle is correct, but calling an observational statement a principle does not turn it into a principle. In this sense, the past discussion of children's numerical cognition failed to be grounded on proper understanding of why it is the case that the  $n$ th item in the count list means  $n$ . Notice that this interpretation is not applicable in the case of the alphabet nor in the case of the sequence of ordinal numerals. It is significant that the set-theoretic characterization of natural number in (4) achieves the desired goal. As will be shown immediately in relation with alternative set-theoretic characterizations of natural number, this is a non-trivial result.

The exceptional phonological complexity of lexical entries for numerals probably warrants the view suggested by Frank et al. (2008) that they should be characterized as a cognitive technology.<sup>8</sup> To arrive at the form in (7c), it is necessary to apply linearization to (7a), match the result with the count list, and then apply deletion to the matched segment of the count list, leaving only the final item. Nothing of this sort is needed for ordinary lexical items, which enjoy complete Saussurian arbitrariness. Cultures that do not bother to go through the three steps in question are destined to lack a full-fledged numeral system. Considered in isolation, (7c) is a manifestation of Saussurian arbitrariness as in the case of ordinary lexical items, but crucially, the phonological information carried by individual digit numer-

<sup>8</sup> The title of their article is misleading, mentioning number as a technology, but the text discussion makes it clear that what is at stake is linguistic expressions, not number concepts.



als also includes (7b), an initial segment of the count sequence, which is related in a non-arbitrary way to the corresponding natural number concept in (7a) as well as to (7c) through the three operations mentioned. No other lexical items map their meaning to their phonological form in terms of a series of well-defined computational processes, in contrast to numerals. As we will see in section 3, learning complex lexical entries for numerals poses tremendous difficulties for children.

## 2.2. Previous set-theoretical characterizations of the natural number concept

Let us now compare the Merge-based conception of natural number with the well-known set-theoretical characterizations of the natural number concept put forth by mathematicians. The issue is how to relate natural language to natural number.

To take Zermelo's system given in (11) first, an important thing to notice is that there is no non-arbitrary way of accommodating the count list under this conception, even if some place holder is used instead of the empty set. Linearization will simply give a single symbol. The count list, therefore, must be related to individual numbers by a brute-force stipulation under Zermelo's system.

(11) Zermelo

- a.  $0 = \emptyset, 1 = \{\emptyset\}, 2 = \{\{\emptyset\}\}, 3 = \{\{\{\emptyset\}\}\}, \text{etc.}$
- b.  $n + 1 = \{n\}$

At the same time, it is possible to reformulate the successor function in (11b) as  $n + 1 = \text{Merge}(n, n)$ .<sup>9</sup>

Turning next to von Neumann's conception summarized in (12), it might not be inconceivable to associate an appropriate initial segment of the count list with each number (or set), since the cardinality of each set coincides with the number value represented, as can be seen rather clearly from (12a'), which paraphrases (12a) for  $n \geq 1$ .

(12) von Neumann

- a.  $0 = \emptyset, 1 = \{\emptyset\}, 2 = \{\emptyset, \{\emptyset\}\}, 3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \text{etc.}$
- a.'  $1 = \{0\}, 2 = \{0, 1\}, 3 = \{0, 1, 2\}, \text{etc.}$
- b.  $n + 1 = n \cup \{n\}$

<sup>9</sup> In syntactic computation, Merge does not put together a non-lexical object with itself, though.

On the other hand, there is no meaningful linguistic operation that corresponds to the successor function in (12b). So, the connection to natural language is quite thin, conceivable only for the relation to the count list. Even this correspondence is not straightforward, given that the set members of (12a') do not match the numerals in the count list. The ordered sequence of natural numbers  $0, \dots, n-1$ , resulting from linearizing the set for  $n$  in (12a'), must somehow be converted into  $1, \dots, n$ . In other words, the  $n$ th item in that sequence is  $n-1$ . There is thus no way of avoiding the direct stipulation that relates the cardinality of a set to the corresponding first segment of the count list (recall the status of the Cardinal Principle).

Thus, Zermelo's and von Neumann's set-theoretical characterizations of natural number do not fully capture their relation to natural language. They therefore fare poorly in this respect, compared with the new proposal outlined above. The difference in particular sets used matters, contrary to Benacerraf (1965), when we try to capture the relation between natural number and natural language.

It is also worth noting that the three series start out identically. For comparison, consider (4'), which is obtained by substituting the empty set for  $i$  in (4) and adding zero.

(4')  $0 = \emptyset, 1 = \{\emptyset\}, 2 = \{\emptyset, \{\emptyset\}\}, 3 = \{\emptyset, \{\emptyset, \{\emptyset\}\}\}, \text{etc.}$

0 and 1 are characterized in the same way as in (11) and (12). Furthermore, the difference between (4') and (12a) becomes visible only from 3 onward. It is therefore impossible, when a sequence of sets is given, to uniquely identify the successor function lying behind it without examining the first four sets. We will encounter a similar situation in the account of children's acquisition of numerals to be presented below.

### 3. Aspects of numeral acquisition explained

We are now ready to consider the question of why the acquisition of numerals by children is a prolonged process. Let us start by reviewing the well-established developmental course that children take in acquiring numerals and discuss how they are to be explained.

#### 3.1. Developmental stages

The developmental stages that children go through when learning numerals are well documented. (13) summarizes the major sequence of events.

## (13) Stages of numeral acquisition

- a. learning the count list (by about  $2\frac{1}{2}$  years)
- b. assigning the meaning to *one* through *three* (or *four*)
- c. learning the meanings of all the larger numerals and also the meaning of counting (by about  $3\frac{1}{2}$  years)

By the age of two and a half, children start to produce a sequence of digit numerals, though the list may contain errors (Fuson 1988; Pollmann 2003; Wynn 1992). At this stage, however, they do not necessarily know the meaning of the numerals that they can recite as a sequence. This fact is revealed when the following set of experiments are conducted:

## (14) Experimental paradigms used to measure acquisition of numerals

- a. Give-a-Number: "Can you give me three frogs?"
- b. Point-to-X: "Can you show me the three frogs?"
- c. What's-on-This-Card: "What's on this card?"

Notice that the experiments in (14) all involve natural language expressions combining a numeral and a noun. The instructions for (14a,b), given above, directly employ such a combination. Children are asked to carry out the instruction in the Give-a-Number test and to point to a matching picture in the Point-to-X test. For (14c), the expected answer is something like "three frogs". The result, therefore, is dependent on whether children understand and are able to utter linguistic expressions that put together a numeral and a noun. Thus, even if the natural number is innate, children fail in these tasks when they do not understand the meaning of the relevant numeral in combination with a noun. Mastering the meaning of numerals and the syntax of numeral-noun combinations is the minimal requirement for adult-like performance.<sup>10</sup>

<sup>10</sup> An anonymous reviewer objects that Almoammer et al.'s (2013) Give-a-Number test omits the noun, claiming that the presence of a noun is not central to this task. This reviewer fails to understand that a sortal notion is indispensable for counting. Not surprisingly, Almoammer et al.'s (2013) What's-on-This-Card test prompts children to supply a noun. The standard linguistic analysis (Lobeck 1995) of cases like *Can you give me two?* is that the head noun is structurally present after *two* but remains unpronounced due to the process of ellipsis. The point can be seen clearly from Le Corre et al.'s (2016) phase 2 version of Give-a-Number test, where an antecedent clause with the noun in question is presented prior to the instruction, making sure that children understand what objects to count. In both of these Give-a-Number experiments, the noun is omitted to suppress the number information carried by nouns. For the record, let me note that the Give-a-Number tests in Barner et al.

It has been shown that these tests identify children's knower level (Le Corre et al. 2006; Wynn 1992). Children first learn the meaning of *one*, becoming one-knowers. They then learn the meaning of *two* (becoming two-knowers), and then that of *three* (becoming three-knowers). After this stage, many children come to understand the meaning of all the other digit numerals in their count list as well as the Cardinal Principle (CP), becoming CP-knowers. A minority of children go through the stage of four-knowers before becoming CP-knowers (Le Corre & Carey 2007). Those who have not become CP-knowers are called subset-knowers. Children's understanding of counting or the Cardinal Principle is tested with the Counting Puppet task, where children are asked to judge whether a verbal instruction is correctly executed by the puppet, who puts a certain number of toys in a can by slowly counting them one at a time. Le Corre et al. (2006) have demonstrated that the performance on this test is more or less correlated with that on the Give-a-Number test for a higher number above three or four, indicating that children come to understand the meaning of counting after learning the first three or four numerals in the count list.<sup>11</sup> It takes about one year to become a CP-knower after becoming a one-knower.

### 3.2. Aspects of development explained

We are now faced with the task of explaining why acquisition of numerals by children proceeds as it does. The challenge is not limited to providing an account of why it takes so long, completed roughly at the age of three and a half. Two additional important characteristics as summarized in (15) also need to be explained.

---

(2009); Le Corre et al. (2006); Le Corre & Carey (2007), phase 1 of Le Corre et al. (2016), and Sarnecka et al. (2007) all use numeral-noun combinations. Wynn (1992) does not supply the needed information, but Wynn (1990) explicitly mentions the use of numeral-noun combinations.

<sup>11</sup> An anonymous reviewer wonders why inclusion of the representation in (7b) as part of the lexical entry does not help counting. The answer is that the Counting Puppet task is designed to test cardinalities beyond the range of subset-knowers (6, 7, and 8 in the case of Le Corre et al. 2006). By definition, subset-knowers have not learned the numerals that correspond to 6, 7, and 8, for example. For small numbers, on the other hand, counting is usually not needed, because of precise representations of distinct individuals, mentioned in footnote 6.

- (15) Significant properties of numeral acquisition
  - a. stepwise development
  - b. the role of “three” or “four”

Section 3.1 has reviewed the results of studies concentrating on English, but the same sequential acquisition of numerals is observed for Japanese, Russian (Sarnecka et al. 2007), Slovenian, Soudi Arabic (Almoammer et al. 2013), and Chinese (Le Corre et al. 2016) as well as for the language of an indigenous farming-foraging group in Bolivia called Tsimane’ (Piantadosi et al. 2014). So this is a cross-linguistically persistent pattern, unaffected by cultural differences. The cut-off point that separates CP-knowers from the rest also seems to be cross-linguistically stable, marked by the first three or four numerals in the count list (Barner et al. 2009; Le Corre et al. 2016; Piantadosi et al. 2014). Let us consider what the proposal presented in section 2.1 has to say about these aspects of numeral acquisition.

The general delay in learning numerals is given a very simple answer: the unusual complexity of the lexical entry for numerals. Recall that two distinct, but related phonological representations are posited for each numeral (except for 1). In addition, (8) is used as a ladder with which to reach (7b). The relation between (7b) and (7c), characterized in terms of deletion, also poses a nontrivial obstacle, as will be shown shortly. Both (7b) and (7c), therefore, need to be memorized as such at the beginning when children start learning numerals. This in turn makes possible the discrepancy, exemplified in (10), between numerals in the count list and those used together with a noun. Blind application of the Cardinal Principle here gives a wrong result.

There is indeed evidence that the complexity of morpho-phonological realization delays lexical learning. Sarnecka et al. (2007) observe that Japanese children lag behind compared with English and Russian learners at an early stage of numeral acquisition, though Barner et al. (2009) have shown that they soon catch up. As a possible factor that causes the initial delay, Barner et al. point to the morphological complexity of numerals due to the presence of a classifier. They simply say that numerals can take on many distinct forms, but a more accurate statement is that Japanese has two completely distinct series of digit numerals, Sino-Japanese and native, and that the shape of the Sino-Japanese series undergoes phonological modification in relation to the immediately following classifier. Furthermore, one major classifier for counting humans mixes the two series. For example, the count list, given in (16a), is made up of the Sino-Japanese series, which are also used with the semi-default classifier *-ko* for inanimate

entities in (16d).<sup>12</sup> Note the gemination in *ik-ko* and *rok-ko*. The native series, together with the default classifier *-tsu* for inanimates, are illustrated in (16b). The human classifier in (16c) starts out with the native series but switches to the Sino-Japanese series at 3, itself changing from *-ri* to *-nin*, too (Watanabe 2010).

(16) Multiple ways of counting in Japanese

- a. *ichi, ni, san, shi, go, roku, etc.* (count list)
- b. *hito-tsu, futa-tsu, mit-tsu, yot-tsu, itsu-tsu, mut-tsu, etc.*  
(counting inanimates with a default classifier)
- c. *hito-ri, futa-ri, san-nin, yo-nin, go-nin, roku-nin, etc.*  
(counting humans with a classifier)
- d. *ik-ko, ni-ko, san-ko, yon-ko, go-ko, rok-ko, etc.*  
(counting inanimates with a semi-default classifier)

Children learning Japanese encounter these bewildering details and have to sort them out. See Okamoto (2015) in this connection. Classifiers are usually used together with a noun, but the numeral-classifier combination can also be used to count entities of an appropriate kind. So there are multiple ways of counting in Japanese in addition to the count list in (16a). It is no wonder that Japanese children are slow at the beginning.<sup>13</sup> Morphological errors may persist after children master the meaning of the numeral correctly. Yamamoto (2005) reports that there are cases where even 6-year-olds produce errors like *\*ichi-nin* and *\*ni-nin* for counting humans.

It is also worth stressing that the presence of two distinct series in the Japanese numeral system makes it reasonable to provide multiple phonological forms for a single lexical entry in general. The Sino-Japanese and

<sup>12</sup> My dialect uses *shi* for 4 in the count list, but there are dialects that recruit *yon* from the native series. The latter, with morpho-phonological variation, is substituted even for classifiers that otherwise select the Sino-Japanese series in all dialects, as in (16d). A distinct numeral for the count list is a phenomenon illustrated in (10) above.

It should be noted in this connection that three digit numerals, *ichi* 'one', *shichi* 'seven', and *hachi* 'eight' end with [tʃi], which is exceptional for disyllabic Sino-Japanese morphemes. See Ito & Mester (2015) and the references cited there. The high concentration of phonologically exceptional morphemes points to the special status of numerals in the lexicon.

<sup>13</sup> Interestingly, according to Yamamoto's (2005) literature survey, the acquisition of the count list (16a) itself does not seem to be delayed. Sarnecka et al. (2007) record no significant difference in counting scores between English and Japanese when children are prompted to count the objects presented by reciting the count list.

native series are completely unrelated morphologically, requiring two forms that correspond to (7c).

Let me hasten to add that the morphological complexity of the numeral-classifier combination is not necessarily the whole story about Japanese children's initial delay, however. Whether or not a numeral is used together with a classifier is a major parametric option that needs to be selected during the course of language development. This choice must be made before children start to learn individual classifiers. And indeed the well-known overgeneralization of the default classifiers *-tsu/-ko* (16b,d) in children's production at an early stage suggests that the decision to use classifiers is separate from acquisition of individual classifiers, as Yamamoto (2005) argues. See also Barner et al. (2009). It is quite likely that it takes some extra time to realize that Japanese is a classifier language, since children need to notice that the morpho-syntax of numerals is accompanied by something additional. If so, it will delay the onset of numeral acquisition.<sup>14</sup>

Returning now to an account of various aspects of numeral acquisition, note that it becomes possible to explain the actual steps that children take in learning numerals when a counting sequence for each number is incorporated as part of the phonological information of the lexical entry as in (7b). First of all, the numeral for 1 does not distinguish between

<sup>14</sup> This story will apply to classifier languages in general. In their study on Chinese children, Le Corre et al. (2016) point to the absence of number morphology on nouns as a key element that delays the learning of the numeral for 1, though they leave room for additional factors in the case of Japanese. Instead, I suspect that the use of a classifier itself may be the cause of the delay. Here is the reason.

One complicating factor in Chinese is the possibility of using a classifier without an overt numeral, as in (i).

(i) Wo xiang mai ben shu.

I want buy CL book

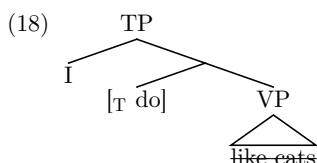
'I want to buy a book.'

Cheng and Sybesma (2005) note that this construction signals singularity. See also Zhang (2014) for further discussion. In addition, Zhang argues that reduplication of classifiers expresses plurality. Thus, Chinese does not lack the grammatical singular/plural distinction, contrary to Le Corre et al.'s factual assumption. It is not clear why the morphological number marking on nouns, but not a grammatical singular/plural distinction elsewhere, facilitates the acquisition of numerals. Incidentally, neither of these complications arises in Japanese. See Watanabe (2014) on the impossibility of using classifiers without an accompanying numeral in Japanese.

(7b) and (7c) and is the simplest item to learn in this respect.<sup>15</sup> The form *one*, *two* in (7b) is minimally distinct from *one*, and therefore is the next item to learn together with the form *two* in (7c). And then comes *one*, *two*, *three*, with *three* added to *one*, *two*, which is already in place. Thus, a nontrivial relation in phonological form (7b) between two adjacent numerals leads children from the simplest item to higher numerals one by one. In other words, the stepwise developmental course is attributable to the phonological complexity of each numeral, once a count sequence itself is associated with each numeral as part of the lexical entry.

After children learn the numeral for 3, it becomes possible for them to figure out the relation between (7b) and (7c), described by the Cardinal Principle, which I suggested should be captured in terms of deletion in section 2.1. To elaborate on this proposal, let me explain what is meant by deletion more precisely. Recall that the new theory of numerals derives (8) by applying linearization to (7a), treating the natural number concept in (7a) as if it were a syntactic object, on the grounds that the Merge operation provides the basis for discrete infinity of natural numbers. Now, constituent deletion is another operation that applies during the course of mapping syntactic objects to phonological form. This deletion operation is considered to lie behind ellipsis phenomena in recent analysis stemming from Chomsky & Lasnik (1993) and Merchant (2001). According to the deletion analysis of ellipsis, the answer in (17) involves the structure in (18), with the phonological content of VP deleted.

- (17) Q: Do you like cats?  
A: Yes, I do.



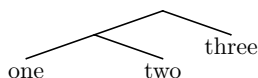
<sup>15</sup> It is an interesting topic for future research to find out whether the use of distinct forms for (7b) and (7c) in languages like German as in (10) poses a challenge for children. For Chinese, experimental results in Le Corre et al. (2016) do not seem to indicate a delay for 2.



acquisition of numerals presents evidence that deletion applies to linearized constituent structure.

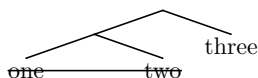
Suppose that the count sequence in (7b) is associated with syntactic constituency, as in (19).

(19)



Conventional tree-like structures like (19) encode the precedence relation as well as constituency. Both of these two pieces of information are needed to relate (7b) to (7c) by means of constituent deletion. Notice that what we want is deletion of the highest constituent within a linearized tree-like structure (or the non-head member of the root node, as an anonymous reviewer suggests). Applying this deletion operation to (19) produces (20).

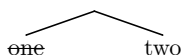
(20)



Recoverability of deletion requires the presence of an antecedent in ordinary cases of ellipsis. In the case of numerals, recoverability is guaranteed by the existence of a count list. This implies that an act of counting with observable objects is an important discourse context that facilitates children's acquisition of numerals. See Gunderson & Levine (2011) for evidence that points in that direction.

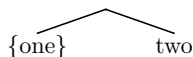
What happens in the case of *two*, as shown in (21)?

(21)



From the learner's viewpoint, it is impossible to tell from the relation between *two* and the sequence *one, two* what operation is actually involved. In the case of (21), it can be deletion of the leftmost item. Even if one recalls that the left branch can be rendered as (21'), the target of deletion can be either the singleton set {~~one~~} or just the member of that singleton set, namely, ~~one~~.

(21')



Notice also that hitting upon the idea of deleting  $\{one\}$  does not necessarily lead children to hypothesize deletion of the highest constituent within a linearized tree-like structure. If the generalization that they reach is that the singleton set in the structure is to be deleted, it will give a wrong result for 3 and above. In other words, observing the relation between *two* and the sequence *one, two* does not guarantee the correct choice on the part of the learner. There is no such ambiguity when both (20) and (21) are taken into account, on the other hand. This difference constitutes a significant dividing line. There is in fact evidence that children who are not yet CP-knowers rely on rote learning. Recall that numerals in the count list can be different from those used with a noun, as exemplified in (10). Hurford (2001) observes that this discrepancy is most commonly found with 2 and rare otherwise. The only other example given is *eins* in German. The forms for (7b) and (7c) thus need to be memorized as such until children become CP-knowers.

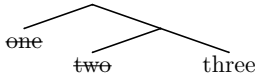
To sum up the discussion so far, the Cardinal Principle, which is just a description of what people do in counting, is explained by means of deletion of the highest constituent within a linearized tree-like structure under the new set-theoretic conception of natural number. The discovery of this deletion mechanism behind the Cardinal Principle by children is delayed until the relevant structure becomes sufficiently complex, in other words, until they learn the lexical entry for the numeral that corresponds to 3. Thus, the characterization of the Cardinal Principle in terms of constituent deletion provides a principled explanation for the fact that the grasp of counting and the meaning of digit numerals in general comes after children become three-knowers. And the existence of rare four-knowers (Le Corre & Carey 2007) simply points to the fact that some children are slower or more cautious than others, waiting to confirm the hypothesis with the proper use of *four*. The whole discussion provides very strong evidence that the Cardinal Principle should not be stipulated as such but needs to be reduced to more elementary items in linguistic computation.

An anonymous reviewer wonders whether the deletion operation in question is optional. The answer is yes, in every language that has a full-fledged system of numerals. When deletion does not apply, we get an initial segment of the count list. The choice is certainly keyed to contexts of use, which may be regarded as stipulative. Technology in general, however, is always like that, except that the question of which choice is functionally

more efficient is expected to play a role. In counting, the count list facilitates one-to-one correspondence with the objects counted. But when the result is put together with a noun, one is inclined to avoid clumsiness. Recall Frank et al.'s (2008) view mentioned above.

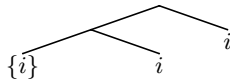
Now, notice that the above account presupposes the left-branching structure as in (19). A proper constituent deletion operation cannot be defined over the right-branching structure, as can be seen from (22).

(22)



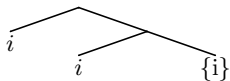
So, this is another detail that children have to figure out. What this amounts to is the relevance of the head parameter in linearization. To be mapped to (19), the set-theoretic object in (7a) must be linearized in such a way as to result in (23).

(23)



Recall that  $i$  is the place-holder for an arbitrary lexical item, which therefore can be regarded as the head of each constituent. Notice that (23) is an instance of head-final structure. Though headedness itself does not play a role in the semantic interpretation of (7a), logical possibilities in linearization necessarily leave room for a choice between (23) and (24).

(24)



Once the head-final option (23) is adopted for linearization of (7a), with the resulting structure linked to an initial segment of the count list, it becomes possible for the working of the Cardinal Principle to be defined as due to constituent deletion. Crucially, the count list presupposes the precedence relation. Without linearization, therefore, the set-theoretic representation in (7a) cannot be related to the count sequence. At the same time, constituency is needed for the proper functioning of deletion. Thus, we need a conventional tree-like structure after linearization, to which ellipsis applies.

It is worth pointing out that the linearized hierarchical structure of the (23) type provides a computational foundation for the count sequence itself, too. Notice that as the Merge operation is repeated to generate larger

natural number concepts, instances of *i* will be added on the right in the corresponding head-final structure. In other words, the growth of the head-final structure mimics the growth of the count sequence. This property is obviously linked to the working of the Cardinal Principle, which is now rendered in terms of deletion of the highest constituent. Thus, the set-theoretic objects in (4)/(7a) and the linearized hierarchical representations that they are mapped to are crucial elements in the cognitive technology that makes use of numerals.

I would also like to note in this connection that children's use of VP ellipsis is considered by developmental scientists to be delayed.<sup>16</sup> Thornton (2010) observes that children start to use VP ellipsis by about two-and-a-half years of age, attributing the delay to various conditions imposed on VP ellipsis, including peculiarities of the English auxiliary system. She bases her discussion on comparison with fragment answers, illustrated in (25).

- (25) Q: What can Daddy cook?  
A: Chili.

Merchant (2004) argues that answers like (25A) involve a full syntactic representation given in (26), where the focused phrase is moved out of TP, which in turn undergoes PF deletion.

- (26) [*chili*<sub>*j*</sub> [<sub>TP</sub> ~~Daddy can cook *t<sub>j</sub>*~~]]

Thornton points out that fragment answers appear earlier than VP ellipsis, by about age two or even earlier, and that the different time course of development is due to the fact that fragment answers are free from additional conditions that VP ellipsis needs to satisfy.

The acquisition of ellipsis shows two things. First and most importantly, the mechanism of constituent deletion itself is already in place when children need it to become CP-knowers. Notice that children start to use VP ellipsis, which is considered to be relatively late in coming, roughly at the same time as they start to recite the count list. They can therefore recruit constituent deletion when learning individual numerals. Second, the deletion mechanism cannot always be used to derive the adult form if it takes time to learn some additional properties of the relevant ellipsis structure. This reminds us of the truism that adult-like performance requires every relevant piece to be present.

<sup>16</sup> Thanks are due to Arhonto Terzi for directing me to questions about the acquisition of ellipsis.

To summarize, the protracted numeral acquisition by children is attributed to complex lexical entries. In addition, the Cardinal Principle is reduced to constituent deletion, which lies at the heart of the numeral system. Its acquisition is delayed till children become three-knowers because of the logic of the structural complexity that necessitates its employment. The proposed account of development claims that there is a qualitative difference between steps (13b) and (13c): the former involves rote learning, while the latter hinges on the realization that the representation in (7b), given by linearizing (7a) and matching the result with an initial segment of the count list, can be trimmed by constituent deletion to leave the last item as the representation in (7c).

It is interesting to end the discussion by noting that the proposed picture of numeral acquisition receives empirical support from the study on bilingual children by Wagner et al. (2015), which shows that one-knower, two-knower, and three-knower stages in one language are independent of the corresponding stages in the other, whereas there is a strong correlation between the two languages in the timing of becoming a CP-knower. Knowledge of lexical entries is not transferrable from one language to another, but the operation of constituent deletion can be shared. The account proposed in this article gives a concrete characterization of qualitatively distinct aspects of numeral acquisition.

#### 4. Conclusion

The constructivist view, dominant in the psychology literature, has not yet provided an account of how children come to internalize a set-theoretic characterization of natural number, be it Zermelo's, von Neumann's, or the one proposed here, as part of their unconscious knowledge. The nativist position obviates the need for such an account, but it must still specify what kind of cognitive structure is involved. To borrow the distinction drawn by Giaquinto (2015), the past discussion has been confined to the domain of finite cardinal number, failing to reach the realm of natural number. This article has put forth a new proposal concerning the relation between natural number and natural language, supporting the idea that discrete infinity of natural number arises from the innate combinatorial mechanism of language. Specifically, a slightly revised version of Chomsky's (2008) Merge-based conception of natural number is combined with the hypothesis that semantic and phonological representations in the lexical entry of numerals are related by operations such as linearization and constituent deletion that are employed in linguistic computation. More concretely, the

set-theoretic object representing a natural number concept is linearized first, and then the resulting hierarchical structure is matched with the corresponding initial segment of the count sequence. Finally, the linearized structure, enriched with the count sequence, undergoes constituent deletion, leaving the final element of that sequence. Thus, the meaning of a digit numeral is mapped to its two phonological representations through a series of well-defined computational operations. The Cardinal Principle describes the last step of this process. Counting is made possible by establishing the one-to-one correspondence between the counted objects and an initial segment of the count sequence incorporated as part of the linearized hierarchical structure. The whole story thus sheds new light on various aspects of numerals that are simply taken for granted in the past. The proposed theory of numerals also enables us to show that the innateness of natural number is compatible with the delayed acquisition of numerals. What it takes time to learn are complex lexical entries of numerals with multiple phonological representations. Furthermore, by the logic of structural complexity, children have to wait till they master the word for 3 to figure out that constituent deletion lies behind the Cardinal Principle. The significance of the word for 3 is given a straightforward structural account for the first time under the proposed system of numerals.

It is worth stressing that these results are made possible by linguistic manipulation of two-membered sets that are put in one-to-one correspondence with natural numbers. The proposal focuses on an obvious fact that the relation between natural number and language is mediated by the system of numerals, whose lexical entries hold the key. In order for linguistic computation to be able to handle set-theoretic objects for natural number concepts encoded in lexical entries of numerals, these objects themselves must be of the kind that allows such computation to function properly. It is in this respect that Chomsky's Merge-based conception of natural numbers turns out to be empirically superior to the alternatives by Zermelo and von Neumann.

I would also like to point out that the mediation of natural language expressions in psychology experiments intended to explore human concepts has been overlooked or treated very carelessly. The experiments used to measure children's knower-level, summarized in (14), for example, all make crucial use of verbal instructions involving numerals. If the delay in the acquisition of numerals is simply a matter of learning lexical items as suggested here, the view that advocates the non-innate nature of natural number loses its empirical basis. The same is true with a recent discussion of the successor function by Cheung et al. (2017), who claim that its

manifestation in children's behavior takes roughly two years after they become CP-knowers. The task used to assess children's understanding of the successor function is called the infinity task, which starts by asking children, "What is the biggest number you can think about?" After a couple of related questions, the experimenter goes on to ask children whether it is possible to keep adding one to the biggest number that they can name. If Chomsky's proposal or the modification of it in this article is on the right track, the successor function is  $n + 1 = \{n, i\}$ . It is a bit too naïve to think that translation of this function into plain English is straightforward for children. Notice, for example, that in this experiment, the adjective *big* is no longer limited to the physical size and is extended to the abstract domain of numbers. Children need to understand this extended sense of *big* to display an adult-like behavior. Children also have to figure out the correct abstract meaning of *adding one*, which is an application of Merge under our proposal. Given all these complications, a delay is expected. An important lesson from the discussion in this article is that an appropriate linguistic analysis of verbal instructions used in experiments is a prerequisite for evaluating the data obtained. It can hardly be said that this requirement has been met in the past discussion of natural number.

Lastly, let me turn to interesting by-products of this study. One is the demonstration that set-theoretic objects in narrow syntax are initially converted into conventional tree-like representations, which constituent deletion targets. Ordinary cases of ellipsis do not care about how linearization works at the ellipsis site. The operation of deletion behind the Cardinal Principle, however, is sensitive to the order of numerals in the count sequence. In order to obtain the correct result, linearized hierarchical structure is indispensable.

Note also that the proposed account has important implications for the head parameter. Kayne (1994) and subsequent work have pursued the idea that only head-initial structures are available for linguistic computation. The reduction of the Cardinal Principle to the operation of constituent deletion, however, requires the head-final linearization of set-theoretic structures that correspond to natural number concepts. We are led to conclude that head-final structures are possible in principle. There are indeed two logically possible ways of giving the precedence relation to the two members of each syntactic object formed by Merge. If one of them is to become unavailable, additional factors must be invoked to rule it out. Kayne (2013) points to the probe-goal relation in syntactic computation as such a factor. Being the purest instance of discrete infinity made possible by Merge, the proposed set-theoretic characterization of natural number

does not involve the probe-goal relation, nor any other complications. It is therefore not surprising to find head-final structures realized in the system of numerals.

### Acknowledgements

This article has a long gestation period, which more or less coincided with my father's struggle after he was struck by a stroke on May 20, 2008. The idea of combining linearization with modification of Chomsky's (2008) proposal came early and was presented at a workshop held at Meikai University in December 2009. It took me some more years, however, to come up with the use of deletion to derive the Cardinal Principle. Everything was put together in a presentation at the Advances in Biolinguistics workshop of the 19th International Congress of Linguists in Geneva in July 2013. I would like to thank the audiences on these occasions as well as anonymous reviewers for thought-provoking comments. The research reported here is supported by Grant-in-Aid for Scientific Research (C) 22520492 and (C) 15K02591 from the Japan Society for the Promotion of Science. This article is dedicated to the memory of my father, who passed away on December 5, 2016.

### References

- Almoammer, Alhanouf, Jessica Sullivan, Chirs Donlan, Franc Marušič, Rok Žaucer, Timothy O'Donnell and David Barner. 2013. Grammatical morphology as a source of early number word meanings. *Proceedings of the National Academy of Sciences* 110. 18448–18453.
- Barner, David, Amanda Libenson, Pierina Cheung and Mayu Takasaki. 2009. Cross-linguistic relations between quantifiers and numerals in language acquisition: Evidence from Japanese. *Journal of Experimental Child Psychology* 103. 421–440.
- Bell, Eric Temple. 1937. *Men of mathematics*. New York: Simon and Schuster.
- Benacerraf, Paul. 1965. What numbers could not be. *Philosophical Review* 74. 47–73.
- Carey, S. 2009. *The origin of concepts*. Oxford: Oxford University Press.
- Cheng, Lisa and Rint Sybesma. 2005. Classifiers in four varieties of Chinese. In G. Cinque and R. S. Kayne (eds.) *The Oxford handbook of comparative syntax*. Oxford: Oxford University Press. 259–292.
- Cheung, Pierina, Miriam Rubenson and David Barner. 2017. To infinity and beyond: Children generalize the successor function to all possible numbers years after learning to count. *Cognitive Psychology* 92. 22–36.
- Chomsky, Noam. 2008. On phases. In R. Freidin, C. P. Otero and M. L. Zubizarreta (eds.) *Foundational issues in linguistic theory. Essays in honor of Jean-Roger Vergnaud*. Cambridge, MA: MIT Press. 134–166.
- Chomsky, Noam and Howard Lasnik. 1993. The theory of principles and parameters. In J. Jacobs, A. von Stechow, W. Sternefeld and T. Vennemann (eds.) *Syntax: An international handbook of contemporary research*. Vol. 1. Berlin & New York: Mouton de Gruyter. 506–569.



- Chrisomalis, S. 2010. Numerical notation: A comparative history. Cambridge: Cambridge University Press.
- Cohen Kadosh, Roi and Ann Dowker (eds.). 2015. Oxford handbook of numerical cognition. Oxford: Oxford University Press.
- Dehaene, Stanislas. 1997. The number sense. Oxford: Oxford University Press.
- É. Kiss, Katalin. 2010. Structural focus and exhaustivity. In M. Zimmermann and C. Féry (eds.) *Information structure. Theoretical, typological and experimental perspectives*. Oxford: Oxford University Press. 64–88.
- Feigenson, Lisa, Stanislas Dehaene and Elizabeth Spelke. 2004. Core systems of number. *Trends in Cognitive Sciences* 8. 307–314.
- Frank, Michael C., Daniel L. Everett, Evelina Fedorenko and Edward Gibson. 2008. Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition* 108. 819–824.
- Fuson, Karen C. 1988. *Children's counting and concepts of number*. New York: Springer-Verlag.
- Gelman, Rochel and Charles Randy Gallistel. 1978. *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- George, Alexander and Daniel J. Velleman. 2002. *Philosophies of mathematics*. Malden, MA & Oxford: Blackwell.
- Giaquinto, Marcus. 2015. Philosophy of number. In Cohen Kadosh & Dowker (2015, 17–31).
- Gunderson, Elizabeth A. and Susan Levine. 2011. Some types of parent number talk count more than others: Relations between parents' input and children's cardinal-number knowledge. *Developmental Science* 14. 1021–1032.
- Gunderson, Elizabeth A., Elizabet Spaenen, Dominic Gibson, Susan Goldin-Meadow and Susan C. Levine. 2015. Gesture as a window onto children's number knowledge. *Cognition* 144. 14–28.
- Hilbert, David. 1983. On the infinite. In P. Benacerraf and H. Putnam (eds.) *Philosophy of mathematics* (2nd ed.). Cambridge: Cambridge University Press. 183–201.
- Hiraiwa, Ken. 2016. The faculty of language and core systems of number. Manuscript. Meiji Gakuin University/MIT.
- Huang, Yi Ting, Elizabeth Spelke and Jesse Snedeker. 2013. What exactly do numbers mean? *Language Learning and Development* 9. 105–129.
- Hurford, James R. 1975. *The linguistic theory of numerals*. Cambridge: Cambridge University Press.
- Hurford, James R. 2001. Languages treat 1–4 specially. *Mind and Language* 16. 69–75.
- Ito, Junko and Armin Mester. 2015. Sino-Japanese phonology. In H. Kubozono (ed.) *Handbook of Japanese phonetics and phonology*. Berlin: de Gruyter Mouton. 289–312.
- Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kayne, Richard S. 2010. *Comparisons and contrasts*. Oxford: Oxford University Press.
- Kayne, Richard S. 2013. Why are there no directionality parameters? *Studies in Chinese Linguistics* 34. 3–37.
- Kayne, Richard S. 2016. Some thoughts on one and two and other numerals. Manuscript. New York University.

- Kennedy, Christopher. 2013. A scalar semantics for scalar readings of number words. In I. Caponigro and C. Cecchetto (eds.) *From grammar to meaning: The spontaneous logicity of language*. Cambridge: Cambridge University Press. 172–200.
- Le Corre, Mathieu and Susan Carey. 2007. One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition* 105. 395–438.
- Le Corre, Mathieu, Peggy Li, Becky H. Huang, Gisela Jia and Susan Carey. 2016. Numerical morphology supports early number word learning: Evidence from a comparison of young Mandarin and English learners. *Cognitive Psychology* 88. 162–186.
- Le Corre, Mathieu, Gretchen Van de Walle, Elizabeth M. Brannon and Susan Carey. 2006. Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology* 52. 130–169.
- Lipton, Jennifer S. and Elizabeth S. Spelke. 2006. Preschool children master the logic of number word meanings. *Cognition* 98. B57–B66.
- Lobeck, Anne. 1995. *Ellipsis: Functional heads, licensing, and identification*. Oxford: Oxford University Press.
- Menninger, Karl. 1969. *Number words and number symbols*. Cambridge, MA: MIT Press.
- Merchant, Jason. 2001. *The syntax of silence*. Oxford: Oxford University Press.
- Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27. 661–738.
- Nieder, Andreas. 2016. Representing something out of nothing: The dawning of zero. *Trends in Cognitive Sciences* 20. 830–842.
- Okamoto, Yukari. 2015. Mathematics learning in the USA and Japan. In Cohen Kadosh & Dowker (2015, 415–429).
- Overmann, Karenleigh A., Thomas Wynne and Frederick L. Coolidge. 2011. The prehistory of number concept. *Behavioral and Brain Sciences* 34. 142–144.
- Piantadosi, Steven T., Julian Jara-Ettinger and Edward Gibson. 2014. Children's learning of number words in an indigenous farming-foraging group. *Developmental Science* 17. 553–563.
- Piantadosi, Steven T., Joshua B. Tenenbaum and Noah D. Goodman. 2012. Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition* 123. 199–217.
- Pollmann, Thijs. 2003. Some principles involved in the acquisition of number words. *Language Acquisition* 11. 1–31.
- Rips, Lance J., Amber Bloomfield and Jennifer Asmuth. 2008. From numerical concepts to concepts of number. *Behavioral and Brain Sciences* 31. 623–642.
- Sarnecka, Barbara W. to appear. Learning to represent exact numbers. *Synthese*.
- Sarnecka, Barbara W., Valentina G. Kamenskaya, Yuko Yamana, Tamiko Ogura and Yulia B. Yudovina. 2007. From grammatical number to exact numbers: Early meanings of 'one', 'two', and 'three' in English, Russian, and Japanese. *Cognitive Psychology* 55. 136–168.
- Thornton, Rosalind. 2010. Verb phrase ellipsis in children's answers to questions. *Language Learning and Development* 6. 1–31.
- Wagner, Katie, Katherine Kimura, Pierina Cheung and David Barner. 2015. Why is number word learning hard? Evidence from bilingual learners. *Cognitive Psychology* 83. 1–21.

- Watanabe, Akira. 2010. Vague quantity, numerals, and natural numbers. *Syntax* 13. 37–77.
- Watanabe, Akira. 2014. 1-deletion: Measure nouns vs. classifiers. *Proceedings of Japanese/Korean Linguistics* 22. 245–260.
- Wynn, Karen. 1990. Children's understanding of counting. *Cognition* 36. 155–193.
- Wynn, Karen. 1992. Children's acquisition of the number words and the counting system. *Cognitive Psychology* 24. 220–251.
- Yamamoto, Kasumi. 2005. The acquisition of numeral classifiers: The case of Japanese children. Berlin & New York: Mouton de Gruyter.
- Zhang, Niina Ning. 2014. Expressing number productively in Mandarin Chinese. *Linguistics* 52. 1–34.