The Linguistic Origin of the Next Number

Charles Yang*
University of Pennsylvania

April 2016

> Still glides the Stream, and shall for ever glide;
> The Form remains, the Function never dies.
> Wordsworth (1820)

Mother:	I'll tell her the details of the party.
	There was six children who came.
Child:	Ten actually.
Mother:	Who came to your party?
Child:	Actually not ten. Ten hundred.
Mother:	Ten hundred.
Child:	Actually not ten hundred. Infinity.

(5;0: the CHILDES corpus)

1 Languages and Numbers

Discrete infinity is the hallmark of both natural languages and natural numbers. This has led some observers (e.g., Chomsky 1980, Hurford 1987) to conjecture that our linguistic and numerical capacities have a shared evolutionary origin. Such a position is not without difficulties (Gelman and Butterworth 2005). All human societies have infinite languages but not all cultures have an infinite number system, which apparently limits the speaker's performance on exact numerical tasks (Gordon 2004, Pica et al. 2004 though see Butterworth et al. 2008). All human children acquire a language effortlessly: many language-specific features are firmly in place by the beginning of language production (Brown 1973, Pierce 1992, Ramus et al. 1999, Yang 2013). By contrast, while pre-linguistic infants can solve simple addition and subtraction problems for very small numbers ("2=1+1", "3=2+1"; Wynn 1992a), numerical development follows a protracted period of several years (Fuson 1988, Wynn 1990, Feigenson and Carey 2003, Condry and Spelke 2008, Davidson et al. 2012). Moreover, home signers, who do not have a linguistically systematic

*The ideas presented here emerged at the workshop on language and number at the Lorentz Center (March 8-11 2016). I thank the organizers Pierre Pica and Johan Rooryck for their invitation and the workshop participants especially David Barner, Stan Dehaene, Randy Gallistel, Rochel Gelman, and Elizabeth Spelke for discussions. Subsequent conversations with Lila Gleitman, Peggy Li, and Herb Terrace also contributed to my understanding of the issues presented here. Minimal changes were made in April 2018 to update a few references.

input model, spontaneously invent a combinatorial gestural system similar to spoken/sign language syntax (Goldin-Meadow and Mylander 1998, Goldin-Meadow and Yang 2017), yet they do not develop exact numerosity representations (Spaepen et al. 2011).

Nevertheless, this programmatic note will advocate a strong identity thesis that connects language and numbers developmentally as well as evolutionarily. I claim that the acquisition of numbers, as concepts, follows the mechanisms of language acquisition. Number concepts are like words; their phonological forms are given by the morphosyntax of the numeral system, and their semantic contents are provided by the successor function (SF), that each integer has a successor that increments by exactly one (Russell 1919, Gelman and Gallistel 1978). The successor function is thus a generative procedure that recursively assigns semantic meanings to number concepts. Its acquisition, I claim, is enabled by the acquisition of the numeral system, which recursively assigns phonological forms to number concepts: the ordering it imposes on numbers is the crucial step in the semantic induction of numbers. My proposal is outlined as follows:

(1) a. Base case: the child has innately available two concepts “1” and “2” for which the successor function holds (i.e., $2 = 1 + 1$; Wynn 1990, 1992a);

b. Numeral acquisition: the child acquires the numeral system such that the next numeral is formally predictable from the previous one thereby establishing an ordered list of number concepts;

c. Induction: the property that holds between the 1st and 2nd concepts on such an ordered list (i.e., the successor relation between 1 and 2) is extended to any two successive concepts (i.e., the successor function for N and $N + 1$).

As I discuss in Section 5, my proposal builds on a large body of previous theories of numerical development. The key innovation here is that the successor function is enabled by a *productive* numeral system:[1] morphosyntactic infinity in the linguistic system provides the anchoring for numerical infinity in the conceptual system.

Because the numeral system is fundamentally linguistic, the critical step in the acquisition of numbers becomes a problem of language acquisition. Specifically, the child needs to identify the productive rules for numeral formation in the face of the unpredictable and thus rote-learned numeral words. For instance, in the English numeral system, *seventeen*, *twenty-five* and *eighty-two* are formally predictable whereas *three*, *twelve*, and *thirty* are not. The acquisition of rules along with exceptions is the sort of problem that all language-learning children face and generally excel at. I propose that a theory of language learning, dubbed the Tolerance Principle (Yang 2016), provides the causal, and quantitatively precise, link between the numeral system and the successor function.

My argument will be given as follows. Section 2 reviews the key features of the Tolerance Principle, which accounts for how children overcome exceptions to learn rules. Section 3 reviews cross-linguistic studies of children's counting to show that the Tolerance Principle makes accurate predictions for when children discover their numeral system and why it takes so long. Section 4 presents preliminary evidence that the successor function seems to become available as soon as — but not before – the acquisition of the numeral system. Section 5 gives a more detailed statement of my proposal in (1) and highlights how the acquisition of numbers is fundamentally

[1] I use the term *productive* as opposed to *generative* to describe an infinite numeral system. Generative processes describe input-output relations which may be restricted to a finite number of items. Productive processes, on the other hand, are open-ended and infinite.

linguistic. In the concluding section, I will articulate a new cultural and evolutionary perspective on the connection between natural language and natural number.

2 How Children Generalize

Linguistic rules often take a long time to learn. The most prominent example in the language development literature is the acquisition of English past tense. Children initially are locked in a protracted stage during which the irregular verbs are inflected correctly (Marcus et al. 1992, Maratsos 2000). Then rather abruptly, over-regularization errors (e.g., *think-thinked*, *know-knowed*) start to appear. This marks the emergence of the productive regular rule ("add -d") which can apply to all verbs, affecting even the irregulars that never appear with "-d" in the input. An important question in the study of English past tense is to determine the "tipping point" at which the "add -d" rule becomes productive: the child must learn enough regular verbs to overcome the irregulars. Similar questions arise for the acquisition of linguistic rules in general, many of which have exceptions.

The Tolerance Principle, which is summarized in Yang (2016), is a theory that provides for the threshold of exceptions for a productive rule. In the case of English past tense, the regular "add -d" rule becomes productive only if the child has acquired a sufficiently large number of verbs such that the number of exceptions (i.e., the irregulars) falls below the threshold. The motivation for Tolerance Principle comes from computational efficiency considerations of language use. It holds that when faced with alternative organizations of linguistic data, the child chooses a more efficient, i.e., faster, grammar. For the acquisition of rules, the choice is between a productive rule with a list of exceptions or a fully lexicalized list that dispenses with the rule altogether. By comparing the cost of full listing and the cost of a rule plus exceptions, we can derive a threshold value of exceptions below which postulating a productive rule results in faster processing time. Under general assumptions about word frequencies (i.e., Zipf's Law; 1949), it is possible to prove the following:

(2) **Tolerance Principle**:
If R is a productive rule applicable to N candidates, then the following relation holds between N and e, the number of exceptions that could but do not follow R:

$$e \leq \theta_N \text{ where } \theta_N := \frac{N}{\ln N}$$

Yang (2016) presents dozens of empirical case studies to show that the Tolerance Principle makes accurate predictions about where productivity arises in language and where it collapses. The slow growth of the $N/\ln N$ function suggests that for a productive rule to emerge, there must be an overwhelming number of rule-following items to overcome the exceptions. Recent experiments with Kathryn Schuler and Elissa Newport using the artificial language paradigm have produced near-categorical support for the numerical predictions of the Tolerance Principle (Schuler et al. 2016). Children between the age of 5 and 7 were presented with 9 novel objects with labels. The experimenter produced both "singular" and "plural" forms of those nouns as determined by their quantity on a computer screen. In one condition, five of the nouns share a plural suffix and the other four have individually specific suffixes. In another condition, only three share a suffix and the other six are all individually specific. Thus, the nouns that share the suffix are the regulars and the rest of the nouns are the irregulars. The choice of 5/4 and 3/6 was by design: the Tolerance Principle predicts the productive extension of the shared suffix in the 5/4 condition because 4 exceptions are below the threshold ($\theta_9 = 4.1$), but no generalization in the 3/6 conditions. In the latter case, despite the statistical dominance of the shared suffix as the most frequent suffix, the six exceptions exceed the threshold. When presented on additional novel items in a Wug-like test, nearly all children in the 5/4 condition generalized the shared suffix on 100% of the test items in a process akin to the productive use of English *-ed*. In the 3/6 condition, almost no child showed systematic usage of any suffix, much like native speakers when queried with "gapped" (ineffable) morphological items (e.g., the past participle of *stride* whose past tense is *strode*, and numerous other cases in morphologically complex languages; see Halle 1973, Baerman and Corbett 2010, Yang 2016).

The Tolerance Principle is a parameter-free model in the sense that it is not tuned to fit any linguistic data or experimental results. It makes a categorical prediction about productivity on two input variables (N and e). I now give a simple example to show how the Tolerance Principle accounts for the emergence of the "add -d" rule in English past tense; the application to the acquisition of numeral systems follows the same process.

Suppose an English-learning child knows $e = 120$ irregular verbs, roughly the total number found in a five million word corpus of child-directed English (MacWhinney 2000). We can infer the minimum value of N, the number of all verbs (regular plus irregular), such that $\theta_N = N/\ln N \geq 120$. A simple calculation shows that the value is about 800. That is, if the child has acquired 120 irregular verbs, they will need at least 680 regular verbs to establish the productivity of the "add -d" rule. This condition is easily met: the five million word corpus of child-directed English contains over 900 regular verbs inflected in past tense. Thus, an English learner will be justified to conclude that the "add -d" rule is productive and can be extended to novel items (e.g., *rick-ricked*; Berko 1958).

In fact, the prediction can be made more precisely, potentially at the individual level. After all, productivity is determined by the numerical values of N and e, which are determined by an individual learner's vocabulary of regulars and irregulars. Consider "Adam", the poster child for English past tense acquisition (Pinker 1995). Adam produced the first instance of over-regularization error — "What dat feeled like?" — at the age of 2;11. In the transcript of almost a year prior to that point, not a single irregular past tense is incorrect. If we take over-regularization as the moment when "add -d" becomes productive (Marcus et al. 1992), then it must be the case that by this point, Adam has acquired a sufficiently large number of regulars to overwhelm the irregulars. To test this prediction, I extracted every verb stem in Adam's transcripts until 2;11 (MacWhinney 2000). There are $N = 300$ verbs in all, out of which $e = 57$ are irregulars. This is very close to the predicted $\theta_{300} = 53$, and the small discrepancy may be due to the undersampling of the regular verbs which tend to be lower in frequency and are more likely to be missed in a modest sample. The critical point to note here is that Adam needed a filibuster-proof majority of regular verbs to acquire the "add -d" rule: this is strongly consistent with the predictions of the Tolerance Principle.

The formulation of the Tolerance Principle suggests that the statistical composition of regular and irregular items in the learner's vocabulary plays a decisive role in the emergence of productivity. The "add -d" rule is learned relatively late because irregular verbs are among most frequent verbs in English. In a five million word corpus of child-directed English, 54 of the top 100 most frequent verbs in past tense are irregulars; the child must acquire a substantial verb vocabulary before learning that “add -d” is regular. By contrast, the irregular plural nouns (e.g., *child-children*) are far less frequent: only 6 of the top 100 most frequent plural nouns are irregular. The “add -s” rule is therefore acquired much earlier: “Adam”, for instance, produced his first token of over-pluralized noun (*peoples*) a full three months before the attestation of *feeled*.

The Tolerance Principle can be applied to numeral systems, which are also a mixture of regularly formed expressions along with rote-learned items. It makes precise predictions on when children acquire a productive numeral system, which will be manifested in their counting range. This, I then argue, leads to the acquisition of the successor function and the system of natural numbers.

3 To 73 and Beyond

While reading a classic paper (Fuson et al. 1982) on English-learning children’s development of counting, I was suddenly overcome with excitement. The authors chose to present their data in a particular fashion (Table 1).

Age/grade[a]	$n < 10$	$10 \leq n < 14$	$14 \leq n < 20$	$20 \leq n < 30$	$30 \leq n < 72$	$72 \leq n < 101$	$101 \leq n < 201$	$201 \leq n$
3 years 6 months to 3 years 11 months	17	44	22	17	0	0	0	0
4 years to 4 years 5 months	0	41	35	12	12	0	0	0
4 years 6 months to 4 years 11 months	0	12	47	18	12	12	0	0
5 years to 5 years 5 months	0	6	25	13	44	13	0	0
5 years 6 months to 5 years 11 months	0	6	22	17	44	11	0	0
Kindergarten	0	7	11	30	26	4	22	0
First grade	0	0	3	14	7	21	48	7
Second grade	0	0	0	0	8	3	31	58
Third grade	0	0	0	0	0	4	25	71

Table 1: Percentage of age groups producing accurate sequences of variable lengths (from Fuson et al. 1982: p38).

Here children are grouped according to their counting range. The lower values are divided by the decade but why is 72 a “key point” (Fuson et al. 1982: p39) for breakdown? The paper does not provide any specific justification but Karen Fuson helpfully clarified in an email:

> You can see two major learning tasks in the chunks in the table: learning the teens count and learning from 30 to 72. It can take a long time to learn the teens count because of all the irregularities in English. Fifteen seems to be especially difficult. Then children get the pattern 21 to 29, and they may get this before they consistently know all of the teen num bers. But the irregularities in the teens and especially ten, eleven, twelve, interfere with learning the repeat of the twenties pattern with 30. There is nothing to signal that the count starts repeating at ten, so children count ... twenty nine, twenty ten, twenty eleven, etc. Inhibiting this can take time. Once they finally solidly have the 30 31 ... 39 pattern, they have to learn the decades forty, fifty, sixty, which are irregular and do not seem to come from four, five, six (sixty does of course) and *that is one reason children take off after 72.* (Personal communication. March 31st, 2016. Emphasis added)

There is a reason why children take off after 72, which is in fact predicted by the Tolerance Principle. The acquisition of an open-ended counting list is only possible if the child has mastered a productive numeral system, which in turn must meet the condition of productivity in the face of the rote-learned items. Consider the English numerals below 100:

(3) **one two three four five six seven eight nine ten eleven twelve thirteen** 14 **fifteen** 16 17 18 **twenty** 21 22 23 24 25 26 27 28 29 **thirty** 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 **fifty** 51 52 ...

Here the rote-learned numerals are in boldface, 17 in all. According to the Tolerance Principle, the acquisition of the productive numeral system in English requires the child to have learned at least N numerals such that $\theta_N = N/\ln N \geq 17$: the smallest value of N is 73. In other words, if a child can count to 73, they are guaranteed to "take off".[2]

Although the Tolerance Principle has proven more effective than I had expected (or hoped), this calculation is surely too fantastic to be exact. The tipping point of 73 can only be regarded as an approximation. I have not said anything about how the child discovers the numeral rules but only counted the number of items (17) that would not be predictable in the target-like system. No doubt children eventually arrive a such rules but how they do so in an incremental fashion needs further investigation and it is likely that they are guided by the constraints on numeral systems in the world's languages (Hurford 1975). Here we also expect to find individual differences in the construction of the rules (and thus counting lists), as is generally the case in the acquisition of linguistic rules such as morphology (Freyd and Baron 1982, Maratsos 2000). Nevertheless, since everyone knows how to count forever but no one learns that by literally counting forever, some upper bound on the numeral vocabulary must be sufficient for the acquisition of the numeral system. The Tolerance Principle provides a rough estimate of where the transition to infinity may be. An immediate prediction is that once such a transition has taken place, children may start producing "over-regularization" errors, such as *ten-hundred* noted at the beginning of this paper. Presumably the child did not know the rote-learned word *thousand* well enough but the fact that *ten-hundred* was produced at all suggests that he understood English numerals to be infinite, just as irregular error such as *feel-feeled* marked the productivity of "add -d".

For any language, the Tolerance Principle can predict the transitional point of numeral acquisition using only the number of rote-learned words. Consider a much simpler (i.e., more regular) numeral system, that of Mandarin Chinese:

(4) **一 二 三 四 五 六 七 八 九 十 十一** 十二 十三 十四 十五 十六 十七 十八 十九 **二十** 二十一 二十二 二十三 二十四 二十五 ...

The numerals from 1 to 10 must be rote-learned. The numeral 11 ("ten-one") is also unpredictable as it is positionally differentiated from 20 ("two-ten") by convention. Taken together, there are at most 11 exceptions; the Tolerance Principle suggests that the Chinese learner needs to acquire at least 42 numeral words to discover the productive process of numeral formation ($N_{42} = 11$). In fact, it has been known for quite sometime now that "four-year-olds in China made very rapid progress in generalizing number names up to 100 after they could count to approximately 40, where few U.S. children grasped the morphological patterns that would allow

[2]Strictly speaking, the Tolerance Principle does not require the 73 numbers to be consecutive as long as they are unique, although it is almost surely that children can only learn 73 unique numerals through counting.

them to count to 100" (Miller et al. 2005:p168). The simplicity of the numeral system readily accounts for the fact that within the same age group, Chinese-learning children count significantly higher than American children (Miller and Stigler 1987); see also Miura et al. (1988) for other East Asian languages and Song and Ginsburg (1987) for contrasts in Korean children acquiring two numeral systems with different degrees of regularity. Especially interesting is a study by Miller et al. (1995). Chinese and American children were assessed for their counting range and tested on object counting. As found previously, Chinese children count significantly higher than American children as soon as the counting range goes in the teens, which is where the two systems begin to diverge in regularity. More significantly, Chinese children hold a significant advantage over American children on the task of object marking that involves a large numbers (14-17) of items: American children were successful 33% of the time as opposed to 60% for Chinese children. This result is already suggestive that the mastery of counting, and the complexity of the numeral system, may be causally connected to the acquisition of number concepts and children's performance on numerical tasks, which is of course the major theme developed here.

Two additional predictions can be made under the current approach to numerical development. First, and rather straightforwardly, we account for the protracted development of numbers (Wynn 1992b). On the present account, this is because the acquisition of the numeral system is protracted. The statistical distribution of numerals is similar to that of English verbs in past tense, only even more irregular-heavy in the highest frequency range; see Dehaene and Mehler (1992) for a cross-linguistic study of similar results. Figure 1 shows the distribution of the numeral words in a five million word corpus of child-directed English. The most frequent regularly formed numeral is "nineteen" and the more frequent rote-learned numerals overwhelmingly dominate: just the three words "one", "two" and "three" account for over 75% of all tokens. With the help of Kyle Gorman, I also extracted all the numeral expressions from the same corpus.[3] Remarkably, about 95% of the numeral tokens, spread over the 17 types in (6a), are rote-learned. The acquisition of a numeral system, then, can only proceed slowly, much like the acquisition of the productive "-ed" rule in English past tense, where a learner's early verb vocabulary will consist of many high frequency irregular verbs. If, as proposed, the acquisition of SF is enabled by the acquisition of a productive numeral system, we can readily account for the correlation found between children's numerical knowledge and their numerical vocabulary (Negen and Sarnecka 2012).

Second, and more interestingly, the Tolerance Principle makes strong predictions about the distribution of children's counting ranges. Roughly speaking, we expect a ʊ-like distribution. At the one end, there should be children that cannot go beyond the rote-learned numerals spread over an extended range, as they have not yet reached the productivity threshold. At the other end, there should be children who have acquired the productive numeral system. Critically, we expect a no-child's-land between the tipping points — approximately 73 for English (Fuson et al. 1982) and 42 for Chinese (Miller et al. 2005) –and the ceiling, i.e., an infinite system, typically capped at 100 in empirical studies. This was exactly what Lipton and Spelke (2005) found. In their study, over half of the 5-year olds can count up to 100 but interestingly, they did not find any child who could count to 80 but not all the way to 100.

Indeed, cross-linguistic studies have consistently pointed to the regularity of numeral systems and its effect on children's counting and related numerical tasks (for English vs. French,

[3]For instance, the numeral word analysis in Figure 1 treats the adult utterance "thirty-nine" as "thirty"and "nine". The numeral analysis, which uses a regular expressin extractor based on the English system of numeral formation, treats it as a single number 39. The corpus contains only a finite number of numeral words but many more numerals.

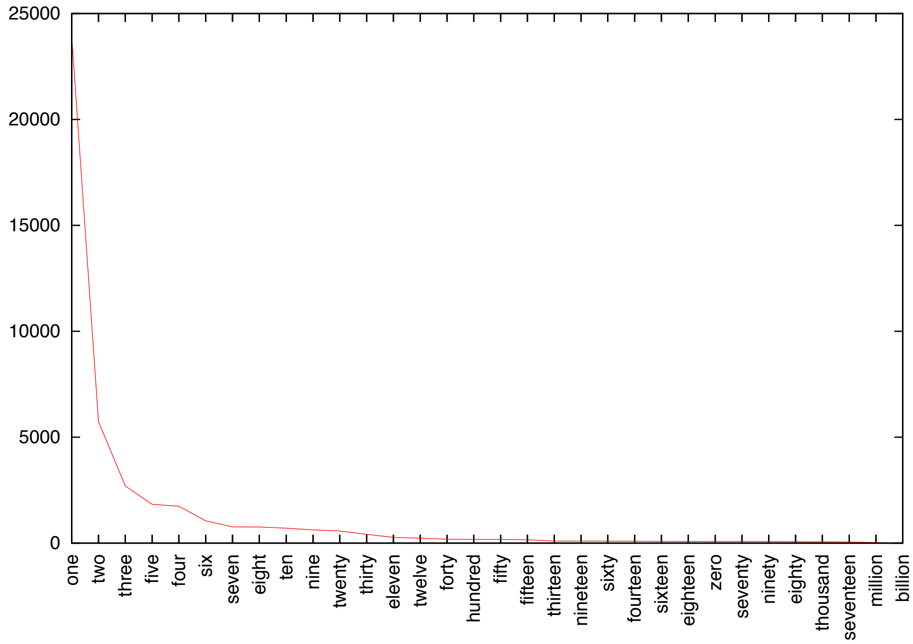

Figure 1: Distribution of English numeral words in child-directed speech.

Lefevre et al. 2002; Welsh, English, and Tamil, Dowker et al. 2008; the, base-5, Belgian Sign Language vs. Belgian French, Leybaert and Cutsem 2002). Future research may follow the methods of Lipton and Spelke (2005), Wagner et al. (2015), Cheung et al. (2017) to systematically assess the children's counting and numerical development across languages. If the current proposal is correct, we expect to observe parallel trajectories, with a quantal change at the point predicted by the Tolerance Principle. In Section 4 I review some preliminary evidence to this effect, before giving a detailed account of the inductive process in Section 5.

4 From Counting to Successor Function

I should make clear that I do not claim that learning to count with a numeral system necessarily or immediately leads to the acquisition of the successor function. Being able to count to 10 or even 20 in a language like English is pretty useless for the purpose of establishing a productive numeral system. Analogously, if a child knows only 10 verbs out of which 9 are irregular, there is no way for them to learn the "add -d" rule is regular. The claim is that the successor function becomes available only after the child can count to a sufficiently high number such as productivity is guaranteed: How high that number is can be provided by the Tolerance Principle. I believe that this claim is supported by the findings on counting and successor function in the previous literature.

In the so-called subset-knower stage, children who can recite a count list may fail to give the exact number of objects that is lower than their counting range (the Give-N task; Wynn 1990). Nor can children in this stage describe precisely the number of objects on cards (Le Corre et al.

2006; see also Gelman 1993). What follows is the CP-knower stage, named after Gelman and Gallistel's Cardinality Principle (1978), where children can reliably produce any number within their counting range and can use the last number to represent the cardinality of the set. For some authors, the CP-knower stage is identified with the acquisition of SF (Wynn 1990, Condry and Spelke 2008, Sarnecka and Carey 2008, Carey 2009). But children at the beginning of the CP-knower stage typically have a very short counting list, while a true understanding of the successor function ought to extend to arbitrary numbers.

Davidson et al. (2012) provided a systematic study of the connection between CP-knowing and the successor function. They asked whether children knew that adding one item to a set corresponds to the next word on the counting list (the Unit task). Unlike previous studies (e.g., Le Corre and Carey 2007, Sarnecka and Carey 2008) that focused on small values of cardinality, these authors tested children on small (4, 5), medium (13, 14, 15), and large numbers (24, 25). They found that CP-knowers very often fail on the Unit task. For instance, there are children who can count past 30 yet fail on the Unit task for numbers below 10. Overall, for children in the low (10-19) and medium (20-29) counting range, their modal response on the Unit task is only at chance even for small numbers. By contrast, children in the high (≥ 30) counting range performed better across the board. This study casts doubt on the proposal that CP-knowing necessarily leads to the successor function; see also Le Corre (2014) for similar findings on a smaller range of CP values. Moreover, Izard et al.'s recent study (2014) using a non-linguistic task shows that three-year-old children who can track a set of 5 to 6 objects nevertheless fail to understand the change of cardinality when an object is added to or subtracted from the set: it is possible that CP-knowing does not necessarily lead to successor function even for very small numbers.

Having a sufficiently long counting list, however, does seem to activate SF in numerical development. To the best of my knowledge, Cheung et al. (2017) is the first systematic study of counting range and the acquisition of SF when assessed on a large range of numeral values, and it lends direct support for the current proposal. Building on Davidson et al. (2012), Cheung et al. found an almost two year gap between becoming CP-knowers and learning the successor function, the latter of which typically occurs between 5 and 6. These authors tested the successor function with the Successor Task (after Sarnecka and Carey 2008). Children were told the number of objects that the experimenter had put in an opaque box: "I have N bears. I'm putting N bears in the box". Another object (e.g., bear) was added to the box, and the child was to choose between $N+1$ and $N+2$ as the answer. The task was carried out on four ranges of N: Small (4, 5, 7), Medium (12, 16, 18), Large (23, 24, 28, 31, 35, 36) and Very Large (53, 57, 76, 77). One hundred children were tested (mean age 5;6; range = 4;0–6;11); all were CP-knowers. The children were divided in four groups by counting range: Low Counters (8-19), Medium Counters (20-39), High Counters (40-79), and Very High Counters (80-100). The results are as follows. While all children performed above chance on the Small Number test, only half of Medium and High Counters consistently answered correctly for numbers in their respective counting range. Most revealingly, the Very High Counters as a group performed near ceiling on all numbers on the Successor Task. It is as if 80 marks a decisive phase transition into the successor function region.[4] The discrete nature of this development, as well as the critical point at which such transition takes place, are strongly consistent with the approach developed here.

Before proceeding, I should also note that it is possible for children to develop exact numeral

[4]This work does not provide a finer classification among the High Counters (40-79). A stronger prediction is that children in the range of 73-79 ought to behave like the Very High counters.

knowledge for a finite, and likely small, range of numbers. Many children are explicitly taught how to count, with fingers, small objects, or abacus (the CHILDES corpus records numerous examples). They may learn that, for instance, 8 follows 7, 13 follows 12, etc. Thus, children may appear as if to have acquired a "bounded" successor function for a small set of numbers. Examples of a bounded successor relation include the days of the week, the months of the year, etc., as Wednesday follows Tuesday, and October follows September. Such knowledge would be item-specific and learned by rote, rather than what is entailed by the unbounded successor function. I suspect this is the beginning stage of CP-knowing: children appear to have developed the successor function for very small numbers tested in earlier studies (e.g., Wynn 1990, Condry and Spelke 2008, Sarnecka and Carey 2008), but fail to generalize beyond those, not even to numbers within their counting range (Davidson et al. 2012). As we shall see in the next section, such item-based knowledge is in fact a typical stage of language development and is expected under the current proposal that grounds numerical capacity in language acquisition.

To fully test the predictions of the current proposal, more studies like Cheung et al. (2017) are needed for additional languages. We predict critical points for the acquisition of language specific numeral systems, which can be verified by children's counting range. If a productive numeral system is indeed the cause for the successor function, then we expect to see dramatic and across-the-board improvement in the performance of numerical tasks much like the Very High Counters in Cheung et al.'s study (2017).

I will now spell out how the numeral system facilitates the inductive processes in the acquisition of the successor function.

5 Linguistic Acquisition and Numerical Induction

5.1 Rule learning as induction

To learning a language is to learn its rules, which provide a systematic mediation between form and meaning. To a learn rule is to carry out an induction: the capacity to form abstractions such that the properties of a finite number of items can be extended to an entire class with infinitely many members. For instance, when the child knows the English noun phrase rule "NP → Det Noun", they know that two classes of words, determiners and nouns, can be combined freely. During language acquisition, the child will establish such rules based on a finite number of examples: obviously no one can ever hear all possible determiner-noun combinations. Induction immediately follows the successful acquisition of the rule.

The inference process can be turned the other around. If the child knows the NP rule and encounters an expression such as "the dax", then "dax" will be instantly analyzed as a noun as befitting the word category that follows the determiner (Shi and Melançon 2010). In other words, forms can be used to deduce meanings, as shown by Roger Brown in a pioneering study (1957) and more recently in the literature known as syntactic bootstrapping (Gleitman 1990): If a novel word appears between two noun phrases, it is most likely a verb, and it will have the transitive semantics of agency.

Of course, the child must learn the language-specific rules that enable the induction from finiteness to infinity and the inference from form to meaning. In many instances, the rules are available from very early on (Valian 1986, Golinkoff et al. 1987, Yang 2013): this is important for theories such as syntactic bootstrapping but it deprives us the opportunity to observe the transi-

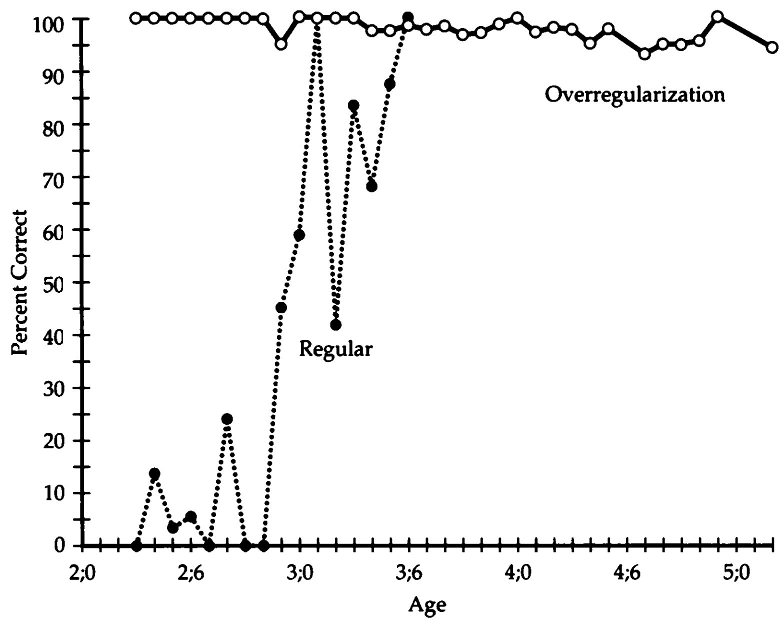

Figure 2: Past tense marking for regular and irregular verbs (from Pinker 1995: p116).

tion from a finite list of items to a general class. Occasionally, however, we do capture the moment when a rule becomes productive: the effect of generalization is quite dramatic and can only be described as a moment of epiphany. I return once again to the acquisition of English past tense.

As noted earlier, the acquisition of the "add -d" rule can be protracted process: Adam produced his first instance of over-regularization just before his third birthday. But the sudden emergence of productivity triggers a wholesale change to the child's grammar. It is widely observed that children acquiring languages such as English have a strong tendency of using nonfinite matrix verbs when the finite form is obligatory; see Wexler (1994), Legate and Yang (2007), Phillips (2010) for reviews. The examples in (5) are naturally occurring examples from the CHILDES database where the verb should have been marked for past tense. Figure 2 shows Adam's longitudinal development of past tense marking.

(5) The dollies we *have* last week.
I *go* under the tunnel yesterday.
He *walk* around and he said and he's my frog.

Adam's regular past tense marking ("Regular") was initially very sporadic: that is, almost all instances where the target form should be "He walked around" surfaced as "He walk around" instead. These forms are almost certainly rote-learned and item-specific despite the appearance of regular inflection. Then, and apparently abruptly, the rate of regular past tense marking increased significantly, exactly around the point where he produced the first instance of overregularization (the first dip in "Overregularization"). When Adam discovered that the "add -d" is productive, he started to use it across the board for verbs, with the irregulars as collateral damage. The "add -d" rule, which defines a formal relation between pairs of words (i.e., *X-Xed*), becomes the inductive basis about the meanings of such pairs: that the *walked* denotes the semantics of *walk* in past time can be extended to pairs such as *rick-ricked*, the novel verb created in Berko (1958)'s classic study and its past tense realization effortlessly created by young children.

In what follows, I propose that the induction of numbers follows exactly the same logic and process of linguistic rule acquisition.

5.2 Successor induction: From form to meaning

In a nutshell, I suggest that the numeral system establishes an ordered list of number concepts: the predictability in the *forms* of successive numbers provides the inductive base for the *semantics* of successive numbers, which enables child to generalize the semantic relation between *some* successive numbers to *all* successive numbers, thereby acquiring the successor function.

On this view, the acquisition of numbers (as concepts) is similar to the acquisition of words.[5] To learn a word is to learn its form and meaning; thus, to acquire a number is to learn its phonology and semantics. I use the Arabic numeral (e.g., 79) to denote a number concept and use ϕ and ψ to represent its phonology and semantics respectively. For instance, $\phi(79)$ = "seventy-nine", and $\psi(79) = \psi(78) + 1$, which is provided by the successor function.

One reason for supposing that number concepts are acquired like words comes from chilren's interpretation of numerals. Children appear to follow the Mutual Exclusivity constraint (Clark 1987, Markman and Wachtel 1988), a well-established principle for word learning,[6] which has the effect of the one-to-one correspondence principle in the study of numerical cognition (Gelman and Gallistel 1978). The Mutual Exclusivity constraint is operative in very young infants (e.g., Halberda 2003): if concepts are like words, we expect children to know different numerals (e.g., "fifty-nine" and "seventy-eight") should denote different quantities. This would be so even if children do not know the exact quantity of these numerals as long as they know these *are* numerals, presumably from the syntactic cues ("there are N bears"), which is fundamentally the same behavior when a child regards "dax" and "bem" as the label for different objects. Mutual Exclusivity accounts for findings such as those in Lipton and Spelke (2006)'s study. Children were tested on two tasks. In the mapping task, they were instructed to identify a card that has N rectangles against a distractor card that only contains half or twice the rectangles. In some of the trials, N is beyond their counting range, which is assessed by the method of Lipton and Spelke (2005: see Fn **??**). In the logic task, they were told that there are N objects in a set, again with some values of N greater than their counting rage. These objects then undergo various transformations, some are quantity perserving (e.g., stirring, taking half of the objects away then putting them back) and others are quantity changing (e.g., removing one, removing half). Children were then asked if there were still N objects. Consistent with previous findings, children were at chance with the mapping task, showing that they did not know the exact quantity represented by N. But for the logic task, they were nearly categorical in recognizing that stirring and putting half back did not change N but removing one and removing half of the objects did. Importantly, since the values of N are beyond the children's counting range, and they were only at chance in distinguishing N

[5]Indeed, there are similarities in the brain activation patterns during the processing of words and exact numbers (Dehaene et al. 1999).

[6]Mutual Exclusivity is understood as a very strong preference but not a categorical rule so as to allow (near) synonyms such as *pail* and *bucket*. It is also possible for multiple numerals to refer to the same quantity, as in the case of *twelve* and *dozen*, but such pairs are rare and do not represent the general case.

from $N/2$ or $2N$, they clearly do not know the precise semantics of N. Yet, as one would expect from the Mutual Exclusivity constraint, they understand that N, or more precisely, the phonological form of N, cannot refer to multiple quantities.

Like all inductive processes, the induction of numbers requires a base case that must be independently established. I suggest that the base case is provided innately (Wynn 1992a). Children, and possibly other animals, innately know the semantics of 1 and 2: $\psi(1) = 1, \psi(2) = \psi(1) + 1$. It is possible that children's innate knowledge about exact numbers extends slightly higher than 2 (e.g., 3=2+1), but 1 and 2 alone are sufficient for the purpose of my argument. For these numbers, all the child needs to learn is their phonological form: $\phi(1) =$ "one", $\phi(2) =$ "two", or $\phi(1) =$ "一", $\phi(2) =$ "二".

The inductive step is enabled by the numeral system, which is a set of morpho-syntactic rules R such that $\phi(N+1) = R(\phi(N))$. That is, the form of the next number is always predictable on the basis of the previous number. The Mutual Exclusivity constraint ensures that the semantics of N and $(N+1)$ must be different, i.e., they represent different quantities. The inductive step extends the property that holds for some successive members on the ordered list to successive members on the same list. Specifically, what is true of 1 and 2, $\psi(2) = \psi(1) + 1$ (that 2 is exactly 1 plus 1), is generalized to $\psi(N+1) = \psi(N)+1$. On this view, only a few concepts that represent numbers need to be innate: minimally, 1 and 2. The numeral system (R) and the inductive process provide a way of constructing infinitely many number concepts in both form and meaning: $\phi(N+1) = R(\phi(N))$, which is a linguistic expression, and $\psi(N+1) = \psi(N) + 1$, which is a discrete quantity.

Thus, the acquisition of numbers involves a qualitative conceptual change (Carey 2009, Sarnecka and Lee 2009). Prior to the discovery of a productive numeral system, children's number concepts are in effect item specific. That is, they may know, from explicit instructions (e.g., counting with objects), that 9 is exactly one more than 8: it gives the appearance of knowing the successor function but is in fact rote-learned and does not generalize beyond direct experience. Likewise, they may be able to produce, on their counting list, a numeral word such as *seventeen*, which follows the target system of numeral formation but is in fact holistically stored.[7] This stage is similar to how children represent regularly inflected verbs prior to the emergence of "add -d" (Figure 2): although they may have learned that *talked* refers the act of talking in the past, this knowledge is limited to the particular verb *talk*. The generalization that all verbs express the meaning of past by adding "-d" only becomes available when the rule "add -d" becomes productive. For past tense, this happens when the child has learned a sufficiently large number of regular verbs to overcome the irregulars. For numbers and their successor relation, I suggest that this takes place when the counting list has gotten sufficiently long: for English, this is around 73.

The proposal developed here, which I believe is novel, shares many features with previous theories of numerical development. The Tolerance Principle, which I have argued to be the key mechanism for the acquisition of numerals, provides a formal mechanism for theories of counting that have consistently emphasized and documented the effect of linguistic regularity (e.g., Fuson 1988, Miller et al. 1995, Nunes and Bryant 1996). Like Gelman and Gallistel (1978), Gel-

[7] I leave open the possibility that children may be able to extract regularities within a bounded range. For instance, an English-learning child may attempt to construct productive rules within the subdomain of teens and decades. A recursive application of the Tolerance Principle suggests that they are unlikely to succeed. For the 10 numbers in the teens (11-20), 5 are not predictable and must be rote-learned (11, 12, 13, 15, 20), which exceeds the threshold $\theta_{10} = 4.2$. The decade numerals have 4 rote-learned items (10, 20, 30, 50), so the child needs to reach at least 90 to discover the productive rule of *-ty*.

man (1993), Leslie et al. (2008), I regard numbers as discrete symbolic representations and the acquisition of an ordered list is essential to the development of numbers. Unlike these authors, I do not regard (all) number concepts to be innately available or independently of language, only to be "triggered" (a la Fodor 1975) by experience, including the mapping them to words. As discussed in Section 6, I claim that numbers and the successor function are the product of language and language acquisition, which also provide a way to recursively construct numbers from very few innate concepts. In this sense, my proposal is similar to Carey (2009). Like Carey, I assume that the acquisition of a counting list is the key to the development of numbers, which is characterized by a developmental discontinuity when the successor function is acquired. I differ from Carey on what enables such a change and how long the list needs to be. For Carey, this involves an inductive process ("Quinian bootstrapping") on semantic grounds which need relatively few exemplars. Under the present proposal, the induction is facilitated by the formal system of numerals, which typically requires dozens of numeral expressions for most languages. My position on the relationship between language and number is probably closest to Spelke's theory (2003; cf. Bloom and Wynn 1997). Spelke also regards the productive use of language as the center of numerical cognition. For her most recent account (2016), the role of language is realized in the acquisition of noun phrases that make reference to objects and/or sets of objects, including the use of conjunction, quantifiers, as well as numerical expressions. This aspect of linguistic knowledge interacts with other domains of core language, such as the approximate number system, to provide the child with the system of natural numbers. Since Spelke's and my proposal make use of different linguistic structures for the acquisition of natural numbers, future empirical work may test and distinguish these positions.

6 Language, Culture, and Evolution

Before concluding, I must confess a slight sense of unease by the strongly Whorfian character of my proposal. For English-learning children with a small numeral vocabulary, which presumably contains mostly of the 17 rote-learned items, the successor function is impossible. By this logic, I am forced to claim that monolingual speakers of a language without a productive numeral system *cannot* acquire the successor function and do not have the knowledge of exact numbers. While this is consistent with previous findings on Piraha (Gordon 2004) and Munduruku (Pica et al. 2004), I emphasize that I do not claim that these speakers lack the capacity of a natural language system. In fact, there is suggestive evidence that they do have such a capacity. First, Kenneth Hale found that for speakers of Warlpiri, which only has number words *one*, *two*, *few*, and *many*, "the English counting system is almost always instantaneously mastered by Warlpiris who enter into situations where the use of money is important (quite independently of formal Western-style education" (cited in Gelman and Butterworth 2005). Second, Spelke (2016) reports a study by Véronique Izard in which Munduruku speakers were introduced to an exact number system using the combinations of Munduruku number words. Once the task was made clear to the subjects, mononlingual Munduruku speakers showed some ability of representing exact numeral values when supported with combinatorial rules; see Fn 7 for related discussion.

I suggest that there is another way of looking at numerical cognition and language, one which does not commit one to a strong form of linguistic determinism, which has had a very mixed track record to say the least. The Whorfian hypothesis maintains a causal relationship between language and the *non-linguistic* aspects of perception and cognition; numerical capacity is typically considered to fall into the latter category. This somewhat awkward perspective is avoided if numbers are *actually* linguistic. If so, cross-linguistic differences in numerical cognition are really just cross-linguistic differences in the numeral systems. Thus, the fact that some languages have a productive numeral system while other do not says nothing about the generative capacity of these languages nor the cognitive capacity of the speakers of these languages. Some examples from more familiar languages can make this point succinctly.

(6) a. Maria's neighbor's friend's house
 b. Marias Haus
 c. *Marias Nachbars Freundins Haus
 Maria's neighbor's friend's house

In English, noun phrase allows unlimited embedding of possessives as in (6a). In German and in fact most Germanic languages (Roeper 2011), the embedding stops at level 1 (6b) but not further (6c). Presumably, German children do not allow the counterpart to (6a) because they do not encounter sufficient evidence that enables such an inductive generalization, whereas English children do: an interesting acquisition problem in its own right. But no one would suggest that German children lack the capacity for infinite embedding on the basis of the examples in (6). The fact that not all languages have recursive expressions in every domain is not particularly surprising (Everett 2005, Nevins et al. 2009, Legate 2011).

I do not wish to trivialize the centrality of numerical cognition and its profound impact on human history and civilization. Clearly, language alone is not sufficient for numerical capacity to emerge. Indeed, many components in the present theory of numerical induction appear to be evolutionarily ancient and shared across species. These include the exact knowledge of very small quantities (Gallistel 1990, Hauser et al. 1996, 2000, Sulkowski and Hauser 2001) and the ability to establish an ordered number list (Brannon and Terrace 1998, Terrace et al. 2003, Matsuzawa 2009). Even the Mutual Exclusivity constraint, which ensures the one-to-one mapping between numerals and the quantities that they denote, may not be specific to language either and is possibly present in non-human species (Markman 1992, Kaminski et al. 2004). Therefore, once language became available as a biological capacity, numerical capacity also became available even though it presumably took thousands of years of human social and cultural evolution to be activated (Carey 2009). Despite having the evolutionary prerequisites for numbers, non-human species never develop an infinite number system precisely due to the absence of a linguistic system. The current proposal takes the shared origin of language and numbers a step further: the mechanisms for number acquisition are exactly the same mechanisms for language acquisition. Viewed in this light, the numerical difference between a Chinese and a Piraha speaking child is no more remarkable than the syntactic difference between an English and German speaking child.

References

Baerman, M. and Corbett, G. G. (2010). Defectiveness: Typology and diachrony. In Baerman, M., Corbett, G. G., and Brown, D., editors, *Defective paradigms: Missing forms and what they tell us*, pages 19–34. Oxford University Press, Oxford.

Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2–3):150–177.

Bloom, P. and Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of child language*, 24(03):511–533.

Brannon, E. M. and Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science*, 282(5389):746–749.

Brown, R. (1973). *A first language: The early stages*. Harvard University Press, Cambridge, MA.

Brown, R. W. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, 55(1):1–5.

Butterworth, B., Reeve, R., Reynolds, F., and Lloyd, D. (2008). Numerical thought with and without words: Evidence from indigenous australian children. *Proceedings of the National Academy of Sciences*, 105(35):13179–13184.

Carey, S. E. (2009). *The origin of concepts*. Oxford University Press, New York.

Cheung, P., Rubenson, M., and Barner, D. (2017). To infinity and beyond: Coutning procedures precede learning the logic of the natural numbers. *Cognitive Psychology*, 92:22–36.

Chomsky, N. (1980). *Rules and representations*. Columbia University Presss, New York.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In MacWhinney, B., editor, *Mechanisms of language acquisition*, pages 1–33. Erlbaum, Hillsdale, NJ.

Condry, K. F. and Spelke, E. S. (2008). The development of language and abstract concepts: the case of natural number. *Journal of Experimental Psychology: General*, 137(1):22.

Davidson, K., Eng, K., and Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, 123(1):162–173.

Dehaene, S. and Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, 43(1):1–29.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, 284(5416):970–974.

Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistic influences on mathematical development: How important is the transparency of the counting system? *Philosophical Psychology*, 21(4):523–538.

Everett, D. (2005). Cultural constraints on grammar and cognition in pirahã. *Current anthropology*, 46(4):621–646.

Feigenson, L. and Carey, S. (2003). Tracking individuals via object-files: evidence from infants' manual search. *Developmental Science*, 6(5):568–584.

Fodor, J. A. (1975). *The language of thought*, volume 5. Harvard University Press.

Freyd, P. and Baron, J. (1982). Individual differences in acquisition of derivational morphology. *Journal of Verbal Learning and Verbal Behavior*, 21(3):282–295.

Fuson, K. C. (1988). *Children's counting and concepts of number.* Springer, Berlin.

Fuson, K. C., Richards, J., and Briars, D. J. (1982). The acquisition and elaboration of the number word sequence. In *Children's logical and mathematical cognition*, pages 33–92. Springer.

Gallistel, C. R. (1990). *The organization of learning.* The MIT Press.

Gelman, R. (1993). A rational-constructivist account of early learning about numbers and objects. *Learning and motivation*, 30:61–96.

Gelman, R. and Butterworth, B. (2005). Number and language: how are they related? *Trends in cognitive sciences*, 9(1):6–10.

Gelman, R. and Gallistel, C. R. (1978). *The child's understanding of number.* Harvard University Press.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1):3–55.

Goldin-Meadow, S. and Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, 391(6664):279–281.

Goldin-Meadow, S. and Yang, C. (2017). Statistical evidence that a child can create a combinatorial linguistic system without external linguistic input: Implications for language evolution. *Neuroscience and Biobehavioral Reviews*, 81(Part B):150 – 157.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., and Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14(01):23–45.

Gordon, P. (2004). Numerical cognition without words: Evidence from amazonia. *Science*, 306(5695):496–499.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1):B23–B34.

Halle, M. (1973). Prolegomena to a theory of word formation. *Linguistic Inquiry*, 4(1):3–16.

Hauser, M. D., Carey, S., and Hauser, L. B. (2000). Spontaneous number representation in semi–free–ranging rhesus monkeys. *Proceedings of the Royal Society of London B: Biological Sciences*, 267(1445):829–833.

Hauser, M. D., MacNeilage, P., and Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Sciences*, 93(4):1514–1517.

Hurford, J. R. (1975). *The linguistic theory of numerals.* Cambridge University Press, Cambridge.

Hurford, J. R. (1987). *Language and number: The emergence of a cognitive system.* Blackwell, Oxford.

Izard, V., Streri, A., and Spelke, E. S. (2014). Toward exact number: Young children use one-to-one correspondence to measure set identity but not numerical equality. *Cognitive psychology*, 72:27–53.

Kaminski, J., Call, J., and Fischer, J. (2004). Word learning in a domestic dog: evidence for" fast mapping". *Science*, 304(5677):1682–1683.

Le Corre, M. (2014). Children acquire the later-greater principle after the cardinal principle. *British Journal of Developmental Psychology*, 32(2):163–177.

Le Corre, M. and Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105(2):395–438.

Le Corre, M., Van de Walle, G., Brannon, E. M., and Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive psychology*, 52(2):130–169.

Lefevre, J.-A., Clarke, T., and Stringer, A. P. (2002). Influences of language and parental involvement on the development of counting skills: Comparisons of french-and english-speaking canadian children. *Early Child Development and Care*, 172(3):283–300.

Legate, J. A. (2011). Warlpiri wh-scope marking. *Syntax*, 14(2):97–121.

Legate, J. A. and Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition*, 14(3):315–344.

Leslie, A. M., Gelman, R., and Gallistel, C. (2008). The generative basis of natural number concepts. *Trends in cognitive sciences*, 12(6):213–218.

Leybaert, J. and Cutsem, M.-N. V. (2002). Counting in sign language. *Journal of Experimental Child Psychology*, 81(4):482 – 501.

Lipton, J. S. and Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development*, 76(5):978–988.

Lipton, J. S. and Spelke, E. S. (2006). Preschool children master the logic of number word meanings. *Cognition*, 98(3):B57–B66.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Lawrence Erlbaum, Mahwah, NJ, 3rd edition.

Maratsos, M. (2000). More overregularizations after all: New data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen and Xu. *Journal of Child Language*, 27:183–212.

Marcus, G., Pinker, S., Ullman, M. T., Hollander, M., Rosen, J., and Xu, F. (1992). *Overregularization in language acquisition*. Monographs of the Society for Research in Child Development. University of Chicago Press, Chicago.

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In Gunnar, M. R. and Maratsos, M., editors, *Modularity and Constraints in Language and Cognition: The Minnesota Symposium on child psychology*, pages 59–102. Lawrence Erlbaum Associates, Inc.

Markman, E. M. and Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, 20(2):121–157.

Matsuzawa, T. (2009). Symbolic representation of number in chimpanzees. *Current opinion in neurobiology*, 19(1):92–98.

Miller, K. F., Kelly, M., and Zhou, X. (2005). Learning mathematics in China and the United States. In Campbell, J. I. D., editor, *Handbook of Mathematical Cognition*, pages 163–178. Psychology Press.

Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science*, 6(1):56–60.

Miller, K. F. and Stigler, J. W. (1987). Counting in chinese: Cultural variation in a basic cognitive skill. *Cognitive Development*, 2(3):279–305.

Miura, I. T., Kim, C. C., Chang, C.-m., and Okamoto, Y. (1988). Effects of language characteristics on children's cognitive representation of number: Cross-national comparisons. *Child Development*, pages 1445–1450.

Negen, J. and Sarnecka, B. W. (2012). Number-concept acquisition and general vocabulary development. *Child development*, 83(6):2019–2027.

Nevins, A., Pesetsky, D., and Rodrigues, C. (2009). Pirahã exceptionality: A reassessment. *Language*, 85(2):355–404.

Nunes, T. and Bryant, P. (1996). *Children doing mathematics*. Wiley-Blackwell.

Phillips, C. (2010). Syntax at age two: Cross-linguistic differences 1. *Language Acquisition*, 17(1-2):70–120.

Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science*, 306(5695):499–503.

Pierce, A. (1992). *Language acquisition and syntactic theory: A comparative analysis of French and English*. Kluwer, Dordrecht.

Pinker, S. (1995). Why the child holded the baby rabbit: A case study in language acquisition. In Gleitman, L. R. and Liberman, M., editors, *An invitation to cognitive science, Vol. 1: Language*, pages 107–133. MIT Press, Cambridge, MA.

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73:265–292.

Roeper, T. W. (2011). The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics*, 5(1-2):057–086.

Russell, B. (1919). *Introduction to mathematical philosophy*. Allen and Unwin, London.

Sarnecka, B. W. and Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, 108(3):662–674.

Sarnecka, B. W. and Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of experimental child psychology*, 103(3):325–337.

Schuler, K., Yang, C., and Newport, E. (2016). Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. In *The 38th Cognitive Society Annual Meeting*, Philadelphia, PA.

Shi, R. and Melançon, A. (2010). Syntactic categorization in French-learning infants. *Infancy*, 15(5):517–533.

Song, M.-J. and Ginsburg, H. P. (1987). The development of informal and formal mathematical thinking in korean and us children. *Child Development*, pages 1286–1296.

Spaepen, E., Coppola, M., Spelke, E. S., Carey, S. E., and Goldin-Meadow, S. (2011). Number without a language model. *Proceedings of the National Academy of Sciences*, 108(8):3163–3168.

Spelke, E. S. (2003). What makes us smart? core knowledge and natural language. *Language in mind: Advances in the study of language and thought*, pages 277–311.

Spelke, E. S. (2016). Core knowledge, language, and number. *Language Learning and Development*, (To appear).

Sulkowski, G. M. and Hauser, M. D. (2001). Can rhesus monkeys spontaneously subtract? *Cognition*, 79(3):239–262.

Terrace, H. S., Son, L. K., and Brannon, E. M. (2003). Serial expertise of rhesus macaques. *Psychological Science*, 14(1):66–73.

Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology*, 22(4):562.

Wagner, K., Kimura, K., Cheung, P., and Barner, D. (2015). Why is number word learning hard? evidence from bilingual learners. *Cognitive Psychology*, 83:1 – 21.

Wexler, K. (1994). Optional infinitives, verb movement and the economy of derivation in child grammar. In Lightfoot, D. and Hornstein, N., editors, *Verb movement*, pages 305–350. Cambridge University Press, Cambridge.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36(2):155–193.

Wynn, K. (1992a). Addition and subtraction by human infants. *Nature*, 358(6349):749–750.

Wynn, K. (1992b). Children's acquisition of the number words and the counting system. *Cognitive psychology*, 24(2):220–251.

Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, 110(16):6324–6327

Yang, C. (2016). *The price of linguistic productivity: How children learn to break rules of language.* MIT Press, Cambridge, MA.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology.* Addison-Wesley, Cambridge, MA.