

## **UC Merced**

### **Proceedings of the Annual Meeting of the Cognitive Science Society**

#### **Title**

Explaining the Number Hierarchy

#### **Permalink**

<https://escholarship.org/uc/item/5917s4b8>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

#### **Authors**

Malouf, Robert

Ackerman, Farrell

Seyfarth, Scott

#### **Publication Date**

2015

Peer reviewed

# Explaining the Number Hierarchy

Robert Malouf (rmalouf@mail.sdsu.edu)

Department of Linguistics and Asian/Middle Eastern Languages, San Diego State University

Farrell Ackerman (fackerman@ucsd.edu)

Scott Seyfarth (sseyfarth@ucsd.edu)

Department of Linguistics, University of California, San Diego

## Abstract

Greenberg's (1963) Universal 34 states that "No language has a trial number unless it has a dual. No language has a dual unless it has a plural." We present an associative model of the acquisition of grammatical number based on the Rescorla-Wagner learning theory (Rescorla & Wagner, 1972) that predicts this generalization. Number as a real-world category is inherently structured: higher numerosity sets are mentioned less frequently than lower numerosity sets, and higher numerosity sets always contain lower numerosity sets. Using simulations, we demonstrate that these facts, along with general principles of probabilistic learning, lead to the emergence of Greenberg's Number Hierarchy.

**Keywords:** Linguistics; Language acquisition; Learning; Computer simulation

## Introduction

In many languages, the number of items in the referent of a noun phrase is obligatorily encoded by an inflectional feature on the head noun or other lexical category in a clause. For example, in English we have distinction between *the book*, which must refer to a single book, and *the books*, which must refer to more than one.

While many languages show the same singular vs. plural distinction that English does, this is not the only attested system. Another fairly common type of language distinguishes between the SINGULAR (exactly one), the DUAL (exactly two), and the PLURAL (more than two). For example, in Upper Sorbian (a Slavic language spoken in Germany), we find singulars like *hród* 'castle' and *džětam* '(I) work', duals like *hrodaj* 'two castles' and *džětamoj* '(we two) work', and plurals like *hrody* 'castles' and *džětamy* '(we) work' (Stone, 1993; Corbett, 2000). In Hmong Daw (a Hmong-Mien language of Laos and southern China), personal pronouns distinguish three persons and three grammatical numbers (Jaisser, 1995, 118):

	SING	DUAL	PLURAL
1ST	<i>kuv</i>	<i>wb</i>	<i>peb</i>
2ND	<i>koj</i>	<i>neb</i>	<i>nej</i>
3RD	<i>nws</i>	<i>nkawd</i>	<i>lawv</i>

Beyond the singular and dual, some languages even distinguish a TRIAL (= exactly three) number. This four-way distinction is found, for example, in the subject agreement prefixes for human referents in Larike, a Central Malayo-Polynesian language spoken on Ambon Island (Laidig, 1993; Siewierska, 2013):

	SING	DUAL	TRIAL	PLURAL
1.INCL	—	<i>itua-</i>	<i>itidi-</i>	<i>ite-</i>
1.EXCL	<i>au-</i>	<i>arua-</i>	<i>aridu-</i>	<i>ami-</i>
2	<i>a-/a-</i>	<i>irua-</i>	<i>iridu-</i>	<i>imi-</i>
3.HUM	<i>ma-/mei-</i>	<i>matua-</i>	<i>matidu-</i>	<i>mati-</i>
3.NONHUM	<i>i-</i>			<i>iri-</i>

While cross-linguistic typological studies have revealed that there is a certain amount of variation among the grammatical number systems of diverse languages, that variation has been found to fall within fairly strict limits. Greenberg's (1963) Universal 34 states that "No language has a trial number unless it has a dual. No language has a dual unless it has a plural." This Number Hierarchy can be expressed as a chain of implicational relations: Trial → Dual → Plural → Singular. Greenberg's typological studies over a small corpus of 34 languages encouraged him to postulate numerous universal generalizations, both as absolutes and as tendencies. In succeeding years many of his preliminary claims have had to be modified or abandoned. The Number Hierarchy, however, has proved to be remarkably robust.<sup>1</sup>

Many of the unattested language types in this domain are silly: no language distinguishes between a prime vs. a composite number of referents, or has a special suffix indicating exactly 47 items. The non-existence of such languages does not demand an explanation. However, some of the language types that we do not find seem a priori more likely. For example, while the two-way SING/NOT-SING (i.e., plural) distinction is very common, the superficially similar DUAL/NOT-DUAL is essentially unattested. Along the same lines, while dual marking is fairly common, trial is rare, quadral (= exactly four) is only marginally attested, and markers for specific numbers greater than four are never found (at least in spoken rather than signed languages). It is less obvious why no known languages use these conceptually plausible number marking systems.

## Universals in linguistic theory

One of the fundamental goals of linguistic theory is to offer explanations for why certain patterns recur cross-linguistically and others do not. Debate around the status of and explanations for language universals, recently highlighted by Evans and Levinson's (2009) provocative argument against innatist language-specific representational bases

<sup>1</sup>It is not quite exceptionless, though, and the actual typological situation is somewhat more complicated (Croft, 1990; Corbett, 2000; Evans, 2012).

for universals, has been focused on two alternative models.

One hypothesis regarding the source of implicational universals like Greenberg’s Universal 34 is that they follow from universal and innate properties of the human language faculty. These properties are specific to language and do not follow from any more general properties of human cognition or culture. By this view, universals like the Number Hierarchy reveal something about the underlying structure of human language and constraints on possible surface organization (Noyer, 1992; Harbour, 2014). Specifically, under this view Universal Grammar provides a range of possible mental representations for grammatical number systems which are able to capture attested systems but not the unattested ones.

As an alternative, one could attempt to derive universals like the Number Hierarchy from language usage or general complexity properties (e.g., Croft & Poole, 2008; Miestamo, 2009). This approach has proven to be quite fruitful for a range of putative universals. But so far, the Number Hierarchy has resisted language-external explanation. For example, it is not immediately obvious how a Singular-Plural system is any simpler than a Dual-NotDual system, yet the first is the most common type and the second is (virtually) unattested (but see Evans, 2012). In fact, Seuren (2013) offers the Number Hierarchy as a best example of a universal which is very unlikely to submit to an explanation that does not depend on language-specific properties of Universal Grammar. This is not an argument against grounding the Number Hierarchy in properties of general cognition, merely an observation that no convincing explanation along those lines has yet been proposed. In what follows we develop a learning model sensitive to realistic recurrences of set size and plausible assumptions about the representation of numerosities.

## Model

Children learn the concept of numerosity before they learn morphosyntactic expressions of number (Barner, Thalwitz, Wood, & Carey, 2007; Wood, Kouider, & Carey, 2009; Clark & Nikitina, 2009) and well before they master number name meanings (Slusser & Sarnecka, 2011). Learning the distinction between singular and plural sets does not appear to be dependent on morphological marking (Li, Ogura, Barner, Yang, & S.Carey, 2009).

Clark and Nikitina (2009) consider the use of *two* as a general purpose plural marker by English-speaking two- and three-year-old children. They argue that this arises at least in part because *two* is the numeral larger than one that is used most often by adults in child-directed speech. Furthermore, their observations indicate that the trajectories of individual children in learning grammatical number can vary quite a bit and are sensitive to the frequency of forms in adult speech. This frequency-sensitivity suggests that a general associative-learning process is playing an important role in the development of grammatical number.

Ramscar, Dye, Popick, and O’Donnell-McCarthy (2011) developed a model of children’s acquisition of number names

(*one, two, three, etc.*) based on the Rescorla-Wagner theory of associative learning, which as a side effect predicts the subitizing limit (Kaufman, Lord, Reese, & Volkman, 1949), a constraint on the human ability to recognize the number of items in a set without explicit counting.

## Rescorla-Wagner learning

Rescorla and Wagner’s (1972) learning model, rooted in Pavlovian learning theory, seeks to explain the way that associative learning gradually builds connections between perceptual cues and specific outcomes over the course of many learning trials. Early work in this direction (Hull, 1943) showed that associative learning follows a negatively accelerated learning curve. R-W’s model extends this approach in a way that can deal with compound cues  $AX$ . When the cues  $A$  and  $X$  both occur with  $O$  on a learning trial, the update for the association weights is given by:

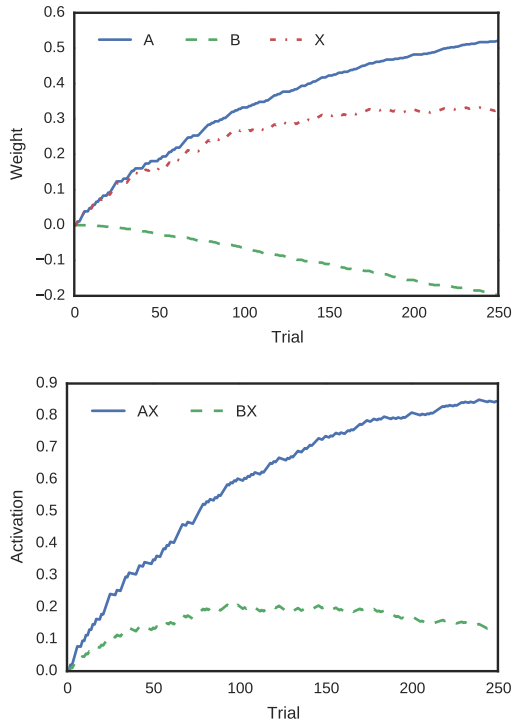
$$\begin{aligned} V_{AX} &= V_A + V_X \\ \Delta V_A &= \alpha_A \beta (\lambda - V_{AX}) \\ \Delta V_X &= \alpha_X \beta (\lambda - V_{AX}) \end{aligned}$$

$\Delta V_A$  is the change in the strength of the association between cue  $A$  and the outcome  $O$  after a learning trial in which  $A$  and  $O$  occur together,  $V_A$  is the previous weight of the association prior to the current trial, and  $\lambda$  is the maximum conditioned response. The learning rate depends on the salience of the cue ( $\alpha_A$  and  $\alpha_X$ ) and the salience of  $O$  ( $\beta$ ), and the learning step size depends on the previous association weight of the compound  $V_{AX}$ .

One consequence of this model is that a cue which always occurs when  $O$  does may not end up with a strong association if some other cue is a better predictor of the outcome. Suppose the compound cue  $AX$  consistently occurs with  $O$ , while the cue  $BX$  consistently does not. We expect the sum  $V_{AX}$  to approach  $\lambda$  and  $V_{BX}$  to approach 0, and as shown in Figure 1 that is what happens. Looking at the total strength of the prediction in the lower graph,  $V_{BX}$  initially grows with  $V_{AX}$ , then begins to fall towards 0. Looking at the individual cue weights in the upper graph, we see that initially  $A$  and  $X$  have roughly equal weights and that  $X$  is competing with  $A$  and  $B$  to explain the (non-)occurrence of  $O$ . After about 50 trials, though,  $A$  wins the competition, and the learner has been able to discriminate among the cues that are present with  $O$  to find the ones which are the best predictors (Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010; Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011). In this model, learning is driven by prediction error:  $V_{AX}$  is the learner’s expectation prior to a trial, and  $(\lambda - V_{AX})$  or  $(0 - V_{AX})$  is how much that expectation is violated on a positive or negative learning trial.

While it is not without problems, the R-W model has been enormously influential in the development of animal associative learning models. R-W learning has also been successfully applied to model human causal reasoning (Lober & Shanks, 2000; Danks, 2007), is closely related to both connectionist (Gluck & Bower, 1988; Shanks, 1991) and

Figure 1: Discriminative learning: when the (hypothetical) compound cue *AX* occurs with *O* but the cue *BX* does not, the model learns that *A* and not *X* is a predictor of *O*



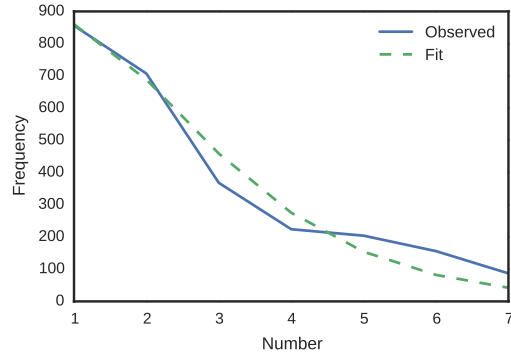
information theoretic (Gallistel, 2002) models, and is formally equivalent to the perceptron (Dawson, 2008). In the domain of language, R-W learning has been applied to a range of problems, including modeling acquisition of plural forms (Ramscar & Yarlett, 2007; Ramscar, Dye, & McCauley, 2013) and number names (Ramscar et al., 2011), and word recognition (Baayen et al., 2011).

### Input model

In order to construct a simulation of grammatical number acquisition based on R-W learning theory, we first need a representation of the input cues that a learner would be exposed to. The ability to subitize, or recognize the numerosity of small sets, is either innate or developed very early and can be taken as a semantic primitive (Wynn, 1992; Piantadosi, Tenenbaum, & Goodman, 2012). Following Ramscar et al. (2011), we include numerosity of the set and all subsets as cues. For example, a learner who encounters a set of one item is only exposed to the cue ‘1’—a single-item set.<sup>2</sup> A learner who encounters a set of two items is exposed to the cue ‘2’—a double-item set—and also to the cue ‘1’, since encountering a double-item set entails encountering a single-item (sub)set. A learner who encounters a set of three items is exposed to triple-item, double-item, and single-item sets,

<sup>2</sup>The learner is also exposed to the cues associated with the item’s other properties such as shape and size, which are ignored here for simplicity.

Figure 2: Frequency of number mentions in COCA (Davies, 2008–; Ramscar et al., 2011) and fitted values



etc. The learner’s task is to determine which sets (the cues) best predict the use of each grammatical number marker (the outcomes).

R-W learning is also potentially very sensitive to statistical properties of the input. Thus, we also need an accurate probabilistic model of the frequencies at which learners would be exposed to various inputs. Specifically, we need to know how often children encounter talk about sets of a given numerosity while learning grammatical number.

In their simulation of number name acquisition, Ramscar et al. (2011) report counts of number names from 1 to 7 used as prenominal modifiers in the Corpus of Contemporary American English (COCA) and the Corpus del Español. The distribution of English counts is shown in Figure 2. Using the VGAM package in R (Yee, 2010), we found that a zero-truncated negative binomial distribution (a mixture of Poissons with no zero counts and gamma-distributed rates to account for overdispersion) provided a good fit to the counts in both languages. Therefore, we generated random inputs for our simulations by sampling from a zero-truncated negative binomial fit to the English counts (size= 3, prob=0.6).

### Results

For the first set of simulations, we ran 250 iterations of R-W learning with randomly generated inputs as the cues and the correct grammatical number for each input as the outcome.<sup>3</sup> The trajectories for singular and plural outcomes are shown in Figure 3. Learning what the singular means is easy: 1 is the only cue that has a positive association with the outcome SG. Plural is a bit more difficult, though: a single-item subset is always present in sets of higher numerosity. The learner therefore always encounters a cue of 1 when they encounter any set and any number-marker outcome. But, after about 50 iterations, the model has succeeded at discriminating between 2 and 1 and has identified the former as a good predictor of the outcome PL. The total activations in the lower graph show that the model is correctly distinguishing singular and plural sets after about 150 iterations.

<sup>3</sup>Source code for simulation of R-W learning is available at <http://github.com/rmalouf/learning>.

Figure 3: Simulated learning of a **singular/plural** distinction

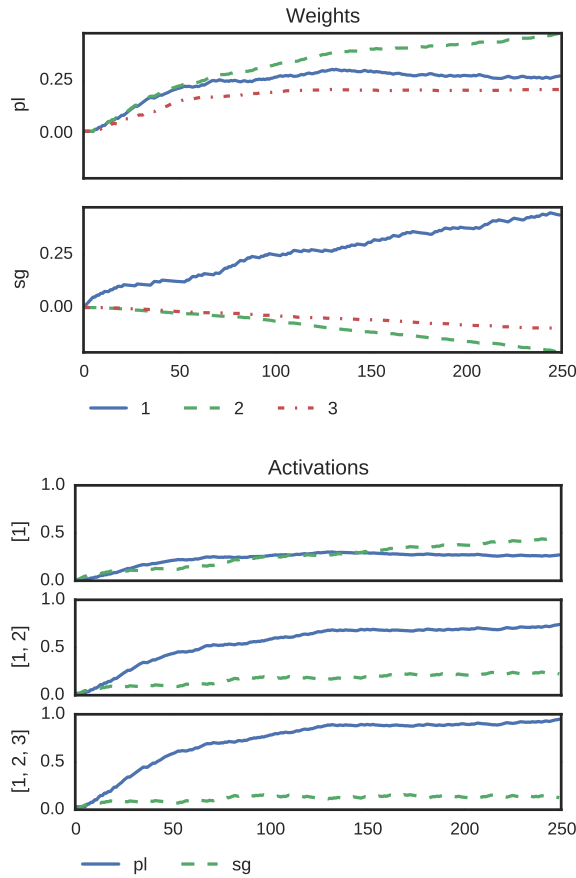
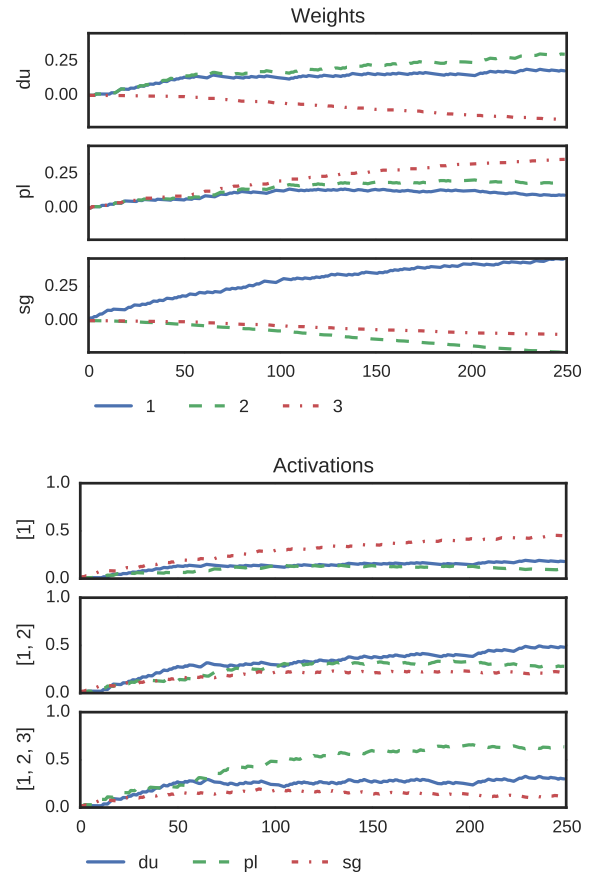


Figure 4: Simulated learning of **singular/dual/plural**



We repeated the simulation for a singular/dual/plural system, in Figure 4. Learning proceeds in much the same way as in the previous case, though in this system there are two discrimination tasks to be solved: the learner must learn that the cue 1 is not associated with DU nor is it associated with PL. Additionally, it must learn that 2 is not associated with PL, even though all plural sets include 2 as a member. As before, the learner is able to correctly label singular, dual, and plural sets after about 150 iterations.

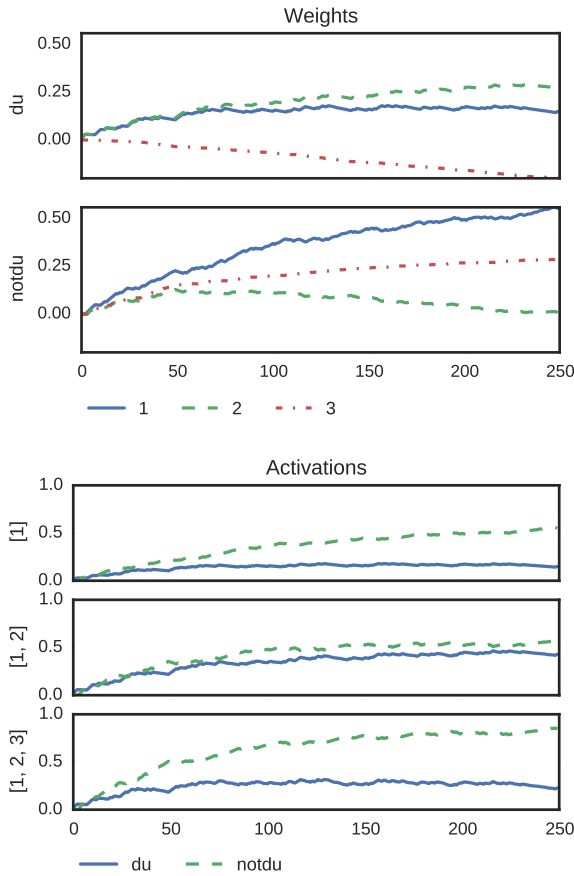
The next simulation considers the task of learning a dual/non-dual distinction, a grammatical number system type that is not found among human languages. The results are in Figure 5. Parallel to the singular/plural system, in this system the model needs to learn that 2 is a good cue for the outcome DU but 1 is not, even though both cue 1 and cue 2 are always present with outcome DU. Unlike the singular/plural system, however, the model must also cope with the fact that 2 sometimes occurs with NONDU, for tripleton and higher sets. Furthermore, the model needs to learn that 3 is a good cue for NONDU. This is challenging, because 1 quickly becomes a strong cue for NONDU in this system, and 3 never occurs without 1.

After 250 iterations, the model has only begun to discriminate 1 and 2 (in the topmost graph), and the activation for

DU never does rise above NONDU for sets of numerosity 2. Note that this is not to say that the dual/non-dual distinction is unlearnable in general, only that it was not learned in 250 iterations. This suggests that some measure of relative degree of learnability is a determinant for the observed distributions of number marking.

The results in Figures 3–5 are traces of individual simulations. The actual learning trajectory and the final weights in the run depend on the particular randomly-generated training examples that are provided in each learning trial. To see how learning progresses in a population, we performed another set of simulations. In these experiments, we combined 100 learners, each following its own trajectory. After each learning trial, we calculated the fraction of the population which had mastered the appropriate grammatical number system: e.g, for the singular/dual/plural system, a ‘correct’ learner would have to assign the highest total activation to SG for a singleton set, DU for a set with two items, and PL for higher numerosities. As shown in Figure 6, the singular/plural and singular/dual/plural systems are learned quickly and reliably by all members of the population. The same is not true for the dual/non-dual system. Some members of the population learn it quickly, but after 250 iterations less than 40% of the population can use the system correctly. The dual/non-dual system

Figure 5: Simulated learning of a **non-dual/dual** distinction



offers a much more difficult discrimination task to the learner than do either the singular/plural or the singular/dual/plural systems.

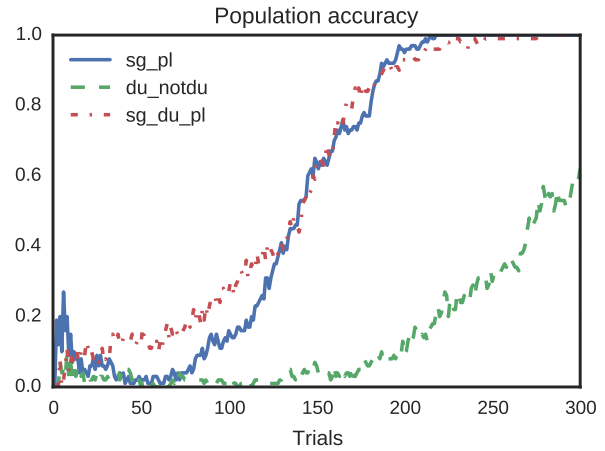
To test what features of the system lead to this contrast in learning difficulty, we performed two more sets of population-based simulations. In one, we removed subset cues from the inputs, including only the numerosity of the set as a whole. Given this input representation, the singular/plural distinction is learned very quickly, but both the singular/dual/plural and the dual/non-dual systems are learned very slowly (and at similar rates).

In the second set of simulations, we altered the input probability distribution to make sets of numerosity two more frequent than singletons (in contradiction to the real observed frequencies in Figure 2). In this case, all three systems are learned easily and quickly, and there is no distinction between the attested and unattested systems.

## Discussion

In general, these results suggest that the implicational organization of number distinctions identified in Universal 34 may emerge from the interaction of a reliably common and skewed distribution of set sizes with particular assumptions concerning the representation of numerosities. The simulations repli-

Figure 6: Fraction of a simulated population of learners that have mastered each system after  $k$  trials



cate the expectations based on attested grammatical marking strategies: SG versus PL is learned both early and easily, while further distinguishing the DU takes longer, but is attainable. In contrast, the differentiation of DU from NON-DU proves a difficult discrimination to make. Given the success of the simulations in achieving empirically attested patterns without positing language specific representations, it was crucial to explore the contribution of two basic assumptions of the model, namely, our representation of numerosity and our assumption concerning the role of exposure to the frequencies of particular numerosities. When we expunged subset cues from the representations, singular/dual/plural patterned with dual/non-dual. This suggests the facilitating role of subset organization for the emergence of singular/plural and singular/dual/plural versus dual/non-dual. When we altered the frequencies for experienced numerosities so that duals occurred more often than singulars, the difference between all three patterns of number organization, i.e., singular/plural, singular/dual/plural, and dual/non-dual disappeared: all were learned equivalently well. This suggests that, as hypothesized, real world experience has an important shaping influence on the organization of grammatical number systems.

While our model explains aspects of the intriguing restrictions on the morphosyntax of cross-linguistic number marking, it only defines the *broad* constraints on number organization in natural language. It ignores the specific strategies that particular languages employ within the space of options permitted by Universal 34. Further research is required to understand the dynamics which motivate, sustain and alter the system of discriminations observed in specific languages. It is in this interdependence between general constraints on the organization of linguistic phenomena and the particularities of their encodings that an understanding of natural language is most likely to be found.

## Conclusion

In this paper, we present a view of grammatical number learning based on associative learning (Rescorla & Wagner, 1972; Ramskar & Yarlett, 2007; Baayen et al., 2011). We show that this model also provides a possible language-external explanation for aspects of the Number Hierarchy. In particular, number as a real-world category is inherently structured in two ways: sets of higher numerosity are mentioned less frequently than sets of lower numerosity, and sets of higher numerosity always contain sets of lower numerosity. Using corpus-based computational simulations (Ramskar et al., 2011), we demonstrate that these facts, in interaction with general principles of probabilistic learning, plausibly lead to languages which violate the Number Hierarchy being much more difficult to learn than languages which follow it, which in turn motivates the emergence of the Number Hierarchy as an implicational universal. This shows that even fairly abstract properties of grammatical systems, when viewed from a developmental perspective, can be seen to have a physical or cognitive origin external to language.

## Acknowledgments

We thank David Barner, Michael Ramskar, and the participants at the 2014 workshop on “Corpus resources for quantitative and psycholinguistic analysis” held at Eszterházy College, Eger, Hungary for helpful criticisms.

## References

- Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*, 438–482.
- Barner, D., Thalwitz, D., Wood, J., & Carey, S. (2007). On the relation between the acquisition of singular-plural morpho-syntax and the conceptual distinction between one and more than one. *Developmental Science*, *10*(3), 365–373.
- Clark, E. V., & Nikitina, T. V. (2009). One vs. more than one: Antecedents to plural marking in early language acquisition. *Linguistics*, *47*(1), 103–139.
- Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press.
- Croft, W. (1990). *Typology and universals*. Cambridge: Cambridge University Press.
- Croft, W., & Poole, K. T. (2008). Inferring universals from grammatical variation: multidimensional scaling for typological analysis. *Theoretical Linguistics*, *34*, 1–37.
- Danks, D. (2007). Causal learning from observations and manipulations. In *Thinking with data* (pp. 359–388). New York: Lawrence Erlbaum Associates.
- Davies, M. (2008–). *The corpus of contemporary american english: 450 million words, 1990-present*. (Available online at <http://corpus.byu.edu/coca/>)
- Dawson, M. R. W. (2008). Connectionism and classical conditioning. *Comparative Cognition and Behavior Reviews*, *3*, 1–115.
- Evans, N. (2012). Even more diverse than we had thought: The multiplicity of Trans-Fly languages. In N. Evans & M. Klamer (Eds.), *Melanesian languages on the edge of Asia: Challenges for the 21st century*. University of Hawai'i Press.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–492.
- Gallistel, C. R. (2002). Frequency, contingency and the information processing theory of conditioning. In P. Sedlmeier & T. Betsch (Eds.), *Etc. Frequency processing and cognition* (pp. 153–171). Oxford: Oxford University Press.
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*(3), 227–247.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of grammar* (pp. 73–113). Cambridge: MIT Press.
- Harbour, D. (2014). Paucity, abundance, and the theory of number. *Language*, *90*(1), 185–229.
- Hull, C. L. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.
- Jaisser, A. (1995). *Hmong for beginners*. Berkeley: Center for South and Southeast Asia Studies, University of California at Berkeley.
- Kaufman, E., Lord, M., Reese, T., & Volkman, J. (1949). The discrimination of visual number. *American Journal of Psychology*, *62*(4), 498–525.
- Laidig, W. D. (1993). Insights from Lariké possessive constructions. *Oceanic Linguistics*, *32*, 312–351. doi: 10.2307/3623197
- Li, P., Ogura, T., Barner, D., Yang, S.-Y., & S. Carey. (2009). Does the conceptual distinction between singular and plural sets depend on language? *Developmental Psychology*, *45*(6), 1644–1653.
- Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review*, *107*, 195–212.
- Miestamo, M. (2009). Implicational hierarchies and grammatical complexity. In G. B. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 80–97). Oxford: Oxford University Press.
- Noyer, R. (1992). *Features, positions and affixes in autonomous morphological structure* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
- Piantadosi, S., Tenenbaum, J., & Goodman, N. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, *123*, 199–217.
- Ramskar, M., Dye, M., & McCauley, S. (2013). Error and expectation in language learning: The curious absence of ‘mouses’ in adult speech. *Language*, *84*(4), 329–347.
- Ramskar, M., Dye, M., Popick, H. M., & O'Donnell-McCarthy, F. (2011). The enigma of number: Why children find the meanings of even small number words hard to learn and how we can help them do better. *PLoS One*, *6*(7), e22501.
- Ramskar, M., & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science*, *31*, 927–960.
- Ramskar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*(6), 909–957.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning II* (p. 62–99). Appleton-Century-Crofts.
- Seuren, P. A. M. (2013). *From Whorf to Montague: Explorations in the theory of language*. Oxford: Oxford University Press.
- Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 433–443.
- Siewierska, A. (2013). Third person zero of verbal person marking. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/103>
- Slusser, E. B., & Sarnecka, B. W. (2011). Find the picture of eight turtles: A link between children's counting and their knowledge of number word semantics. *Journal of Experimental Child Psychology*, *110*(1), 38–51.
- Stone, G. (1993). Sorbian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic languages* (p. 593–685). London: Routledge.
- Wood, J., Kouider, S., & Carey, S. (2009). Acquisition of singular-plural morphology. *Developmental Psychology*, *45*(1), 202–206.
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*, 220–251.
- Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, *32*(10), 1–34. Retrieved from <http://www.jstatsoft.org/v32/i10/>