To appear in *Journal of Experimental Psychology: General*

Title

The use of number words in natural language obeys Weber's law

Authors

Luca Rinaldi [a,b,1] and Marco Marelli [a,b,1]

[a] *Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy*

[b] *NeuroMI, Milan Center for Neuroscience, Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy*

[1] To whom correspondence may be addressed. E-mail: luca.rinaldi@unimib.it (L.R.) or marco.marelli@unimib.it (M.M.)

Abstract

It has been suggested that the origins of number words can be traced back to an evolutionarily ancient approximate number system, which represents quantities on a compressed scale and complies with Weber's law. Here, we use a data-driven computational model, which learns to predict one event (a word in a text corpus) from associated events, to characterize verbal behavior relative to number words in natural language, without appeal to perception. We show that the way humans use number words in spontaneous language reliably depends on numerical ratio - a clear signature of Weber's law - thus perfectly mirroring the human and non-human psychophysical performance in comparative judgments of numbers. Most notably, the adherence to Weber's law is robustly replicated in a wide range of different languages. Together, these findings suggest that the everyday use of number words in language rests upon a pre-verbal approximate number system, which would thus affect the handling of numerical information not only at the input level but also at the level of verbal production.

Keywords

numerical cognition; language evolution; approximate number system; Weber's law; distributional models

Introduction

Humans throughout history have regularly created symbolic systems, like written numerals, for representing quantities in such a way that goes beyond the limits of perception (Menninger, 1969). The past few years have witnessed a fervent debate about how humans manage to do so (Gelman & Butterworth, 2005). Most theories of numerical cognition have proposed that an evolutionary ancient mechanism - the approximate number system (ANS) - is exploited to represent exact numerosities in language (Cantlon, Platt, & Brannon, 2009; Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004; Gallistel & Gelman, 1992). These accounts assume that the ANS is a pre-verbal system suited for processing all types of magnitude, whether they are physical or symbolic (Cantlon et al., 2009; Dehaene, 1997; Feigenson et al., 2004; Gallistel & Gelman, 1992). One of the hallmarks of the ANS, indeed, is that the ability to discriminate between two perceptual magnitudes depends on their ratio, a clear signature of Weber's law (Cantlon et al., 2009; Leibovich, Ashkenazi, Rubinsten, & Henik, 2013). This behavioral dependency is paralleled by neurophysiological evidence, as neuronal tuning in the human and macaque parietal regions also obeys Weber's law (Nieder & Dehaene, 2009). Weber's law states that the difference in intensity needed to discriminate two stimuli (also known as "just noticeable difference") is proportional to their objective intensities, or $\Delta I/I = K$, where ΔI represents the difference threshold, I is the initial stimulus intensity, and K is a constant that signifies the resulting sensitivity to changes in stimulus intensity. Somewhat surprisingly, Weber's law well describes not only comparisons of visual arrays of dots by humans, but also comparisons of number symbols (Moyer & Landauer, 1967). This experimentally established fact suggests that humans make unconscious reference to the corresponding ANS representations of quantity when making judgments of symbolic magnitudes (Butterworth, 1999; Dehaene, 1997; Feigenson et al., 2004; Moyer & Landauer, 1967). This possibility is also supported by brain imaging research showing that the activity of the human intraparietal sulcus is modulated to the same extent by the numerical ratio of both symbolic and non-symbolic numerosities (Cohen Kadosh, Lammertyn, & Izard, 2008; Fias, Lammertyn, Reynvoet, Dupont, & Orban, 2003; Holloway, Price, & Ansari, 2010; Kaufmann et al., 2005). The intraparietal region is indeed associated with an abstract, supramodal representation of numbers, hosting a core neural circuity that is systematically activated during numerical magnitude processing (Eger, Sterzer, Russ, Giraud, & Kleinschmidt, 2003; Nieder & Dehaene, 2009).

The idea that sophisticated non-symbolic numerical knowledge depends on mechanisms with a long phylogenetic history is substantiated by research on other animal species (Barnard et al., 2013; Cantlon & Brannon, 2006; Ditz & Nieder, 2016), preverbal infants (Libertus & Brannon, 2010; Starr, Libertus, & Brannon, 2013; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005) and people belonging to remote cultures with extremely limited or no numerical vocabulary (Gelman & Gallistel, 2004; Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004). One additional piece of evidence comes from human developmental research (Cordes & Gelman, 2005; Gallistel & Gelman, 1992). A non-verbal numerical reasoning system would support counting, addition, and subtraction by young children and infants, already at 6 months of age (Barth, La Mont, Lipton, & Spelke, 2005; Barth et al., 2006; Cordes & Gelman, 2005; Gelman, 2011). Further, various studies have shown that the ANS lays the cognitive foundation for symbolic mathematical skills in both children (Gilmore, McCarthy, & Spelke, 2010; Halberda, Mazzocco, & Feigenson, 2008; Mazzocco, Feigenson, & Halberda, 2011; Piazza, Pica, Izard, Spelke, & Dehaene, 2013) and adults (DeWind & Brannon, 2012; Park & Brannon, 2013). Finally, neuroimaging research has revealed that parietal areas respond similarly to numerosity information in both adults and children, thus prior to formal arithmetic education (Cantlon, Brannon, Carter, & Pelphrey, 2006; Temple & Posner, 1998).

Another source of evidence in favor of the ANS hypothesis derives from studies of numerical estimation. In estimation tasks, participants have to label the numerals describing a series of rapidly flashed dot arrays (Whalen, Gallistel, & Gelman, 1999). Crucially, the pattern of error in the participants' verbal estimates has been shown to comply with Weber's law, indicating that numerical estimation requires linking numerals to the ANS (Gallistel & Gelman, 1992; Izard & Dehaene, 2008; Whalen et al., 1999). Similar evidence has been reported in preschool children as young as 5 years of age (e.g., Huntley-Fenner, 2001; Lipton & Spelke, 2005). When taken together, this evidence suggests evolutionary and developmental continuity from a nonverbal analog magnitude system to a symbolic faculty for numbers (Diester & Nieder, 2010). However, a shortcoming of this theoretical assumption is that it faces the obvious difficulty of unveiling how approximate representations about magnitude in the ANS could provide the content of number words, which in turn seemingly have precise and absolute meaning in language.

It is worth noting that some parallel findings have started to cast doubt on the notion that numerical symbols would be grounded in the ANS. The mechanisms subserving the development of a symbolic faculty for numbers are, indeed, rather controversial and have been the subject of intense debate. Although some theories posit innate numerical resources (e.g., Gallistel & Gelman, 1992), not all converge on the idea that verbal numerals are grounded on a number module. Various objections to these proposals have been advanced, primarily advocating that there is no intrinsic link between the ANS and natural language (Laurence & Margolis, 2005; Margolis & Laurence, 2008). For instance, it has been proposed that linguistic quantifiers may bootstrap number acquisition (Barner, 2017; Bloom & Wynn, 1997; Carey, 2009). The idea is that the child learns the meaning of number words, such as “five”, through a bootstrapping process by integrating the verbal counting routine with numerical quantity (Carey, 2009; for a computationally-implemented model see Piantadosi, Tenenbaum, & Goodman, 2012). Accordingly, empirical developmental evidence that undermines the ANS hypothesis is available (Carey, Shusterman, Haward, & Distefano, 2017; Le Corre & Carey, 2007). These data suggest that ANS values are not readily amenable to labeling with words in young children (Sullivan & Barner, 2014).

In addition to this, behavioral studies have shown that symbolic and non-symbolic representations of numbers may be qualitatively different from one another (Krajcsi, 2017; Krajcsi, Lengyel, & Kojouharova, 2016; Lyons, Nuerk, & Ansari, 2015). Similarly, a complete overlap in regions subserving symbolic and non-symbolic numerical quantity processing has been questioned by recent neuroimaging evidence (Bulthé, De Smedt, & Op de Beeck, 2015; Damarla & Just, 2013;

Lyons, Ansari, & Beilock, 2015). In particular, by applying sophisticated pattern-classification algorithms to distributed (multi-voxel) patterns of functional MRI data, these studies showed only a partial commonality of neural patterns associated with symbolic and non-symbolic number processing, with a poor cross-format classification of numerical quantities (Bulthé et al., 2015; Damarla & Just, 2013). Further, while some studies documented a positive correlation between individual differences in tasks tapping into the ANS and symbolic math (Halberda et al., 2008), others did not report such a tight relationship (Holloway & Ansari, 2009).

Moving beyond this controversy related to the ANS, and taking a more general perspective, a complex interaction between biological and cultural factors has likely characterized the evolution of the symbolic faculty of numbers alone (Núñez, 2017). Number words are possibly one of the most paramount cultural tools for expressing numerosities. Different aspects of numerical cognition are indeed saturated with culture (Beller et al., 2018; Núñez, 2009; Núñez, 2017), with proprieties of numeration systems that vary across languages (Bender & Beller, 2012). One of these proprieties is certainly the extent of the system itself. Accordingly, some languages lack number words or contain only a limited set of numerals (Everett, 2017; Hammarström, 2010) and this may, in turn, affect performance in numerical tasks (Dehaene, Izard, Spelke, & Pica, 2008; Gordon, 2004; Pica et al., 2004). Similarly, differences across languages and cultures in how people use number words for counting (Chrisomalis, 2010) have an impact on number processing (Bender & Beller, 2012; Nickerson, 1988; Zhang & Norman, 1995). Further, effects of enculturation on number representation have been observed in the brain (Ansari, 2008). For instance, different language systems such as English and Chinese can shape the way our brain processes symbolic number information: whereas English speakers show more activation in classical language-related regions of the brain, Chinese speakers show selective activation of premotor areas (Tang et al., 2006). This may be related, for instance, to the experience of reading the more complex Chinese characters, as compared to the English ones (Tang et al., 2006). Together, this evidence points to a crucial mediation role played by cultural practices (Calude & Verkerk, 2016; Epps, Bowern, Hansen, Hill, & Zentz, 2012).

All the studies reviewed above are compatible with the idea that Weber's law would regulate the transformation of non-symbolic quantities presented at the perceptual level into internal analog representations, with some of these studies extending such a regulatory process to symbolic quantities. For instance, in performing a standard symbolic numerical comparison task in the written modality, the perceptual input (e.g., two number words) will be handled by a verbal word frame (e.g., Dehaene, 1992). This modality-dependent information would then be scaled, as described by Weber's law, into a corresponding approximate representation. Thus, Weber's law would describe a scaling from perceptual input to internal representations in the ANS (see Fig. 1A for a pictorial description of current theoretical models), with this compliance being possibly superimposed by the internal organization of cortical representations (Dehaene, 2003). Yet, if numerical symbols are grounded on such a pre-verbal ANS, as suggested by most theories of numerical cognition (Cantlon et al., 2009; Dehaene, 1997; Feigenson et al., 2004; Gallistel & Gelman, 1992), we may expect signatures of this system to be detectable not only through tasks requiring the perceptual processing of numerical symbols but also when speakers use number words in spontaneous language.

Two main theoretical reasons may support such a proposal. First, production is a privileged field to explore any relationship between cognitive systems primarily deputed to process sensory-based information, such as the ANS, and language forms. Indeed, language does not constitute an encapsulated system (Garrett, 1989): processes that manipulate perceptual and linguistic signs must, at some levels, share a common set of representational knowledge as well as organizational principles (Osgood, 1971). Accordingly, there is evidence for a contribution to language production by perceptual systems (Glenberg & Gallese, 2012; Osgood, 1971). When translated to numerical processing, this may suggest that the way the brain handles non-symbolic perceptual information about numbers may in turn influence the linguistic system deputed at expressing such information. Second, and critically, language perception and production are inherently tangled together (Liberman & Whalen, 2000; Pickering & Garrod, 2013). For instance, the neural network deputed to speech perception largely overlaps with brain areas involved in language production (Wilson, Saygin, Sereno, & Iacoboni, 2004). In line with this, the way humans process symbolic numerical information (i.e., with the relative compliance with Weber's law) may as well favor the spread of the ANS signatures in spontaneous language. On these bases, we therefore hypothesize that humans may make unconscious reference to the corresponding ANS representations not only when processing sensory-based symbolic and non-symbolic magnitudes, but also when expressing information about quantity at the level of output, in language production (see Fig. 1B for the model proposed and tested). Evidence for such a system governing numerical symbolic processing at both input and output levels would provide support to the notion that numerical symbols are grounded in the ANS. More generally, this model would favor a strict interrelation between the faculty of language and the evolutionary ancient system deputed to quantity.

Whether the ANS regulates the spontaneous verbal production about quantity is a possibility that has been indirectly considered by previous research. However, and critically, previous studies have mostly explored this issue through the analysis of simple lexical frequencies from text corpora. In a seminal paper, Dehaene and Mehler (Dehaene & Mehler, 1992) reported that the lexical frequency of numerals 1-9 decreases with numerical magnitude (i.e., with a comparable, albeit less consistent pattern for numerals 10-19). Similar results were observed for Arabic number reading, with the time to process an integer that is a function of the logarithm of the number magnitude and of its lexical frequency (Brysbaert, 1995). This indicates that the internal representation of numerals, as assessed by the lexical frequency of spontaneous verbal production, is more detailed for small numbers than for large ones, thus being overall compressive (Dehaene & Mehler, 1992). This pattern of results was more recently replicated cross-culturally considering larger corpora and extended to the decade words as well as across historical time (Piantadosi, 2014, 2016). While this evidence on lexical frequency provides some indirect evidence for an ANS regulating verbal behavior about quantity, it has also some evident theoretical limitations, as it does not allow to replicate the core signatures of such system (i.e., ratio effect or Weber's law). Indeed, the frequency distribution of number words in language may simply reflect environmental pressures for representing small numerical values, rather than being similar to the internal organization of the ANS (Piantadosi, 2016; Piantadosi & Cantlon, 2017).

Here, we take advantage of distributional systems from computational linguistics to test the hypothesis that the ANS affects the usage of number words in language. Models that represent language behavior as high-dimensional numerical vectors, extracted from large amounts of natural language data, appeared in the field of cognitive science about twenty years ago, with the introduction of the prominent LSA (Latent Semantic Analysis; Landauer & Dumais, 1997) and HAL (Hyperspace Analogue to Language; Lund & Burgess, 1996) models. These vector-space models (VSMs) have since then received considerable attention and have been shown to be impressively high-performing across a wide range of cognitive tasks (for reviews see Jones, Willits, & Dennis, 2015; Lenci, 2018). Perhaps even more intriguingly, recent evidence indicates that these models do not only capture linguistic knowledge, but they can efficiently grasp also (implicit) mental representations. For instance, these text-based models can inform about very implicit information such as social and cultural biases, replicating human performance as assessed by computerized paradigms (Bhatia, 2017b; Caliskan, Bryson, & Narayanan, 2017). Similarly, another recent study demonstrated how these computational models can efficiently capture the representations underlying high-level human judgments (Bhatia, 2017a).

The theoretical foundation of VSMs is the distributional hypothesis, according to which similar words should behave similarly in language, and therefore can be grouped according to their distributional behavior (Harris, 1954). Under this assumption, language behavior is described as the statistical distribution of linguistic items in contexts (Sahlgren, 2008). Accordingly, applying a rigorous operationalization of “context” allows the quantification of a word distribution and, consequently, the linguistic behavior associated with a particular linguistic item. Common definitions of contexts are the documents (i.e. sentences, paragraphs or articles) a word occurs in (Griffiths, Steyvers, & Tenenbaum, 2007; Landauer & Dumais, 1997) or other words within a fixed-size window around it (Jones & Mewhort, 2007; Lund & Burgess, 1996; Mikolov, Chen, Corrado, & Dean, 2013;

Pennington, Socher, & Manning, 2014). In traditional VSMs, the first step in the implementation of the model is to construct a word-by-context matrix. By applying a mathematical encoding of the distributional properties of lexemes, it is thus possible to encode a lexical item as a distributional vector representing its co-occurrences with linguistic context (i.e., the cell entries of such a matrix are the word co-occurrences counts). In the last stage, two words will be similar in linguistic behavior when they have similar global distributional patterns over all contexts. In actual implementations, however, these word-context co-occurrences are typically subject to further processing, with a transition from first- to higher-order relations.

On these grounds, we therefore reasoned that distributional models may be an ideal methodological tool to explore the language behavior associated with the usage of number words (see Brysbaert, 2018). In this study, we specifically used a recent family of distributional models, namely word-embeddings models (Baroni, Dinu, & Kruszewski, 2014; Mandera, Keuleers, & Brysbaert, 2017; Mikolov et al., 2013), to explore the distributional behavior of number words in spontaneous language. In contrast to the typical VSMs described before (i.e., generally referred to as "count" models, as the computational implementation is based on a common counting step), word-embeddings are based on a neural network built on a predictive component (i.e., referred to as "predict" models) (Baroni et al., 2014; Mandera et al., 2017). Briefly, word-embeddings are trained on large collections of texts that document natural language use. Nodes in the input and output layers represent words and a neural network learns to predict a target word on the basis of the lexical contexts in which it appears (i.e., the words it co-occurs with in the text), incrementally updating a set of weights by minimizing the difference between model predictions and observed data at each learning event (i.e., every occurrence of the target word). For example, when presented with the sentence *there is a cat in the garden*, the model updates the distributed representation of the word *cat* to improve its prediction given the context words (*there*, *is*, *a*, *in*, *the*, *garden*). The estimated set of weights will eventually capture linguistic behavior associated with a specific word in distributed terms. These distributed representations, or vectors, can be quantitatively compared by measuring their distance in a multidimensional space. Such representational distance captures the similarity between words in terms of distributional usage within language: similar words will occur in similar contexts, ending up being associated with vectors that are geometrically close. Essentially, the distributional similarity between two words as extracted from language behavior amounts to their degree of mutual substitutability: in language usage, *cat* can be more easily substituted with *kitty* as compared to *snake*. Importantly, as the architecture of word-embeddings is based on a context-prediction principle, words that are used similarly in language will not merely co-occur more often with each other: two words are similarly used in language not because of their mutual co-occurrence score, but rather if they have similar global distributional patterns over all contexts (e.g., Lenci, 2018). In addition to their potential theoretical appeal, word-embeddings have recently proven to be a robust method for predicting human performance in various psychological experiments (Baroni et al., 2014; Caliskan et al., 2017; Mandera et al., 2017).

Whether a data-driven model trained on text corpora is able to isolate specific behavioral profiles about number words in natural language is itself a challenging question. In fact, number words like *four*, *six*, and *nine* all denote a property of a set, namely its numerosity, and are likely used in similar lexical contexts. Here, our goal nonetheless goes a step further. We reasoned that, if numerical symbols are associated to a pre-verbal ANS, this association should be reflected in the way we speak about quantities: the distribution of vectors representing the usage number words should obey the same psychophysical laws that sensory judgments follow, including Weber's law, and should capture the basic regularities of our mental apparatus encoding quantity. As a consequence, a clear ratio signature should be found in language behavior (Dehaene, 1997; Piazza, 2010). At the same time, however, it should be noted that distributional models are entirely based on spontaneous language production. In striking contrast, the main hallmarks of the ANS -including the ratio effect- have been insofar mainly observed in experimental tasks, whereby participants are forced to perform a numerical magnitude comparison. It is thus possible that ANS signatures may characterize the symbolic faculty for numbers only when an explicit magnitude comparison is required and would not be observable in everyday spontaneous language.

We first considered word-embeddings based on a large-scale written corpus of English and we extracted vector representations for number words from *two* to *twenty*. We then explored whether the extracted distributional similarity between number words in natural language is associated with the actual numerical distance between their corresponding quantities, thus attempting to replicate the typical distance effect observed in number comparison tasks (Moyer & Landauer, 1967). Following a similar rationale, we next probed the presence of a size effect (Dehaene, 2003; Piazza, 2010). Specifically, we tested whether small numbers are more precisely used than larger numbers in spontaneous language behavior. After having identified these two language-based metrics, we explored whether Weber's law can account for the usage of number words in natural language. We next performed control analyses to rule out the possibility that the observed pattern of results was determined by simple lexical frequencies and tested some comparative sets (e.g., months and days of the week). Finally, we also aimed at providing further insight into the complex nature-nurture debate surrounding number words. In particular, we attempted to replicate the compliance with Weber's law cross-linguistically, in nine different languages. These included three Germanic languages (English, Dutch, and German), two Romance languages (Italian and French), a Slavic language (Russian), a Hindustani language (Hindi), a Sino-Tibetan language (Chinese) and an artificial language (Esperanto). Most of the languages were selected based on corpus availability and to probe the possible cross-cultural generalizability of our findings, starting from English. Chinese was selected on purpose based on previous evidence (e.g., Tang et al., 2006) to possibly detect any effect of cultural mediation on symbolic numbers also in natural language usage. Similarly, we also tested Esperanto, which is a relatively recent artificial language created in the late 1870s. In this specific case, therefore, cultural effects should be relatively restricted compared to languages with a long tradition in literacy.

Method

Details on the vector-space model for English

The VSM of reference (Baroni et al., 2014) for the analyses on English was induced from a 2.8-billion-word corpus obtained by a concatenation of ukWaC[1*], the English Wikipedia, and the British National Corpus[†]. The VSM was trained using the word-embeddings approach proposed by (Mikolov et al., 2013), and in particular the Continuous Bag of Words (CBOW) method. Word-embeddings are built on a neural network architecture with one hidden layer (Mikolov et al., 2013). In this architecture, words serve as the input layer from which its context words – the output layer – are predicted (or vice versa; see below for details on the CBOW and SkipGram approaches). The activation values of the hidden layer for a given word in the input layer is then taken as its distributional vector. This network is trained on a corpus, word by word, and so the incremental development of distributional vectors over time is an inherent property of this algorithm. Here, we opted for a parameter setting that provided the best performance across a number of tasks, as described by Baroni et al. (Baroni et al., 2014) in a large-scale optimization study[‡], and mainly in line with the psycholinguistic evaluation by Mandera and colleagues (Mandera et al., 2017): 5-word co-occurrence window, 400-dimension vectors, negative sampling with k=10, subsampling with t=1e^{-5}.

This set of parameters defines the learning procedure used to induce word vectors (Mikolov et al., 2013). CBOW indicates the applied learning procedure: when using CBOW, the obtained vector dimensions capture to what extent a target word is reliably predicted by the contexts in which it appears; conversely, the SkipGram approach would weight how well contexts are predicted by the target word. Co-occurrence window size indicates how large the considered lexical contexts are; in our case, a 5-word window indicates that we estimated predictions concerning 2 words on the left and 2 words on the right of the target word. The number of vector dimensions indicates how many nodes

[*] http://wacky.sslmit.unibo.it
[†] http://www.natcorp.ox.ac.uk
[‡] http://clic.cimec.unitn.it/composes/semantic-vectors.html

are included in the hidden layer of the neural network, essentially specifying the extent of the obtained dimensionality reduction. Negative sampling estimates the probability of a target word by learning to distinguish it from draws from a noise distribution; the parameter *k* specifies the number of these draws. The subsampling parameter *t* specifies a threshold-based procedure that limits the impact of very frequent, uninformative words.

From this VSM, we extracted vector representations for number words (numerals) ranging from *two* to *twenty*. *one* was excluded by the items list because of its polysemy: this word is indeed often used in contexts where it does not directly denote a quantity (e.g., when it is used as a pronoun), resulting in a less specific vector encoding different aspects along with numerical information. In a subsequent analysis, we also tested our model on a wider range of number words, from *two* to *fifty* (see the Supplemental Material). Finally, for the analyses on ordinal information, we extracted vector representations for months (*January*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*, *December*) and days of the week (*Monday*, *Tuesday*, *Wednesday*, *Thursday*, *Friday*, *Saturday*, *Sunday*).

Details on the cross-linguistic evaluation

VSMs for the considered languages were obtained from a variety of sources and applying a variety of parameter settings. The English space was based on Baroni et al. (2014): CBOW model, 5-word window, 400 dimensions, trained on a concatenation of ukWaC, BNC, and English Wikipedia. The Italian space was released by Marelli (2017): CBOW model, 9-word window, 400 dimensions, trained on itWaC. The Dutch space was obtained from Mandera et al. (2017): CBOW model, 9-word window, 200 dimensions, trained on a concatenation of SONAR and a subtitle corpus. The German space was ad-hoc trained, adopting the same parameter setting used for English: CBOW model, 5-word window, 400 dimensions, trained on a concatenation of deWaC and German Wikipedia. The VSMs for French, Russian, Chinese, Hindi, and Esperanto were obtained from an online NLP resource[§]: SkipGram model, 5-word window, various vector dimensionalities, trained on the respective Wikipedia. The datasets used for the evaluation of Esperanto and Russian were slightly smaller, as some of the numbers (*eighteen* and *nineteen* for Esperanto, *nineteen* for Russian) were not found in the corresponding VSMs.

Results

Distance effect in the vector-space model

We first explored whether the linguistic behavior in the use of number words is associated with the actual numerical distance between the associated quantities. The distance effect is a typical empirical finding observed in number comparison tasks, whereby participants are instructed to indicate the smaller/larger of two visually presented numerals or non-symbolic numerosities. In most experimental designs used in the literature, the two compared stimuli are simultaneously presented on the left and right sides of a computer screen, respectively (for the sequential task variant, see Price, Palmer, Battista, & Ansari, 2012). A distance effect is reflected by an enhancement of performance (i.e., a decrease in reaction times and error rate) as the distance between two numbers increases (Moyer & Landauer, 1967). That is, responses are consistently faster and more accurate for larger (e.g., *two – nine*) than for smaller (e.g., *two – three*) numerical distances. One influential view, initially offered by Restle (Restle, 1970), is that the distance effect would reflect the placement of numbers on an analog continuum. The difficulty in discriminating between two numerals would depend on the distributional overlap of their representations on the analog continuum. Accordingly, the distributions would overlap more for numerals that are numerically close, giving rise to the observed distance effect (for an alternative view, see Verguts, Fias, & Stevens, 2005).

To investigate whether number words in natural language also follow the distance effect (i.e., whether closer numbers are used more similarly than distant ones), we estimated the degree of

[§] https://github.com/Kyubyong/wordvectors similarity between their corresponding model-induced vectors. Specifically, for each word pair, we obtained a distance metric (vector distance: VD) based on the cosine of the angle formed by vectors representing the distributional use of these words, as a proxy for real numerical distance. VD was computed using the following formula, where $\vec{x}$ and $\vec{y}$ are the *n*-dimensional vectors associated to a pair of number words:

$$VD(\vec{x},\vec{y}) = 1 - \frac{\sum_{i=1}^{i=n} x_i \times y_i}{\sqrt{\sum_{i=1}^{i=n} x_i^2} \times \sqrt{\sum_{i=1}^{i=n} y_i^2}}$$

Hence, the closer the distance between two vectors in the multidimensional space, the more similarly the number words will be used in natural language. Vectors were extracted by the model based on Baroni et al. (2014) and trained on a concatenation of the British National Corpus, English Wikipedia, and ukWaC (a web-based English corpus), as described before. We considered the vectors of number words from *two* to *twenty* (number *one* was excluded from the model as it is a polysemous word) (Fig. 2A). Pearson correlation analysis between VD for each pair of vectors representing number words and the real distance between numbers yielded a positive, strong association [$r(171)=.7573$, $P=.0001$] (Fig. 2B) (for more detailed results see Supplemental Material). Therefore, the similarity in linguistic behavior, as induced by the model, between a target number word (e.g., *five*) and the other number words decreases as a function of real numerical distance (e.g., *five* being used in spontaneous language more similarly to *four* and *six*, compared to *three* and *seven*). From a complementary standpoint, this means that in the sentence *there are five birds on the tree*, the model will estimate that *four* and *six* are more suited substitutes of *five* than more distant number word candidates (e.g., *two*, *three*, *seven*, *eight*, etc.).

Despite in distributional models words that never occur together can nevertheless end up with very similar distributional profiles (Landauer & Dumais, 1997), we directly tested whether the observed pattern of results was due to high relatedness between some numbers in idiomatic expressions (e.g., as in the sentence *he plays soccer two or three times a week*). We thus extracted the raw co-occurrences between numerals. That is, for each pair of number words, we computed how many times both elements were observed in the same sentence across the whole corpus. These counts were found to be negatively correlated with numerical distance, even though the effect was relatively small [$r(171)=-.2555$, $P=.0007$]. Most critically, the partial correlation between VD and real numerical distances while controlling for word co-occurrence counts remained large and almost unaffected [partial correlation: $r(168)=.7376$, $P=.0001$], ruling out the possibility that our findings were due to examples of number pairs used in idiomatic expressions.

We then tested the uniformity of the observed distance effect by computing the standardized residuals between VD and real numerical distances (i.e., by fitting a linear regression with the former as the independent variable and the latter as the dependent one). Only 8 data points (on a total of 171 number pairs) had residuals falling outside +/-2 SD. When these data points were removed, the correlation between VD and real numerical distances remained large and of the same extent [$r(163)=.8404$, $P=.0001$]. This suggests that the effects observed were not driven by small subsets of numbers.

We finally probed the robustness of the association by testing this model against 10,000 simulations, in which we correlated the extracted VD with real numerical distances altered by minimal deviations (i.e., ±1). The correlation between model estimates and numerical distances was indeed larger than the one between the same estimates and minimally altered distances in 96% percent of the cases (9,565 simulations out of 10,000). This simulation approach was also applied to evaluate the precision of the word-embeddings model, that is, determine at which level word-embeddings estimates are not able anymore to discriminate between the actual numerical distances and the noisy ones. We thus considered 200-noise levels, each one represented by a deviation ranging from ±2 to ±0.01. For each noise level, we tested model predictions in 10,000 simulations. Fig. S1 in Supplemental Material shows that the model estimates remain robust in predicting actual numerical distance across a wide range of noise levels. The two data sets (actual distances vs. noisy distances) were not discriminable only for very low noise level (±0.01), but even for a noise level of ±0.02 the model performance, although low, was above chance level [$\chi^2(1)=8.91$, $P=.0028$]. The presence of such a tight association indicates that language-based estimates are specifically related to their corresponding numerical distances, rather than reflecting rough signals that accidentally express numerical relations. Thus, the way number words are used in language production is similar to the way humans process symbolic and non-symbolic quantities in psychophysical experiments. The presence of a distance effect in the VSM further suggests a spontaneous connection with the ANS in daily language production.

Size effect in the vector-space model

We then investigated whether numerosity is represented in language as a fluctuating mental magnitude. The size effect indicates that discrimination of a pair of numbers with equal numerical distance worsens as their numerical size increases (Moyer & Landauer, 1967), a pattern that has been also observed in single-digit naming tasks (Cohen Kadosh, Tzelgov, & Henik, 2008). Accordingly, current conceptualizations of the ANS assume not only that the representation of each number is approximate, but also that this imprecision increases proportionally with magnitude (Dehaene, 2003; Piazza, 2010). If number words in natural language call upon the ANS, we then should expect numerals denoting smaller numbers to be more precisely used than numerals denoting larger numbers. In distributional models, this relationship should be captured by the variance between the dimensions of the vectors representing number words (vector variance, VV), computed with the following formula, where $\vec{x}$ is the n-dimensional vector associated to the number word, $x_1 \ldots x_n$ its n components, and μ the average value of these latter values:

$$VV(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

As in the characterization of vector entropy in traditional distributional models (Marelli & Baroni, 2015), VV refers to the specificity with which the target word can be predicted by the linguistic context in which it typically appears (Mikolov et al., 2013). Words with higher VV will be associated with many different contexts, whereas words associated with specific contexts will end up being represented through lower VV. In computational terms, therefore, VV can be taken as an indicator of the precision of the usage of each number word in natural language. We found a positive and strong correlation between VV and number magnitude [$r(19)=.8215$, $P=.0001$], thus unveiling that the representation of larger number words in the VSM is more approximate and noisy than smaller numerals, mirroring the evidence from psychophysical experiments (Dehaene, 2003; Piazza, 2010) (Fig.3).

To further probe the uniformity of the observed effect, and in analogy to what we have done for VD, we computed the standardized residuals between VV and real magnitude (i.e., by fitting a linear regression with the former as the independent variable and the latter as the dependent one). Only 1 data point (on a total of 19) had residuals falling outside +/-2 SD (e.g., *twenty*). Importantly, when this data point was removed, the correlation between VV and real magnitude remained large and of the same extent [$r(18)=.8934$, $P=.0001$].

Finally, we also tested whether the size effect was equally present in the two portions of the number line. To this aim, we computed two correlation analyses between VV and magnitude on the sub-datasets 2-10 and 11-20, respectively. The results showed that the size effect was similar in the two portions of the number line [2-10: $r(9)=.5801$; 11-20: $r(10)=.4666$], thus providing additional support to the uniformity of the findings described.

Ratio dependency in the vector-space model

After having identified two language-based metrics that parsimoniously describe numerical distance and numerical magnitude, we tested whether Weber's law can account for the verbal usage of number words, as captured in the data-driven model. Experimental research has shown that the discriminability of two numbers is dependent on their ratio, as this index takes simultaneously into account not only the difference between them but also their absolute size. Specifically, larger numerical ratios (i.e., closer to 1) are associated with poorer performance because of the similarity between the to-be-compared stimuli increases. We thus explored whether the VD of each pair of number words along with the VV of the two number words predicted numerical ratio. In a regression model we found significant effects for both VD [$t(167)=-15.94$, $P=.0001$] and VV [smaller number: $t(167)=16.3$, $P=.0001$; larger number: $t(167)=4.71$, $P=.0001$] on the numerical ratio (Fig. 4) (for more detailed results see Supplemental Material, Table S1). Notably, the variance explained by such a model ($R^2=.7978$) is fully compatible with the variance accounted for by psychological experiments in which participants are asked to perform comparisons of number words (Moyer & Landauer, 1967).

To probe the uniformity of the ratio effect, we computed the standardized residuals of the regression model (e.g., numerical ratio as the dependent variable and three model-based estimates as predictors). Only 8 data points on a total of 171 had residuals falling outside +/-2 SD. When these data points were removed, the variance explained by the model remained large and of the same extent ($R^2=.8686$). This further corroborates the fact that the effects observed were not driven by small subsets of numbers.

Finally, we replicated the ratio effect in a wider range of numbers, including number words from *two* to *fifty*. Even in this case, we found a significant effect for both VD and VV (for more detailed results see Supplemental Material, Figure S2, and Table S2).

Taken together, these results suggest that the use of number words in natural language production obeys Weber's law and it is likely grounded on a noisy ANS. In more concrete terms, the difference in the way we speak about *two* and *four* in everyday language would be very similar to the difference in the way we speak about *four* and *eight* because the ratio subtending these number word pairs is equal, i.e., 0.5. Remarkably, the Weberian nature of numbers seems to originate spontaneously, without the need of any explicit comparative process. The ANS, therefore, would be compulsorily activated whenever we use number words in everyday spontaneous language.

Lexical frequency of number words

Yet, because numerical quantities are known to be associated with lexical frequency of the corresponding number words (i.e., the smaller the number, the more frequent the corresponding word) (Dehaene & Mehler, 1992; Hutchinson & Louwerse, 2014), we tested to what extent our findings merely depend on the model-produced estimates encoding frequency information. As expected (Dehaene & Mehler, 1992; Hutchinson & Louwerse, 2014), number magnitude was found to be negatively correlated with log-transformed lexical frequency extracted from the English corpus used [$r(19)=-.9022$, $P<.0001$]. We therefore first fitted a regression model with log-transformed frequency of both number words as predictors and ratio as dependent variable ($R^2=.7678$). This model served as a strong baseline to test whether our model-based distributional indexes can account for more variance compared to purely lexical frequency parameters. We next sequentially entered VD, VV of the smaller number and VV of the larger number in the baseline model. In each step, we found that the contribution of each distributional index entered was significant and that the variance accounted by the model increased incrementally (VD entered: $R^2=.8342$; VV smaller number entered: $R^2=.84$; VV larger number entered: $R^2=.8477$) (for more detailed results see Supplemental Material, Table S3). These findings indicate that the distributional model described above cannot be reduced to the effect of lexical frequencies. The fact that model estimates are robust with respect to such a strong baseline in capturing numerical aspects speaks in favor of the validity of our proposal: the association between words and the ANS may be reflected in nuanced distributions of language usage.

Comparative sets of words with ordinal relationship

We also tested whether the Weberian adherence of word-embeddings is restricted to number words or rather extends to other word sets characterized by ordinal knowledge. In comparison tasks, the distance effect may, in fact, refer to any influence of the semantic distance between two stimuli (Moyer & Landauer, 1967). Accordingly, similar behavioral effects have been reported for the processing of ordinal information, such as days of the week (Gevers, Reynvoet, & Fias, 2004) and months (Gevers, Reynvoet, & Fias, 2003) or, more generally, any newly learned verbal ordered sequences (Previtali, De Hevia, & Girelli, 2010). Importantly, no analogous size effect should be expected with these words and hence no compliance with Weber's law should be observed. On these grounds, we selected two sets of words that have been empirically tested by previous literature (e.g., days of the week and months). Months were assigned a 1-to-12 value with reference to *January-December*. As the week may start in some calendars from Sunday, while in others from Monday, days of the week were assigned a 1-to-7 value with reference to both *Monday-Sunday* and *Sunday-Saturday* sequences. Results showed that the variance accounted by the models with months and days of the week was much lower compared to the model with number words (months: R^2=.2477; days *Monday-Sunday*: R^2=.1135; days *Sunday-Saturday*: R^2=.3738) (for more detailed results and for control analyses on lexical frequency see Supplemental Material, Table S4). These results, therefore, support the uniqueness of the Weberian nature of number words and rule out the possibility that other correlated features, by chance captured by the distributional model, may explain the current results.

Cross-linguistic generalizability

Importantly, the generalizability of our findings on number words was fully supported by applying the model to different languages. These included two other Germanic languages (Dutch and German), two Romance languages (Italian and French), a Slavic language (Russian), a Hindustani language (Hindi) and a Sino-Tibetan language (Chinese) (for more detailed results see Supplemental Material, Table S1). Hence, compliance with Weber's law was extended to various linguistic corpora and replicated cross-linguistically (Fig. 4). We also applied the model to Esperanto to probe to what extent the faculty for symbolic number representation is a product of cultural evolution. Compared to ethnic languages with stronger cultural transmission, Esperanto is a relatively recent constructed language created in the late 1870s, whose grammar, phonology, vocabulary, and semantics are artificially designed rather than having evolved naturally over time. Evolutionary constraints are therefore limited for Esperanto as compared for instance to English, which has developed over a much longer period of time since its roots can be traced back to the fifth century AD. Even in the case of Esperanto, nevertheless, we found that the use of number words obeyed Weber's law (Fig. 4). Taken together, these findings provide support to the fact that the distributed pattern encoding number word usage is similar across different languages (for a discussion on the possible evolutionary foundations of number see Núñez, 2017).

Discussion

In this study, we employed a machine-learning technique developed in the domain of computational linguistics to explore whether linguistic behavior, and in particular language production, shows signatures reminiscent of the pre-verbal system for numerosity. We hypothesized that, if numerical symbols are grounded on a pre-verbal ANS, a ratio effect should be observable not only when humans process numerical information at the input level, but also when humans use number words in spontaneous language, grounding this prediction in the strict interrelation between language production, language perception, and sensory processes. Our findings provide support to such a theoretical model, by demonstrating that the core signatures of the ANS (namely distance, size and ratio effects) also characterize spontaneous verbal behavior. That is, nuanced distributions of number-word usage in linguistic contexts follow the organizational principles of the ANS. Together, these findings offer new evidence to the hypothesis that our faculty for symbolic numbers harks back to an evolutionarily ancient ANS originally designed for processing non-symbolic quantities (Verguts & Fias, 2004).

In the theoretical framework proposed, the ANS occupies a hierarchically higher position, as this system has a long phylogenetically history and is deputed to process numbers on a non-linguistic basis. In analogy to traditional accounts, we assume that the coarse tuning curves on the number representation would be responsible for the ratio effect observed at the input level (Dehaene &

Changeux, 1993; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; for alternative accounts see, Verguts et al., 2005). The coarse tuning implies that the presentation of a number at the input level would not only activate the target number alone, but also the neighboring numbers because the coding schemes of their neuronal populations are similar (Dehaene & Changeux, 1993). The major overlap of tuning curves for number that are numerically closer would determine the distance effect observed in numerical comparison tasks. Similarly, the organization of numerical representations in the ANS would determine the size effect. Indeed, previous studies suggest a compressive internal scale of tuning curves, with an increasingly coarse encoding of larger numbers (Dehaene & Changeux, 1993; Piazza et al., 2004). The fact that performance in comparison tasks depends on numerical ratio (i.e., incorporating both distance and size effects) is therefore taken as evidence for a scaling from perceptual input into internal representations in the ANS. In short, these signatures would reflect the organization itself of the ANS. Based on this, we hypothesize that the organization of the ANS is reflected in language because of the position of this system in both the hierarchy of cognitive processes and in the evolutionary timeline. That is, the ANS may serve as the foundation for symbolic numerical processing; and, the ANS obviously appeared earlier in evolution as compared to the symbolic faculty for numbers. This would explain why a permeation of the former system can be found in the latter. In addition to this, and as discussed above, the strict interrelation between language production, language perception, and sensory processes may have further strengthened the observed permeation. The result of such considerations is the pattern of linguistic phenomena reported here: when producing language about quantities, the properties of the ANS representations will percolate into linguistic processes and will be expressed as subtle, nuanced modulations of lexical distributions. These lexical modulations, in turn, will be observed in corpora of spontaneous language, through the application of convenient computational methods. How such a mapping between the ANS and the faculty for symbolic number has evolved remains, however, an incredibly interesting puzzle. We still miss a causal account that may explain why a phylogenetically older system permeated into a linguistic sub-system, which has the primary goal of communicating exact quantities and is rooted in 1-to-1 procedures.

One possibility is that this permeation has occurred not simply because these systems encode and process the very same type of information (i.e., quantity), but because of the somewhat similar architectures these systems rely on. As described previously, the ANS is believed to represent numerosity as a fluctuating mental magnitude, taking the form of a number line (Feigenson et al., 2004). Two competing mathematical implementations of the number line have been advanced, with both of them leading essentially to identical predictions (i.e., ratio-dependence) in behavioral tasks (Dehaene, 2001). The linear model represents the mental number line with a linear scale (i.e., equally spaced distributions among numerosities), but with an increasing spread (i.e., standard deviation) of the internal noise of each numerosity (Gallistel & Gelman, 1992). On the contrary, the logarithmic model assumes a logarithmic scale (i.e., unequal space distributions among numerosities, with lower representational distance as magnitude increases) with a fixed amount of noise (Dehaene & Changeux, 1993). In both models, larger numerosities will be coded by distributions that overlap increasingly with nearby numerosities, leading in turn to a ratio-dependent performance. Intriguingly, the architecture of distributional models resembles, in a certain way, how numerosity is represented in the mental number line. Vector-space models, indeed, operationalize the distributional hypothesis by means of a mathematical encoding of the distributional properties of linguistic units (Lenci, 2018). Depending on its distributional pattern (i.e., as induced through word co-occurrences), each lexical unit will therefore take a geometrical position (i.e., encoded in terms of a vector) and populate a vector space. Notably, metrics such as cosine similarity will determine the relative geometrical proximity between words, a parameter that has been demonstrated to thoroughly capture word usage similarity. That is, more similar words will tend to be used in more similar linguistic contexts and, consequently, will end up being represented by vectors that are geometrically closer in the text-based space. Other intrinsic properties of distributional vectors, such as their entropy or variance, have also been shown to be of cognitive interest. For instance, high-entropy and high-variance vectors will be associated with less specific linguistic contexts. The combination of these two vector-space metrics (i.e., cosine distance and vector entropy), therefore, may provide sufficient ground for architectural replication of the number line. The related features between the ANS and the symbolic system for numbers (specifically, as captured by language usage) may have made possible the permeation of the former into the latter.

The main behavioral hallmarks of the ANS, namely the distance effect and the size effect, have been commonly observed in the number comparison task (Verguts et al., 2005). In this experimental design, participants are explicitly instructed to select the larger (or the smaller) of two simultaneously presented symbolic numbers. The common interpretation about the presence of distance and size effects is that the symbolic numbers are converted to percept-like analog quantities, with these analog representations that are in turn processed by the ANS just like any representation of sensory stimuli (Butterworth, 1999; Dehaene, 1997; Feigenson et al., 2004; Moyer & Landauer, 1967). Surprisingly, our findings indicate that similar effects can be observed also when no explicit magnitude comparison is required (e.g., in the sentence *there are six candles on top of the cake* there is only one reference magnitude) and in the context of linguistic production. This implies that the neural and cognitive processes underpinning such an ancient system are not restricted to the transformation of sensory input into internal analog quantity, but rather also operate at the level of active verbal output. The architecture of the ANS would be therefore reflected in the distributional proprieties of spontaneous language. Indeed, language does not often reflect the mere descriptive transcription of perceptual information available to the different senses from the outer world. Rather, language also expresses information that is not contingently available to our perceptual systems. A very large portion of our linguistic production is dedicated to communicating abstract content and information stored in our memory systems. These arguments apply as well to the use of numerals in language.

The interpretation of the present results may be well complemented by studies highlighting the remarkably slow rate of historical lexical replacement of small number words (e.g., especially from *one* to *five*) (Pagel, Atkinson, & Meade, 2007). This pattern of linguistic evolution, documented in different language families (Pagel & Meade, 2017), has been accounted –among several other interpretations- by the phylogenetically older system for processing numerical information. In particular, the strict interplay between brain regions involved in language and those deputed to processing quantity may be responsible for the slow replacement rate of number words (Pagel & Meade, 2017). More specifically, the parietal areas are equipped with a specialized subsystem for quantity that may have provided a stable and fixed location to the language for number, thereby slowing their rates of replacement in comparison to other categories of words that do not rely on such an evolutionary old neural network. The fact that we observed non-linguistic hallmarks of the ANS in purely textual scenarios is in line with this possibility. Yet, it is worth specifying some limitations in the generalizability of our cross-cultural findings. All the languages tested here, except Esperanto, have indeed a long tradition in literacy and have a decimal counting system. To further address the nature-nurture debate, it may be desirable to extend these findings also to small-scale societies (such as native groups in Papua New Guinea), in which non-decimal body-part tallying systems are used (Saxe, 2015; Wassmann & Dasen, 1994), despite the availability of linguistics corpora represents a clear limitation to this approach.

The problem of learning number words has been recently interpreted within the information theory framework (Ramscar, Dye, Popick, & O'Donnell-McCarthy, 2011). Under this theoretical scenario, learning is conceived as a process that has evolved to facilitate the learner in predicting outcomes by weighing and assessing the informativity of relevant cues (Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). In particular, learning would be achieved through a competitive procedure that assigns weighted cues to features in the environment with the ultimate goal of predicting events. Such an approach is well suited also for describing the acquisition of number words. In learning the number words, indeed, an individual will have to predict the most relevant cues (e.g., a set-size in the environment) and associate it to a given number word, by simultaneously discarding all the other non-relevant cues (e.g., other set-sizes or quantity-related features) (Ramscar et al., 2011). As the architecture of our model is consistent with biologically-grounded (and relatively simple) associative learning mechanisms, based on a prediction principle (Baroni et al., 2014; Mandera et al., 2017), our findings converge with these previous data supporting the importance of information structure in number words learning. Yet, our results also suggest that, in principle, it is possible to infer the cognitive architecture subtending the use of number words only from textual input. This possibility challenges prevailing views of number words acquisition, which mantain the way children acquire numerical symbols by mapping them to the abstract representations in the ANS is primarily dependent on (visual) perception (Mussolin, Nys, Leybaert, & Content, 2016). According to these proposals, children may learn the exact quantity subtended by numerals only thanks to recursive perceptual correspondences between sets of objects and linguistic cues, or through the external counting routine (Bloom & Wynn, 1997; Carey, 2009; Gelman & Gallistel, 1986). Findings from our study rather indicate that learning the quantity subtended by number words may be achieved merely through language statistics. This is in line with recent evidence showing that the relationship between the ANS and symbolic numerical reasoning does not depend on visual experience, as this link looks preserved in congenitally blind individuals (Kanjlia, Feigenson, & Bedny, 2018).

Previous works on spontaneous verbal production have explored the mental organization of number concepts through the analysis of lexical frequencies (Dehaene & Mehler, 1992; Piantadosi, 2014, 2016). Yet, this descriptive measure cannot satisfactorily replicate the main signatures of the ANS. In addition to this, we notice that the word frequency effect is thought to arise only at later stages of speech preparation and, specifically, at the level of accessing the phonological code (Jescheniak & Levelt, 1994). The possible connection between language and the ANS should be rather found at higher levels of speech preparation, possibly at the level of conceptual preparation (Levelt et al., 1999). But can behavioral distributional signatures of language usage be informative about these higher-level stages of speech preparation? According to many proposals, VSMs cannot be reduced to methodological tools that simply describe linguistic behavior (Günther, Rinaldi, & Marelli, 2019; Lenci, 2018). Rather, VSMs have been also conceived as a theory explaining how semantic representations are acquired and, thus, as good proxies of human semantic memory (Günther et al., 2019). In line with this possibility, VSMs achieve perfect accuracy on multiple choice tests, such as the TOEFL (Test of English as a Foreign Language; Bullinaria & Levy, 2012); above 95% in tasks indirectly involving semantic similarity processes, such as categorizing nouns (Baroni & Lenci, 2010); and high correlations with human word similarity ratings (Baroni et al., 2014; Bruni, Tran, & Baroni, 2014). Because VSMs are considered to be good proxies for representations of the human semantic memory, one may consequently speculate that the compliance with Weber's law in natural language observed here may lay at the level of conceptual preparation, thus when zooming in on the appropriate lexical concept in the mental lexicon (Levelt et al., 1999). This is a particularly intriguing, but rather provocative proposal. Indeed, it implies that the numerical mental lexicon would be tightly coupled with the ANS, in that when a speaker is retrieving a numerical concept, the approximate quantity subtending the concept itself will be inevitably activated. Indirect evidence convergent with this possibility comes from a recent study with scalp electroencephalography, showing that the same intraparietal sulcus neural populations activated during an experimental arithmetic condition are also activated in natural conditions (i.e., when participants read, hear or speak words with numerical content) (Dastjerdi, Ozker, Foster, Rangarajan, & Parvizi, 2013). Yet, number words are typically used in language to express precise meaning (i.e., quantity). How can, therefore, a nuanced and approximate quantity representation, emerging in linguistic distributions, live together with the intuition of precise, well-defined meanings for numerals? Can our explicit (i.e., the exact meaning of words we are aware of) and implicit (i.e., the way we use words) understanding of number words differ so drastically? These are some of the unsolved and fascinating questions our study is raising. Paradoxically, whereas our data extend the link between the ANS and symbolic number processing one step further, by showing that signatures from the former system can be found in the usage of number words, we still miss a causal and direct account of how this may be possible.

Finally, our work has broad implications for the intense debate surrounding the evolution of the language faculty. Here we have demonstrated that numerical symbols obey Weber's law in natural language, adding to recent evidence that language has evolved to encode perceptual information and physical relations in the world (Louwerse, 2011). We conclude that the way the brain handles information about the environment through sensory systems profoundly affects how language itself is organized.

Acknowledgments

This work was supported by the Regione Lombardia - Fondazione Cariplo (grant 2017-1633 to M.M.). We are grateful to S. Amenta, M. Brysbaert, W. Fias, R.C. Gallistel, F. Guenther, L. Girelli, A. Henik, G. Magri, C. Semenza and N. Stucchi for extremely useful comments on a draft of this paper.

Data availability

The data used in this paper are available online as Supplemental Material.

References

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, *9*, 278–291. https://doi.org/10.1038/nrn2334

Barnard, A. M., Hughes, K. D., Gerhardt, R. R., DiVincenti, L., Bovee, J. M., & Cantlon, J. F. (2013). Inherently analog quantity representations in olive baboons (Papio anubis). *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00253

Barner, D. (2017). Language, procedures, & the non-perceptual origin of number word meanings. *Journal of Child Language*, *44*(3), 553–590. https://doi.org/10.1017/S0305000917000058

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 238–247). https://doi.org/10.3115/v1/P14-1023

Baroni, M., & Lenci, A. (2010). Distributional Memory: A General Framework for Corpus-Based Semantics. *Computational Linguistics*, *36*(4), 673–721. https://doi.org/10.1162/coli_a_00016

Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences*, *102*(39), 14116–14121. https://doi.org/10.1073/pnas.0505512102

Barth, Hilary, La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*(3), 199–222. https://doi.org/10.1016/j.cognition.2004.09.011

Beller, S., Bender, A., Chrisomalis, S., Jordan, F. M., Overmann, K. A., Saxe, G. B., & Schlimm, D. (2018). The cultural challenge in mathematical cognition. *Journal of Numerical Cognition*, *4*(2), 448–463.

Bender, A., & Beller, S. (2012). Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition*, *142*(2), 156–182. https://doi.org/10.1016/j.cognition.2012.05.005

Bhatia, S. (2017a). Associative judgment and vector space semantics. *Psychological Review*, *124*(1), 1–20. https://doi.org/10.1037/rev0000047

Bhatia, S. (2017b). The semantic representation of prejudice and stereotypes. *Cognition*, *174*, 46–60. https://doi.org/10.1016/j.cognition.2017.03.016

Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language*, *24*(3), 511–533. https://doi.org/10.1017/S0305000997003188

Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, *49*, 1–47. https://doi.org/10.1613/jair.4135

Brysbaert, M. (1995). Arabic Number Reading: On the Nature of the Numerical Scale and the Origin of Phonological Recoding. *Journal of Experimental Psychology: General*, *124*(4), 434–452. https://doi.org/10.1037/0096-3445.124.4.434

Brysbaert, M. (2018). Numbers and Language: What's New in the Past 25Years? In *Heterogeneity of Function in Numerical Cognition* (pp. 3–26). Elsevier.

Bullinaria, J. A., & Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: Stop-lists, stemming, and SVD. *Behavior Research Methods*, *44*(3), 890–907. https://doi.org/10.3758/s13428-011-0183-8

Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2015). Visual Number Beats Abstract Numerical Magnitude: Format-dependent Representation of Arabic Digits and Dot Patterns in Human Parietal Cortex. *Journal of Cognitive Neuroscience*, *27*(7), 1376–1387. https://doi.org/10.1162/jocn_a_00787

Butterworth, B. (1999). *The mathematical brain*. London, UK: Macmillan.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Calude, A. S., & Verkerk, A. (2016). The typology and diachrony of higher numerals in Indo-European: a phylogenetic comparative study. *Journal of Language Evolution*, *1*(2), 91–108.

https://doi.org/10.1093/jole/lzw003

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*(5), 401–406. https://doi.org/10.1111/j.1467-9280.2006.01719.x

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, *4*(5), 844–854. https://doi.org/10.1371/journal.pbio.0040125

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*(2), 83–91. https://doi.org/10.1016/j.tics.2008.11.007

Carey, S. (2009). *The Origin of Concepts*. New York: Oxford Univ. Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Carey, S., Shusterman, A., Haward, P., & Distefano, R. (2017). Do analog number representations underlie the meanings of young children's verbal numerals? *Cognition*, *168*, 243–255. https://doi.org/10.1016/j.cognition.2017.06.022

Chrisomalis, S. (2010). *Numerical notation: A comparative history*. *Numerical Notation: A Comparative History*. https://doi.org/10.1017/CBO9780511676062

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, *84*(2), 132–147. https://doi.org/10.1016/j.pneurobio.2007.11.001

Cohen Kadosh, R., Tzelgov, J., & Henik, A. (2008). A synesthetic walk on the mental number line: The size effect. *Cognition*, *106*(1), 548–557. https://doi.org/10.1016/j.cognition.2006.12.007

Cordes, S., & Gelman, R. (2005). The young numerical mind: When does it count? In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 127–142).

Damarla, S. R., & Just, M. A. (2013). Decoding the representation of numerical values from brain activation patterns. *Human Brain Mapping*, *34*(10), 2624–2634.

https://doi.org/10.1002/hbm.22087

Dastjerdi, M., Ozker, M., Foster, B. L., Rangarajan, V., & Parvizi, J. (2013). Numerical processing in the human parietal cortex during experimental and natural conditions. *Nature Communications*, *4*, 2528. https://doi.org/10.1038/ncomms3528

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*, 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S. (2001). Subtracting pigeons: logarithmic or linear? *Psychological Science*, *12*(3), 244–246.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147. https://doi.org/10.1016/S1364-6613(03)00055-X

Dehaene, S., & Changeux, J. P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, *5*(4), 390–407. https://doi.org/10.1162/jocn.1993.5.4.390

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in western and Amazonian indigene cultures. *Science*, *320*(5880), 1217–1220. https://doi.org/10.1126/science.1156540

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29. https://doi.org/10.1016/0010-0277(92)90030-L

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00068

Diester, I., & Nieder, A. (2010). Numerical Values Leave a Semantic Imprint on Associated Signs in Monkeys. *Journal of Cognitive Neuroscience*, *22*(1), 174–183. https://doi.org/10.1162/jocn.2009.21193

Ditz, H. M., & Nieder, A. (2016). Numerosity representations in crows obey the Weber–Fechner law. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1827), 20160083. https://doi.org/10.1098/rspb.2016.0083

Eger, E., Sterzer, P., Russ, M. O., Giraud, A. L., & Kleinschmidt, A. (2003). A supramodal number representation in human intraparietal cortex. *Neuron*, *37*(4), 719–726. https://doi.org/10.1016/S0896-6273(03)00036-9

Epps, P., Bowern, C., Hansen, C. A., Hill, J. H., & Zentz, J. (2012). On numeral complexity in hunter-gatherer languages. *Linguistic Typology*, *16*(1), 41–109. https://doi.org/10.1515/lingty-2012-0002

Everett, C. (2017). *Numbers and the making of us: counting and the course of human cultures*. Harvard University Press.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal Representation of Symbolic and Nonsymbolic Magnitude. *Journal of Cognitive Neuroscience*, *15*(1), 47–56. https://doi.org/10.1162/089892903321107819

Gallistel, C. R., & Gelman, R. (1992). Preverbal counting and computation. *Cognition*, *44*, 43–74. https://doi.org/10.1016/0010-0277(92)90050-R

Garrett, M. F. (1989). Processes in language production. In *Linguistics: The Cambridge Survey III. Language, Psychological and Biological Aspects.* (pp. 69–96). https://doi.org/10.1017/cbo9780511621062.004

Gelman, R. (2011). The case for continuity. *Behavioral and Brain Sciences*, *34*(3), 127–128. https://doi.org/10.1017/S0140525X10002712

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences*, *9*(1), 6–10. https://doi.org/10.1016/j.tics.2004.11.004

Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number*. Cambridge, MA:

Harvard University Press.

Gelman, R., & Gallistel, C. R. (2004). Language and the origin of numerical concepts. *Science*, *306*, 441–443. https://doi.org/10.1126/science.1105144

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*(3), B87–B95. https://doi.org/10.1016/S0010-0277(02)00234-2

Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organised: Evidence from days of the week. *Cortex*, *40*, 171–172.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*(3), 394–406. https://doi.org/10.1016/j.cognition.2010.02.002

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905–922. https://doi.org/10.1016/j.cortex.2011.04.010

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*, 496–499. https://doi.org/10.1126/science.1094492

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244. https://doi.org/10.1037/0033-295X.114.2.211

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*.

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668. https://doi.org/10.1038/nature07246

Hammarström, H. (2010). Rarities in numeral systems. *Rethinking Universals: How Rarities Affect Linguistic Theory*, *45*, 11–53.

Harris, Z. S. (1954). Distributional Structure. *Word*, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

Holloway, I. D., Price, G. R., & Ansari, D. (2010). Common and segregated neural pathways for the processing of symbolic and nonsymbolic numerical magnitude: An fMRI study. *NeuroImage*, *49*(1), 1006–1017. https://doi.org/10.1016/j.neuroimage.2009.07.071

Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': Numerical estimation by 5-7-year-olds. *Cognition*, *78*(3), B27–B40. https://doi.org/10.1016/S0010-0277(00)00122-0

Hutchinson, S., & Louwerse, M. M. (2014). Language statistics explain the spatial-numerical association of response codes. *Psychonomic Bulletin and Review*, *21*(2), 470–478. https://doi.org/10.3758/s13423-013-0492-2

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–1247. https://doi.org/10.1016/j.cognition.2007.06.004

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *204*(4), 824–843. https://doi.org/10.1037/0278-7393.20.4.824

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. https://doi.org/10.1037/0033-295X.114.1.1

Jones, M. N., Willits, J., & Dennis, S. (2015). Models of Semantic Memory. In *Oxford handbook of mathematical and computational psychology* (pp. 232–254). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199957996.013.11

Kanjlia, S., Feigenson, L., & Bedny, M. (2018). Numerical cognition is resilient to dramatic changes in early sensory experience. *Cognition*, *179*, 111–120. https://doi.org/10.1016/j.cognition.2018.06.004

Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., … Ischebeck, A. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: An event-related fMRI study. *NeuroImage*, *25*(3), 888–898. https://doi.org/10.1016/j.neuroimage.2004.12.041

Krajcsi, A. (2017). Numerical distance and size effects dissociate in Indo-Arabic number comparison. *Psychonomic Bulletin and Review*, *24*(3), 927–934. https://doi.org/10.3758/s13423-016-1175-6

Krajcsi, A., Lengyel, G., & Kojouharova, P. (2016). The source of the symbolic numerical distance and size effects. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01795

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Laurence, S., & Margolis, E. (2005). Number and Natural Language. In *The Innate Mind: Structure and Contents* (pp. 216–238). https://doi.org/10.1093/acprof:oso/9780195179675.003.0013

Le Corre, M., & 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. https://doi.org/10.1016/j.cognition.2006.10.005

Leibovich, T., Ashkenazi, S., Rubinsten, O., & Henik, A. (2013). Comparative judgments of symbolic and non-symbolic stimuli yield different patterns of reaction times. *Acta Psychologica*, *144*(2), 308–315. https://doi.org/10.1016/j.actpsy.2013.07.010

Lenci, A. (2018). Distributional Models of Word Meaning. *Annual Review of Linguistics*, *4*(1), 151–171. https://doi.org/10.1146/annurev-linguistics-030514-125254

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196. https://doi.org/10.1016/S1364-6613(00)01471-6

Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental Science*, *13*(6), 900–906. https://doi.org/10.1111/j.1467-7687.2009.00948.x

Lipton, J. S., & Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development*, *76*(5), 978–988. https://doi.org/10.1111/j.1467-8624.2005.00891.x

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*(2), 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, *28*(2), 203–208. https://doi.org/10.3354/ame01683

Lyons, I. M., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping*, *36*(2), 475–488. https://doi.org/10.1002/hbm.22641

Lyons, I. M., Nuerk, H. C., & Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *Journal of Experimental Psychology: General*, *144*(5), 1021–1035. https://doi.org/10.1037/xge0000094

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Marelli, M. (2017). Word-embeddings Italian Semantic spaces: A semantic model for psycholinguistic research. *Psihologija*, *50*(4), 503–520. https://doi.org/10.2298/PSI161208011M

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, *122*(3), 485–515. https://doi.org/10.1037/a0039267

Margolis, E., & Laurence, S. (2008). How to learn the natural numbers: Inductive inference and the acquisition of number concepts. *Cognition*, *106*(2), 924–939. https://doi.org/10.1016/j.cognition.2007.03.003

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, *6*(9). https://doi.org/10.1371/journal.pone.0023749

Menninger, K. (1969). *Number Words and Number Symbols: a Cultural History of Numbers*. Cambridge, MA: MIT Press.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *ArXiv Preprint ArXiv:1301.3781*, 1–12. https://doi.org/10.1162/153244303322533223

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520. https://doi.org/10.1038/2151519a0

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2016). How approximate and exact number skills are related to each other across development: A review. *Developmental Review*, *39*, 1–15. https://doi.org/10.1016/j.dr.2014.11.001

Nickerson, R. S. (1988). Counting, computing, and the representation of numbers. *Human Factors*, *30*(2), 181–199.

Nieder, A., & Dehaene, S. (2009). Representation of Number in the Brain. *Annual Review of Neuroscience*, *32*(1), 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Núñez, R. (2009). Numbers and Arithmetic: Neither Hardwired Nor Out There. *Biological Theory*, *4*(1), 68–83. https://doi.org/10.1162/biot.2009.4.1.68

Núñez, R. E. (2017). Is There Really an Evolved Capacity for Number? *Trends in Cognitive Sciences*, *21*(6), 409–424. https://doi.org/10.1016/j.tics.2017.03.005

Osgood, C. E. (1971). Where do sentences come from? In *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology* (pp. 497–529). Cambridge, MA: Cambridge University Press.

Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*(7163), 717–720. https://doi.org/10.1038/nature06176

Pagel, M., & Meade, A. (2017). The deep history of the number words. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1740), 1–9. https://doi.org/10.1098/rstb.2016.0517

Park, J., & Brannon, E. M. (2013). Training the Approximate Number System Improves Math Proficiency. *Psychological Science*, *24*(10), 2013–2019. https://doi.org/10.1177/0956797613482944

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. https://doi.org/10.3115/v1/D14-1162

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin and Review*, *21*(5), 1112–1130. https://doi.org/10.3758/s13423-014-0585-6

Piantadosi, S. T. (2016). A rational analysis of the approximate number system. *Psychonomic Bulletin and Review*, *23*(3), 877–886. https://doi.org/10.3758/s13423-015-0963-8

Piantadosi, S. T., & Cantlon, J. F. (2017). True Numerical Cognition in the Wild. *Psychological Science*, *28*(4), 462–469. https://doi.org/10.1177/0956797616686862

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, *123*(2), 199–217. https://doi.org/10.1016/j.cognition.2011.11.005

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*(12), 542–551. https://doi.org/10.1016/j.tics.2010.09.008

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. https://doi.org/10.1016/j.neuron.2004.10.014

Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education Enhances the Acuity of the Nonverbal Approximate Number System. *Psychological Science*, *24*(6), 1037–1043. https://doi.org/10.1177/0956797612464057

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/s0140525x12001495

Previtali, P., De Hevia, M. D., & Girelli, L. (2010). Placing order in space: The SNARC effect in serial learning. *Experimental Brain Research*, *201*(3), 599–605. https://doi.org/10.1007/s00221-009-2063-3

Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica*, *140*(1), 50–57. https://doi.org/10.1016/j.actpsy.2012.02.008

Ramscar, 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). https://doi.org/10.1371/journal.pone.0022501

Ramscar, 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. https://doi.org/10.1111/j.1551-6709.2009.01092.x

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, *83*(2), 278–278. https://doi.org/10.1037/h0028573

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*, *20*, 33–53.

Saxe, G. B. (2015). *Culture and cognitive development: Studies in mathematical understanding*. Psychology Press.

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Infants show ratio-dependent number discrimination regardless of set size. *Infancy*, *18*(6), 927–941. https://doi.org/10.1111/infa.12008

Sullivan, J., & Barner, D. (2014). Inference and association in children's early numerical estimation. *Child Development*, *85*(4), 1740–1755. https://doi.org/10.1111/cdev.12211

Tang, Y., Zhang, W., Chen, K., Feng, S., Ji, Y., Shen, J., … Liu, Y. (2006). Arithmetic processing in the brain shaped by cultures. *Proceedings of the National Academy of Sciences*, *103*(8), 10775–10780. https://doi.org/10.1073/pnas.0604416103

Temple, E., & Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(13), 7836–7841. https://doi.org/10.1073/pnas.95.13.7836

Verguts, T., & Fias, W. (2004). Representation of Number in Animals and Humans: A Neural Model. *Journal of Cognitive Neuroscience*, *16*(9), 1493–1504. https://doi.org/10.1162/0898929042568497

Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin and Review*, *12*(1), 66–80. https://doi.org/10.3758/BF03196349

Wassmann, J., & Dasen, P. R. (1994). Yupno number system and counting. *Journal of Cross-*

Cultural Psychology, *25*(1), 78–94. https://doi.org/10.1177/0022022194251005

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal Counting in Humans: The Psychophysics of Number Representation. *Psychological Science*, *10*(2), 130–137. https://doi.org/10.1111/1467-9280.00120

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701–702. https://doi.org/10.1038/nn1263

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. https://doi.org/10.1016/S0010-0277(99)00066-9

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, *8*(1), 88–101. https://doi.org/10.1111/j.1467-7687.2005.00395.x

Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition*, *57*, 271–295. https://doi.org/10.1016/0010-0277(95)00674-3

Figure Legends

Fig. 1. Extending the core signatures of the preverbal numerical system to the level of spontaneous verbal output. (A) An exemplified depiction of how current models describe number processing at the input level (i.e., please note that the direct connection between symbolic processing and the ANS has been questioned by some studies). In a standard numerical comparison task, participants are asked to indicate the smaller/larger quantity of two visually presented number words or visual arrays, respectively. Dashed arrows indicate the input process, while ovals the stages at which numbers are manipulated as written words or visual information, respectively. These input-dependent representations would be then scaled, as described by Weber's law, into internal representations in the ANS. Accordingly, reaction times in discriminating two perceptually presented quantities have been consistently shown to depend on numerical ratio. (B) A graphical depiction of the model proposed and tested, extending the core signatures of the ANS to the output level. According to this model, the ANS would superimpose its specific structure to the way humans use number words in language production and would, thus, regulate both perceptual input and verbal output. This evidence would favor the view that numerical symbols are grounded in the ANS. Notably, and differently from the input level, such ANS signatures would be detectable even in the absence of any comparative process.

Fig. 2. Numerical distance in the vector-space model. (A) Heatmap visualization of the individual linguistic vector distance (VD) between each pair of number words from *two* to *twenty* in the vector-space model. Lighter gradients of color indicate more vector similarity between pairs of number words, as extracted by the computational model. **(B)** The positive correlation between VD and real numerical distances. Results mirror the numerical distance effect typically found in numerical comparison tasks, i.e., an inverse relationship between response times and the distance between two

numbers to be compared. This means that one of the main hallmarks of the ANS (i.e., distance effect) is encoded in spontaneous language statistics.

Fig. 3. Numerical magnitude in the vector-space model. The positive correlation between language-based vector variance (VV) and real numerical magnitude. VV is a measure of the precision of the mental organization of each number word in the vector-space model. Specifically, higher VVs are associated with less specific linguistic contexts and, thus, to less accurate mental representations. The significant correlation between VV and numerical magnitude replicates evidence from psychophysical studies, indicating that the representation of larger number words in the distributional model is more approximate and noisy than smaller numerals.

Fig. 4. Ratio dependency of number words in different languages. Language-based predicted values as a function of real numerical ratio, plotted separately for each language tested. Predicted values were obtained from the linear regression model testing the effects of vector distance (VD) and vector variance (VV) on the numerical ratio. Results showed that model estimates are strongly associated with numerical ratios across all the languages tested. The ratio dependency (expressed in Weber's law) found in language statistics suggests an activation of the ANS whenever we use number words in everyday spontaneous language.

Fig. 1

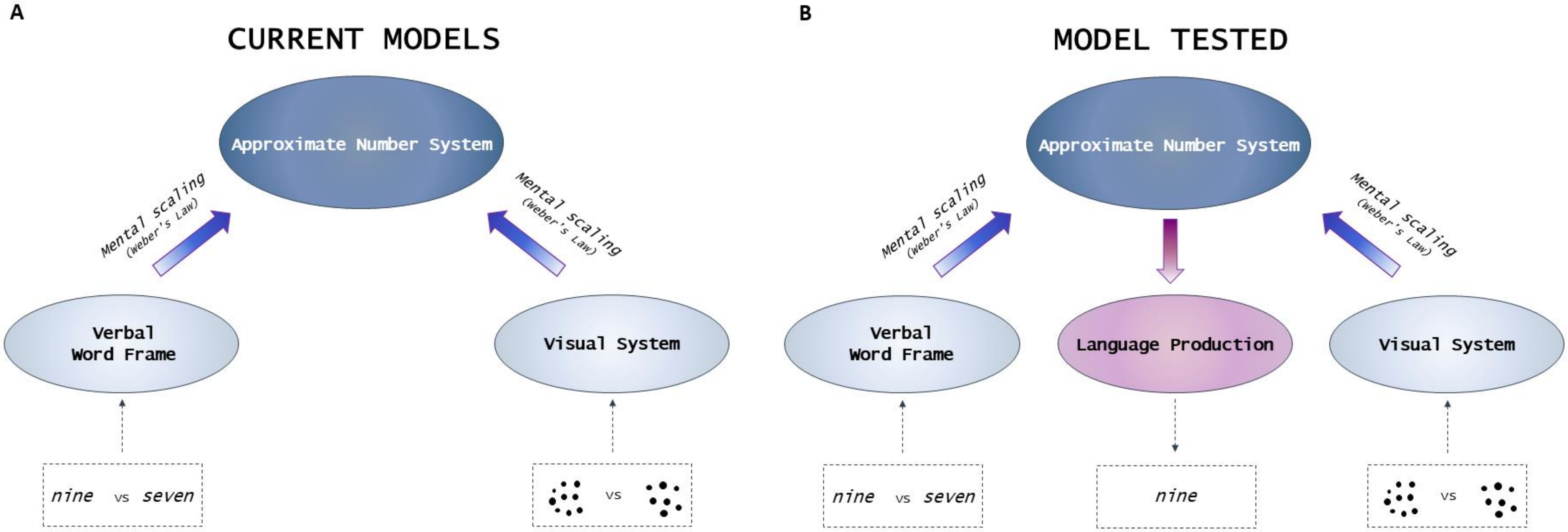

A
CURRENT MODELS
Approximate Number System
Mental scaling
(Weber's Law)
Mental scaling
(Weber's Law)
Verbal
Word Frame
Visual System
nine vs seven
vs
B
MODEL TESTED
Approximate Number System
Mental scaling
(Weber's Law)
Mental scaling
(Weber's Law)
Verbal
Word Frame
Language Production
Visual System
nine vs seven
nine
vs

Fig. 2

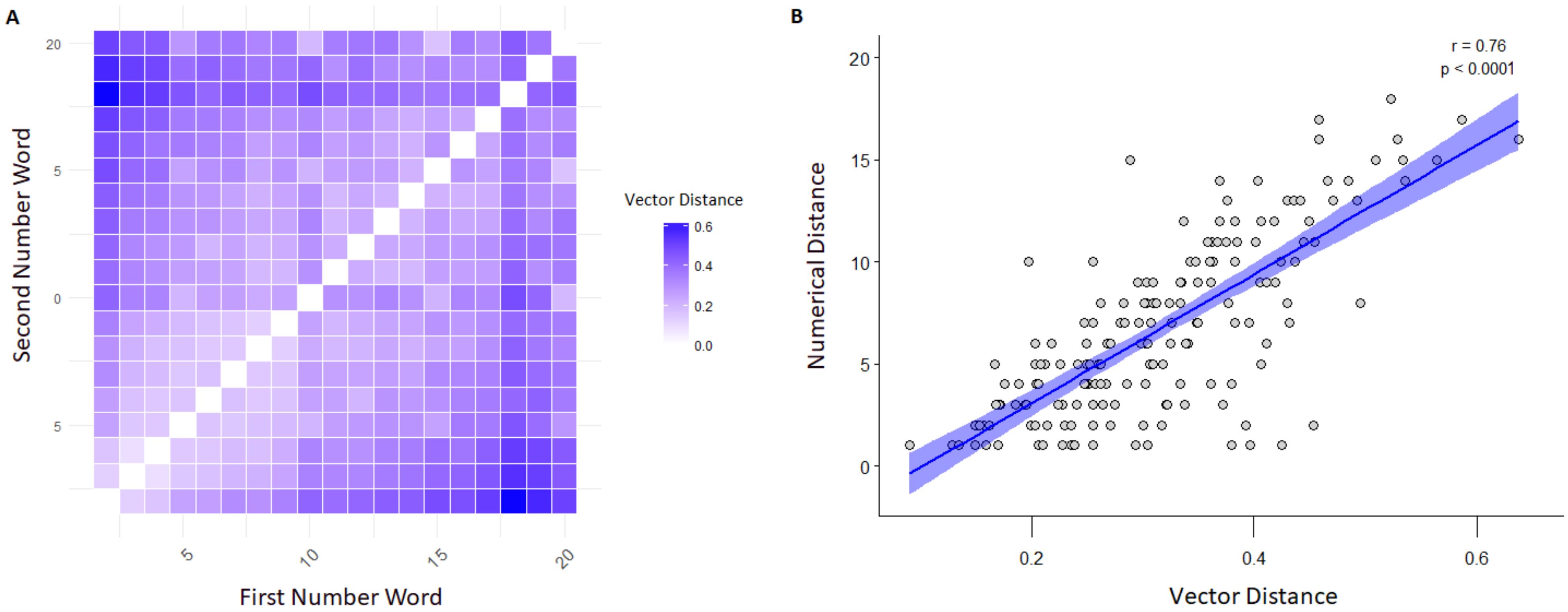

A
Second Number Word
First Number Word
Vector Distance
0.6
0.4
0.2
0.0
B
r = 0.76
p < 0.0001
Numerical Distance
Vector Distance

Fig. 3

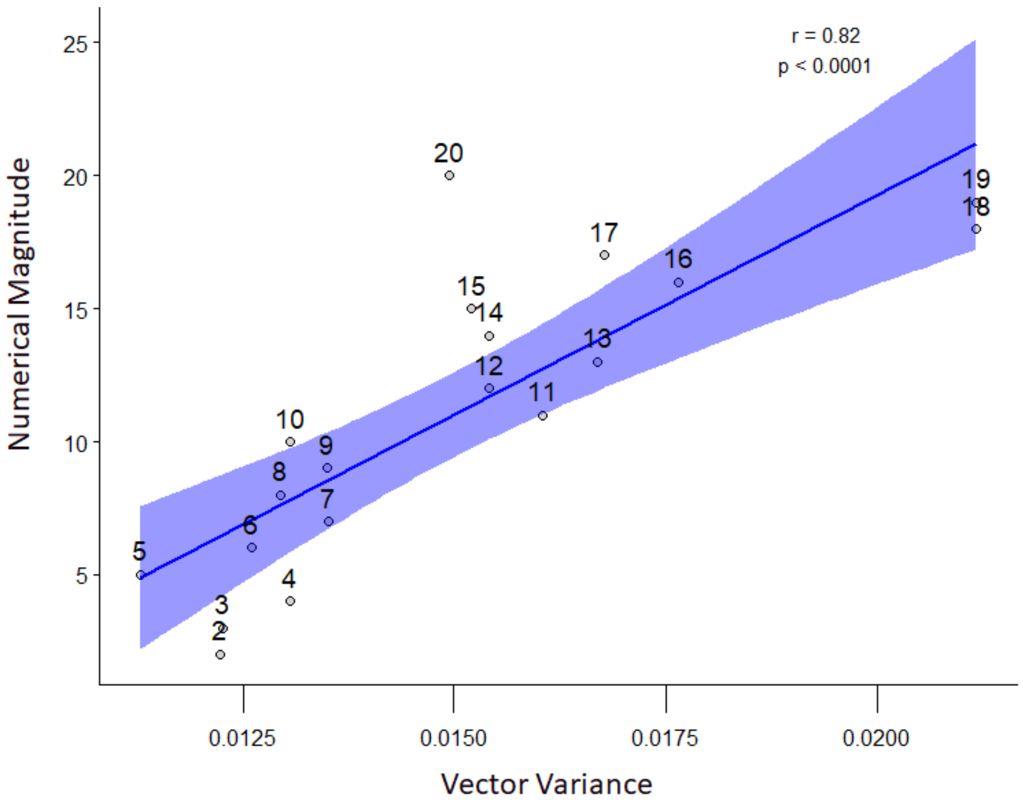
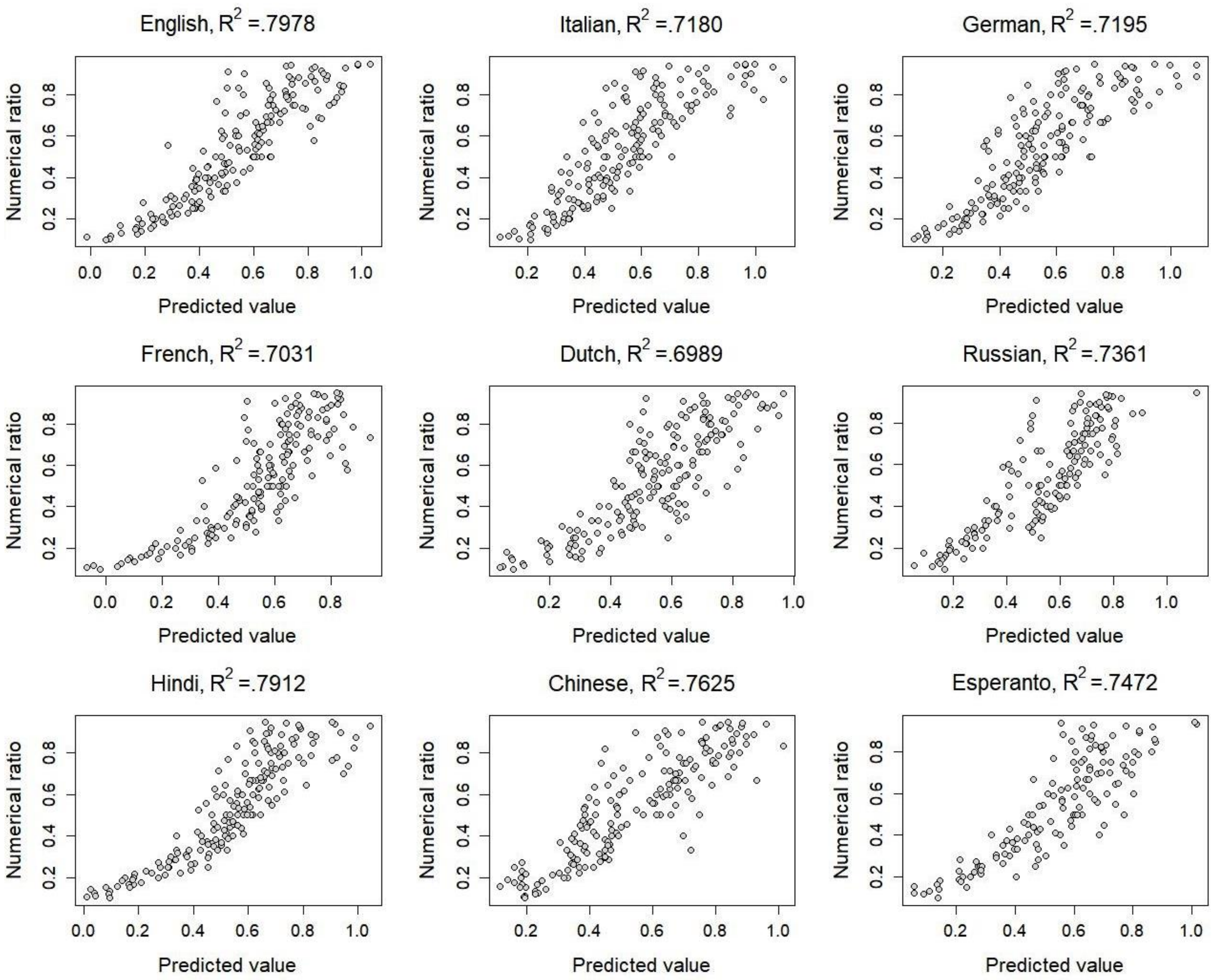

r = 0.82
p < 0.0001
Numerical Magnitude
Vector Variance
25
20
15
10
5
0.0125
0.0150
0.0175
0.0200
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Fig. 4


English, R² =.7978
Italian, R² =.7180
German, R² =.7195
French, R² =.7031
Dutch, R² =.6989
Russian, R² =.7361
Hindi, R² =.7912
Chinese, R² =.7625
Esperanto, R² =.7472
Numerical ratio
Predicted value
0.0
0.2
0.4
0.6
0.8
1.0