Native language effect on number cognition

Nian Liu
University of Oklahoma

Abstract
The present study tends to test the relationship between native language and its speakers’ mathematical processing, specifically, the relationship between linguistic numeral expression and the numeral processing among Chinese, English and Thai speakers. The study discovers that the speakers from different native language backgrounds have different performance on numeral processing with accordance to the linguistic numeral term their native language provided. Therefore, numeral language characteristics have a significant effect on cognitive representation of numbers.

Introduction
Human languages differ from one another in considerable and innumerable ways in how they describe the world. There is evidence that cross-linguistic difference in lexicon and grammar results in nonlinguistic consequences. In recent years, research on linguistic relativity has enjoyed a considerable resurgence, and much new evidence regarding the effects of language on thought has become available using new fine-tuned methods. Various behavioral studies have shown that even the little quirks of language influence speakers’ perception of objects (Boroditsky et al., 2003), substances (Imai & Gentner, 1997; Lucy et al. 2001), space (Levinson, 1996; Pederson et al., 1998) and time (Boroditsky, 2001; Casasanto & Boroditsky, 2008). This is known as linguistic relativity, also commonly referred to as the Sapir-Whorf hypothesis, the notion that linguistic structure affects non-linguistic cognition, such that speakers talk differently ends up think differently.

The current study tests a relatively understudied aspect, that is, whether languages affect the conceptualization of numbers and mathematical processing? Numbers and mathematics seems to be the most objective subject we deal with in daily life. However, although the vast majority of countries adopt the same convention by employing the Arabic notation when writing down a number, in detail the diversity of numeration syntax is striking across languages. For example, it has been reported that some aborigines in the Queensland district of Australia are still using base 2 system—Number 1 is “ganar”, 2 is “burla”, 3 “burla-ganar”, and 4 “burla-burla”. Base 20 also has its adepts in Eskimo and Yoruba. And the traces of it can still be found in French, in which 80 is “quatre-vingt” (four twenties). These languages form

1 The current Sapir-Whorf Hypothesis is the weak version. The strong version which claims that language determines minds (a.k.a. determinism) has long been abandoned.
contrasts with languages express numbers with base 10 place value. For example, East Asian language such as Chinese perfectly aligns with the decimal structure. In such a language, the name of a number is just the decomposition in base ten, thus 17 is “ten-seven” and 35 is “three-ten-five”, while in English the numbers 11 through 20 are denoted by special words (eleven, twelve, thirteen, twenty, etc.)

Does this exuberant diversity of numerical languages lead to practical consequences? Along with the stemming of cognitive linguistics, many an empirical work has been done in the last two decades to answer this question. There have been a number of studies on numerical cognition, showing that differences in number naming systems can affect cognitive development and non-linguistic performance. For example, acquisition studies (Miura, Okamoto, Kim, Steere, & Fayol, 1993; Miura et al., 1994; Miller, Smith, Zhu, & Zhang, 1995; Paik & Mix, 2003) have found that preschool-aged children whose native languages employ more systematic naming systems for their numbers outperform their counterparts who speak languages that use less transparent number naming systems on both number matching and number identification tasks. When asked to demonstrate numbers with combinations of individual unit cube block representing one and long blocks representing ten, Asian-language-speaking children who learned numerical names congruent with base 10 numeration systems (Fuson, 1990) were much more likely to use the blocks of 10 in constructing multi-digit numbers than their non-Asian-language-speaking counterparts, who lacked the access to such transparent numerical naming systems. For instance, when asked to represent number 12, Asian-language-speaking children were more likely to pick one long block and two individual blocks, while their non-Asian-language-speaking counterparts tended to choose twelve individual blocks. The authors of that study argued that “numerical language characteristics may have a significant effect on cognitive representation of numbers” (Miura et al., 1994, p. 410), which in turn may enhance the performance of Asian-language-speaking children on tasks involving the concept of place value.

The types of names given to various symbolic systems, such as numbers, have also been shown to affect the problem-solving abilities of competent symbol users. Seron and Fayol (1994) noticed that the verbal number system in French-speaking Belgium is simpler than the one used in France (in Belgium, 98 is roughly "ninety-eight" but in France, it is "four-twenty-eight"). They reported that second-grade children in France made more errors in Arabic number production than their Belgian counterparts. The effects of number naming systems also extend into adulthood and mathematical performance. For instance, one study showed that adult English speakers have difficulty reversing two-digit numbers ending in 1 (e.g., saying “14” when shown “41”), while Chinese speakers showed no such difficulty, presumably a result of English’s idiosyncratic rules for naming numbers between 11 and 19 (Miller & Zhu, 1991).

However, some other scholars (Saxon & Towse, 1997; Brysbaert et al., 1998) argued
that the influence of numerical language upon cognitive development is minimal and
criticized the methodologies that were previously used, claiming that the significance
of differences between different speakers’ performances in the experiments done by
Miura et al. may result from other factors rather than languages. In sum, the debate is
ongoing about whether differences in number naming systems affect the acquisition
and use of number concepts. Thus, it is meaningful to carry out further studies on the
relationship between numeral language characteristics and cognitive representation of
number.

The cognition of numbers
To pursue the line of language and numeral cognition research, the current study
focuses on the processing of large numbers. More specifically, I’d like to test whether
different linguistic expression of multi-digit numbers affects people’s processing of
these numbers.

Human being’s understanding of large numbers is very different from that of the low
numbers such as 1, 2 and 3. It has been known to psychologists for more than a
century that there is a strict limit on the number of objects that we are able to
enumerate at once (James McKeen Cattel, 1886). When the object number is beyond
three, errors accumulates. Also, the first three numbers are usually the ones that have
particular ordinal forms (E.g., “first”, “second”, “third” vs most ordinals end with “-th”
in English). Moreover, the numbers 1, 2, and 3 are also the only ones that can be
expressed by grammatical inflections instead of words (singular, dual, and trial in
some languages). The etymology of the first three numbers also tells us about the
limit of our numeral cognition to some degree. The Indo-European root of the word
“three” suggests that it might once be the largest number meaning “a lot” and “beyond
all others”. Examples can be found in French très (very), the Italian troppo (too much),
the English word through, and the Latin prefix trans- (Dehaene, 2011). On the other
hemisphere of the world, the word sân “three” in Chinese is used as synonym of
“many” or “all” in lots of idioms. Most interestingly, the structure of ideographic
Chinese characters representing the first three numbers in Chinese is similar,
consisting of one to three horizontal stocks respectively (一, 二, 三), while such
predictable rules cannot be found in expressing numbers bigger than three (for
example, 四 for “four” and 五 for “five”). In a word, it seems that our ancestors’
number sense might be confined to three. In fact, up to this very date, some Australian
aboriginal tribes were reported to have linguistic terms only for the quantities 1, 2,
some and a lot (Ifran, 1998; Gordon, 2004; Butterworth, Reeve, Reynolds & Lloyd,
2008). Pirahã, the Amazonian indigenous language, was arguably found to have no
words for precise numbers, but rather concepts for a small amount and a larger
amount (Everett, 1986; Frank et al., 2008). And this lack of number terms affects
Pirahã people’s performance in a few matching experiments conducted by
researchers.

Human languages move beyond three and linguistic terms assigned to greater
numbers are useful as they allow the speakers to remember and compare information about quantity accurately across space and time. David Hume (year, page) once noted that “I observe that when we mention any great number, such as a thousand, the mind has generally no adequate idea of it, but only a power of producing such an idea by its adequate idea of the decimals, under which the number is comprehended.” Therefore, humans generally have difficulty with large numbers and number terms are linguistic inventions that enable us to mentally represent large numbers to conquer the difficulty of making approximation.

In the era of the Information Revolution, we are facing the trend of using larger and larger numbers more frequently in the areas of astronomy to unit as small as CPU. And it’s interesting that world languages diverge again in expressing large numbers. In general, world languages employ two main systems in clustering multi-digit numerals—the tri-radical magnitude system and the tetra-radical one. The former one divides the number by a block of three digits at one time and the latter one divides multi-digit numbers by a block of four. Based on these two systems, Languages differentiate themselves in terms of the way to express different powers of ten².

For example, English employs the tri-radical magnitude system as most Indo-European languages does—thousand for $10^3$, million for $10^6$ and billion for $10^9$. In the naming of number terms, it is common that when no separate lexical items are available, the concepts are expressed by means of additive, deductive and multiplicative of other available number terms. For instance, in French, ninety is “four twenties plus ten” (quatre-vingt-dix), using the multiplicative and additive methods. So in the situation that no separate lexical terms are assigned to expressing $10^4$ or $10^5$, the language uses combined form of existing terms: the concept of “thousand” with the help of “ten” and “hundred” to form compound words, i.e. ten-thousand and a hundred thousand.

In the contrast, East Asian languages such as Chinese, Japanese and Korean use the tetra-radical system. In Chinese, for example, $10^4$ and $10^8$ are expressed by terms wàn 叁 and yì 亿, whereas $10^8$ is expressed as “hundred wàn” using multiplicative method by combining bǎi “hundred” and wàn “ten thousand”. It is not surprising to find a term of $10^3$ (qiān 千) in Chinese although it is generally a tetra-radical-system language, because a modern language would be inefficient without designing a term for the quantity of a thousand considering its frequent use. So the existence of qiān 千 does not contradict with the multi-digit clustering option that Chinese uses, as no separate lexical term for $10^6$ is found in Chinese. A comparison of the linguistic terms of large number is presented in Table 1 below, showing those Chinese and English provide linguistic terms for large numbers in a systematically different way.

---

² In this paper, scientific notations are used to record large numbers.
Interestingly, Thai makes use of a relatively unusual type of magnitude system: it names each of the multi-digit numeral up to $10^6$, instead of dividing them by blocks of three or four. Only numbers above a million are constructed by prefixing *lan* ล้าน with a multiplier. For example, ten million is สิบ ล้าน (ten *lan*).

It follows that in Thai there are more terms for multi-digit numerals than in English and Chinese. The differences of terms in magnitude system among the three languages are shown clearly in Table 2.

<table>
<thead>
<tr>
<th>Lg\magnitude</th>
<th>10^3</th>
<th>10^4</th>
<th>10^5</th>
<th>10^6</th>
<th>10^7</th>
<th>10^8</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English</strong></td>
<td>thousand</td>
<td>-</td>
<td>-</td>
<td>million</td>
<td>-</td>
<td>-</td>
<td>…</td>
</tr>
<tr>
<td><strong>Chinese</strong></td>
<td>千</td>
<td>万</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>亿</td>
<td>yi4</td>
</tr>
<tr>
<td><strong>Thai</strong></td>
<td>พัน</td>
<td>หมื่น</td>
<td>แสน</td>
<td>ล้าน</td>
<td>-</td>
<td>-</td>
<td>…</td>
</tr>
</tbody>
</table>

Table 2: Linguistic expression of magnitude of ten in English, Chinese and Thai

Linguist Roman Jakobson once noted that: “Languages differ essentially in what they *must* convey and not in what they *may* convey” (Jakobson and Halle 1956). Although all large numbers can be expressed in the three languages discussed above, the differences in available linguistic terms expressing large numbers may very subtly affect how their speakers process large numbers. Therefore, I hypothesize that different ways of linguistically coding magnitude of ten may influence people’s processing of the numbers. More specifically, the processing of these numbers will be
faster if they are coded in accordance to one’s native language. Speakers’ reaction time of reconstructing large numbers will be in accordance with the way how their native languages express them—they will be faster in reconstructing multi-digit numbers to the powers of ten for which their native language has a separate lexical term, and slower in which the native language does not provide a lexical term for. Specifically, Chinese speakers are supposed to show shorter response times in calculating $10^3$ and $10^4$ but longer response time in calculating $10^6$. In the contrast, English speakers will have better performance in the calculation of $10^3$ and $10^6$ than that of $10^4$. There should be no significant response time difference of Thai speakers across the powers of ten calculations as the language assigned linguistic expressions for each of the multi-digit numerals.

Two experiments are designed to test the hypothesis. Experiment 1 tests the idea that whether numeral language characteristics has an effect on cognitive representation of multi-digit numbers. And Experiment 2 tests the pure numeral processing without interference of language serving as baseline for comparison.

**Experiment 1**

**Participants:**
Participants were all college students studying at the University of Hawai‘i. Among them, 12 (6 female and 6 male) were native Chinese speakers, ranging in age from 18 to 31 ($M = 21$, $SD = 4.9$); 14 were native English speakers (8 female and 6 male), aged 18 to 29 ($M = 21$, $SD = 2.5$); and another 12 participants were native speakers of Thai (7 female and 5 male), aged 18 to 27 ($M = 21$, $SD = 1.9$). Participants’ language background was controlled as all of the native speakers of English did not speak Chinese, Japanese or Korean. Also, none of the Chinese or Thai speakers was exposed to English before the age of 12, which is reported to be the critical period of learning a second language. Moreover, all Chinese and Thai speakers have been living in the English-speaking country for less than two years. As for general mathematical abilities, none of the participants majored in mathematics, accounting or engineering, etc.

**Materials:**
Stimuli were questions requesting participants to vocally fill answers into blanks of multi-digit calculations. Sample questions are “Thirty-one thousand = ___ X $10^3$ (answer: 31)”, “Forty-five thousand = ___X $10^4$ (answer: 4.5)”, and “Two million and four hundred thousand = ___ X $10^6$ (answer: 2.4)”. All of numbers in the questions were shown in participants’ native language. Questions are divided in three conditions according to the magnitude of the ten, hence the third, fourth and sixth powers to ten respectively. In total thirty questions were created.

Another thirty filler questions were included in the test. The filler questions were calculation of time units such as “One and half hours = ___ minutes (answer: 90)” and “two weeks = ___days (answer: 14)”. The questions were presented in random
order.

**Procedures:**
In order to disguise the purpose of testing, participants were told that they were going to solve some mathematical problems.

Experiments were designed using E-Prime. In the experiment, participants initiated each trial by pressing the space bar. They then saw a question on a computer screen one at a time. They were instructed to speak the answer as quickly and accurately as possible into a microphone, which was used as a voice key, connected to an E-Prime SR-BOX. After detecting the vocal response, the trial automatically terminated and a fixation cross showed up in the center of the screen. Participants then pressed the space bar to proceed to the next questions when they were ready. Their reaction times were measured in milliseconds and used as independent variable in the analysis. Reaction times were measured from the appearance of the question to the onset of participants’ vocal response. Throughout the experiment a cassette recorder was used to record participant’s answers for later data coding.

Before a run, the experimenter made sure that participants understand the instruction and then they had an opportunity to practice with 6 trails. Microphone sensitivity was adjusted for each participant during this practice phase.

Participants were not allowed to use scratch papers but they were free to use fingers in calculation. In addition, they could not go back to previous questions even if they gave the wrong answer. All instructions were provided in their native language.

**Results**
All filler syllables, tongue clicks, false triggers (sneezes and coughs), and repeated responses (due to their failure to trigger the voice key) were targeted and excluded. Two of the English-group participants were excluded due to their failure of completing the test. Other than that, no participants were excluded because of outlying RTs or low accuracy.

**Intra-language Analysis** For English speakers, the reaction times for the three types of calculation differ significantly, as expected. As we can see from Figure 1, their performances were better in calculation of $10^3$ and $10^6$ than that of $10^4$. 
In order to see a more precise picture, the data were submitted to SPSS for further analysis. One way repeated measures ANOVA with Question Type (Magnitude) as independent variable showed that the response times in different answering different powers of ten questions differ significantly $F(2, 27) = 4.78, p < .05$. Pairwise comparisons showed that the response times for calculation of $10^3$ (marked as $E_3$) ($M=7921, SD=2152$) are significantly shorter than that of $10^4$ ($M=13394, SD=4369$), $t = -2.37, p < .05$. Also, their response times for calculation of $10^6$ (marked as $E_6$) ($M=8873, SD=2207$) is much shorter than that of $10^4$ ($E_4$), $t = -2.01, p < .05$. The type of questions that got shorter response times is with accordance to the linguistic term provided by English for powers of ten numbers.

Chinese speakers’ performance also reflects the pattern that complies with the magnitude system their native language employs: They performed the $10^6$ calculation slower than with the $10^3$ and $10^4$. The mean response times for answering the three types of questions are graphically represented in Figure 3.
The reaction times of Chinese speakers in the three types of calculation differ significantly, $F(2, 27)=3.025, p<.05$. They were slower in calculating $10^6$ (marked as C6) ($M=6881$, $SD=1330$) than in calculating $10^3$ (C3) ($M=4544$, $SD=678$), $t = -2.70$, $p < .05$. However, the response times for calculation of $10^4$ (C4) ($M=5287$, $SD=1679$) is not significantly longer than that of $10^6$, $t = -1.28$, $p = .13$.

**Thai speakers**: We predict that there is no significant time difference among Thai speakers’ response time for three types of questions, however, the one-way ANOVA performed on data of Thai speakers shows $H_0$ is rejected as the reaction times were not significantly equal: $F(2, 27)=1.34$, $p=0.276$. A closer look at the results reveals that the Thai speakers were a little bit slower in calculating $10^6$ than with the other two types of calculations, yet the difference is not statistically significant. We can speculate that this difference may result from the longer response of processing numbers with more digits.

**Inter-language Analysis** We now turn our focus to comparisons among different language groups. The hypotheses are, first, because Chinese language provides its speakers with the separate linguistic numeral expression of $10^5$, which has no equivalent in English, Chinese speakers are supposed to solve the calculation of $10^4$ comparatively faster than English speakers do. And second, the lack of a separate linguistic term of $10^6$ in Chinese may cause comparatively longer response times of Chinese speakers in solving the $10^6$ questions compared with that of the English speakers.

In order to test the two hypotheses, I adopt the method of ratio comparison. Comparing the ratio, rather than the mean response time directly is because ultimately, any comparison between samples of participants drawn from distinct populations runs the risk of unknown confounds—whether those confounds be cultural, individual, sampling, or others. It’s desirable, therefore, to draw ratios from a single language group and then compare them cross-linguistically. This comparison method therefore serves to eliminate uncontrolled factors. Therefore, the ratio of C3/C4 (Chinese speakers’ reaction times of $10^3$ over $10^4$) and E3/E4 (English speakers’ reaction times of $10^3$ over $10^5$) is taken into consideration for the comparison purpose. Moreover, it is the ratio of C3/C4 and E3/E4 being compared rather than C3/E3 and C4/E4 being compared directly is resulted from the fact that C3/E3 is not normally distributed, so that is not suitable to undergo statistical analysis directly. Results shows that it is marginal significant that C3/C4 ($M=869$, $SD=420$) is bigger than E3/E4 ($M=591$, $SD=90$), $t = -1.96$, $p = .061$, confirming that Chinese speakers were comparatively faster in calculating $10^4$ than their English-speaking counterparts. On the other hand, there is a significant difference between C6/C4 ($M=1301$, $SD=140$) and E6/E4 ($M=6625$, $SD=80$), $t = -6.563$, $p < .01$, confirming that the lack of a separate linguistic term of $10^6$ caused slower responses of Chinese speakers in solving the $10^6$ questions compared with the English speakers.
Analysis: Both Chinese and English speakers performed better in the types of questions that corresponding to the linguistic numeral expressions they have in their native languages. More specifically, English speakers calculated $10^3$ and $10^6$ questions faster than $10^4$, corresponding to the fact that the concepts ‘thousand’ and ‘million’ are directly available in the language. Contrastingly, Chinese speakers were faster in solving questions of $10^3$ and $10^4$ than $10^6$, as the former two concepts were expressly explicitly in the language. However, the difference between response times of $10^4$ and $10^6$ were not statistically significant, i.e., Chinese speakers were not statistically significantly faster in solving $10^4$ questions than $10^6$ ones, though the language is equipped with a specific term of $10^4$, but not $10^6$. This may due to mathematical ability influence because processing numbers with more digits will naturally cause longer response time. And the long processing time may weaken or bleach out the difference resulting from linguistic influence. But we should note that the influence is still observable, because otherwise the calculation of $10^6$ would result in longer response time, which was not the case showed by the result. On the other hand, Thai speakers did not show any response time preference in calculating all three types of questions, as all of the three numeral concepts are expressed overtly in their native language. The results support the prediction that linguistic expression of numbers may influence people’s processing of these numbers. Through daily use, people get used to cluster multi-digit numerals according to the way their native language describe it. So if people were asked to perform tasks involving reconstruct multi-digit numbers were in a way they are not familiar with, extra steps may be needed in the processing—they have to “translate” the number from linguistic terms to Arabic numerals prior to performing the calculation.

Experiment 2
Methods
Subjects: subjects were the same as in Experiment 1.

Materials and Procedures: Instructions were the same as in Experiment 1 except that they were given in the participants’ native languages.

Procedures: All the procedures were the same as in Experiment 1 except that the examples were given by Arabic numeral, such as: $39400 = \_ \times 10^3$ (answer: 39.4), $762500 = \_ \times 10^4$ (answer: 76.25) and $8620000 = \_ \times 10^6$ (answer: 8.62). Subjects were responsible for solving 24 questions of calculation (8 for each type) shown with Arabic numerals.

Results: Chinese speakers did not show any significant difference in their response time of solving the three types of questions, $F = 1.06, p = 0.17$. So as English speakers ($F = 1.45, p = 0.24$) and Thai speakers ($F = 0.23, p = 0.79$).

Analysis: The prediction was that there will be no significant difference in speakers’
response time of solving the three types of questions, because in this experiment the native language interference was excluded. Subjects tended to simply count the digit of multi-digit numerals instead of think about them in their native language when they were told to give answers as soon as possible. So the linguistic influence was eliminated in this experiment, which indicated that the numeral processing was supposed otherwise to be universally equal provided that there was no language interference.

**Discussion**

Interestingly, a similar question about language and thinking was asked by Albert Einstein as early as 1941. His discovery that one geometry could be as valid as another for mapping nature was a specialized case of the historical language-and-thinking problem. Einstein dealt with linguistic relativity in a little-known radio speech: “What is it that brings about such an ultimate connection between language and thinking? ...the mental development of the individual and his way of forming concepts depend to a high degree upon language. This makes us realize to what extent the same language means the same mentality (Einstein, 1954).”

So relativity, the philosophical language-and-thought question for thousands of years, was specialized by Einstein as a geometry-and-thought question for the philosophy of mathematics, which indicates the existence of language-mathematics relation theoretically from a scientist’s view.

The mathematical process is supposed to be universal because the value of numerals should not be affected by the way how people express it. But the discovery of differences in processing multi-digit numerals among different language speakers suggests that the different linguistic numeral expression will accelerate or obstruct our procession of numerals. Native language sets certain limits on its speakers’ recognition and cognition process by giving a certain direction and excluding many others.

The present study failed to exclude the non-linguistic influence on numeral processing but the hypothesis does not claim an exclusive role for language as a determinant of mathematical processing. Empirical investigations of linguistic relativity always face the difficult problem of demonstrating the specific contribution made by language in a particular cognition domain. We can work out better ways to test the worth-studying relation of native language and mathematical processing.
References


