Two Distinct Processes in Understanding Number Word Structure

Kexin Que (kexinque@u.northwestern.edu)
Department of Psychology,
Northwestern University, Evanston, IL 60208, USA

Lance Rips (rips@northwestern.edu)
Department of Psychology,
Northwestern University, Evanston, IL 60208, USA

Abstract

Hurford (1975) proposed that number words across languages follow a set of phrase structure rules (syntax), together with a constraint known as the packing strategy. The present study taught participants a novel numeral system to test whether people's mental representations of number words align with this framework. We tested speakers of English and Turkish, which contain irregularities relative to the syntax, and Chinese, which aligns closely with it. Participants' task was to generate large numbers (e.g., 27) in a base-3 artificial number system. In the training conditions, participants learned expressions consistent with both syntax and the packing strategy at different magnitudes (larger vs. smaller numbers); the baseline condition received no training. Across conditions, Chinese speakers were more likely than English or Turkish speakers to produce number words consistent with Hurford's proposal. This advantage was due to greater sensitivity to syntax, not packing. Across languages, training increased sensitivity to syntax relative to baseline, whereas sensitivity to packing did not differ between conditions. Together, these findings suggest that people represent the structure of number words through two distinct cognitive processes rather than through a single unified system.

Keywords: Number; Number-word syntax; Numerals; Packing strategy

Introduction

Number words exhibit a surprisingly similar structure across languages. For example, the Mandarin expression for 3300, 三千三百 (*sān qiān sān bǎi*), is directly parallel to the English *three thousand three hundred*. Similar patterns are found in Turkish (*üç bin üç yüz*), French (*trois mille trois cents*), and many other languages. Hurford (1975, 1987) examined the linguistic structure of number words and proposed that number words across languages universally conform to two principles: a set of phrase structure rules and an arithmetic constraint imposed on the syntax, known as the packing strategy.

The phrase structure rules for numerals describe a recursive internal structure analogous to syntactic structure in natural language. This system is based on four basic categories: Number, Phrase, Digit, and Multiplier, as Figure 1 shows. Number subsumes all number expressions (e.g., *three thousand three hundred*), including Digits (e.g., *one* to *nine* in English) or combinations of a Phrase and optionally another Number (Expression a & b). In turn, a Phrase is composed of a Multiplier (e.g., *-ty*, *hundred*, and *thousand* in English), which may optionally be preceded by a Number (Expression c).

Within this system, the meaning of number words depends on their syntactic structure and their component lexical items. Specifically, the magnitude of a Number is the sum of a Phrase and a Number, whereas the magnitude of a Phrase is the product of a Number and a Multiplier. For example, the meaning of *three thousand three hundred* (3300) is derived as *three thousand* (Phrase) plus *three hundred* (Number). The meaning of the Phrase *three thousand* is derived as *three* (Number) multiplied by *thousand* (Multiplier).

a. Number → Digit
b. Number → Phrase (Number)
c. Phrase → (Number) Multiplier

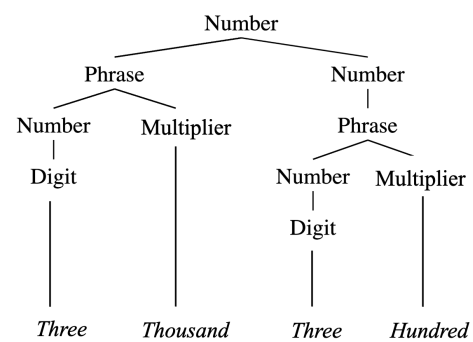

Figure 1: Expressions (a–c) illustrate the phrase structure rules proposed by Hurford (1975). The syntax tree shows the syntactic representation of the number word for the numeral 3300.

The phrase structure rules (syntax) governing number words are doubly recursive over the category Number and therefore tend to overgenerate. For example, although the phrase structure rules license *three hundred one thousand* [for 300 + 3000], it is not a conventional expression. To account for this, Hurford (1975, 1987) proposes a global constraint on the phrase structure rules, called the *packing strategy*. This constraint filters the output of the phrase structure rules, selecting a single well-formed expression for each numerical value while excluding alternative, ill-formed candidates. Formally, the packing strategy (hereafter, packing) requires that the sister of a Number must have the highest possible value.

Under this constraint, expressions such as *three hundred three thousand* are ruled out because, although syntactically well-formed on the pattern of Figure 1, they violate the requirement that the sister constituent of Number (here, a Phrase) denote the highest available value (Phrase = *three hundred* ≯ Number = *three thousand*). The same logic extends to multiplicative constructions. For instance, 300,000 cannot be expressed as *three thousand hundred*, despite conforming to the syntax, because the topmost multiplier (in this case, the second multiplier) does not represent the highest possible value.

Extensive work has characterized the structure of number words (Hurford, 1975, 1987). However, it remains unclear whether people represent this structure in line with Hurford's analysis; that is, through two distinct principles: syntax and packing. This question is important because, although many numeral systems have existed historically, number words are unique in having both a syntax and a syntactic constraint (e.g., Chrisomalis, 2004, 2010, 2020; Comrie, 2011, 2022; Hammarström, 2010). Understanding how people represent the structure of number words therefore contributes to the literature on the development of numeral systems (e.g., Denić & Szymanik, 2024; Liu & Xu, 2019; Prasertsom et al., 2026; Xu et al., 2020; Yang & Regier, 2025). In addition, recent work has increasingly examined the relation between number syntax and language syntax (Hurford, 2003; Ionin & Matushansky, 2006, 2018), as well as the processes underlying the acquisition of number-word structure (e.g., Chu et al., 2020; Guerrero & Park, 2023; Holt & Barner, 2023; 2025a, 2025b; Park et al., 2025; Silvi et al., 2024; Yuan et al., 2019). This body of work further motivates the need to understand how people represent the structure of number words.

Table 1. Number words in Chinese, English and Turkish

Numeral	Chinese	English	Turkish
1	Yi	One	Bir
2	Er	Two	İki
10	Shi	Ten	On
11	Shi Yi	Eleven	On Bir
12	Shi Er	Twelve	On Iki
20	Er Shi	Twenty	Yirmi
21	Er Shi Yi	Twenty One	Yirmi Bir
22	Er Shi Er	Twenty Two	Yirmi Iki

However, this question is complicated by the fact that some languages contain irregularities that deviate from the phrase structure rules (the numeral syntax in Figure 1), which may challenge the universality of people's representations of number words. For example, in English, the teen numbers are irregular or semi-regular, as their forms do not transparently follow the canonical Number structure. In Turkish, the decade names are irregular; as Table 1 shows, the word for twenty (*yirmi*) bears no transparent relation to either two (*iki*) or ten (*on*), and so violates the Phrase structure. In contrast, Mandarin number words adhere strictly to the phrase structure rules.

There are reasons to suspect that speakers with more transparent number syntax have a better understanding of the meaning and structure of number words. For example, early cross-linguistic work has shown that Chinese-speaking children tend to have a higher count range than English-speaking children (Ho & Fuson, 1998; Miller & Stigler, 1987; Miller et al., 1995). More recent studies suggest that children whose number language has more transparent morphological syntax develop an earlier understanding of base-10 compositionality and the idea that each successive numeral represents an increase of one in set size (e.g., Derzhanski & Veneva, 2018; Dowker et al., 2008; Guerrero et al., 2020; Miura & Okamoto, 2013; Schneider et al., 2020; Srinivasan, 2024; Topaloğlu, 2026). Consistent with this advantage, Derzhanski and Veneva (2018) found that features such as alternative base names, phonological changes, and word-order variation make it more difficult for secondary-school students to infer the structure and meaning of number words in unfamiliar languages.

Present Study

Our goal is to examine people's mental representations of the structure of number words. Our central question is whether these representations align with Hurford's framework—that is, whether number words are represented through syntax and packing as two separate processes. To do this, we designed a training study. We described to participants an artificial base-3 language and then asked them to generate number words in that language from an Arabic numeral prompt (e.g., to generate *Fim Gom Lod* from the prompt 27). The language included two digit terms, *Fim* and *Miv*, representing *one* and *two*, and two multiplier terms, *Gom* and *Lod*, representing *three* and *nine*, respectively. Participants in the training conditions learned expressions for either large or small magnitudes that were consistent with both syntax and packing; those in the baseline condition received no training.

Our hypothesis was that if syntax and packing are distinct processes, improvements in syntax should not necessarily be associated with improvements in packing. In contrast, if they reflect a single underlying process, improvements in syntax should accompany improvements in packing.

To examine the universality of people's representations, we recruited native speakers of Chinese, English, and Turkish, which differ in the transparency of their number syntax. Based on prior work (cited earlier), we predicted that Chinese speakers, whose number system closely aligns with the syntax of Figure 1, would show a stronger understanding of the syntactic structure of the novel system than would speakers of English or Turkish. Whether they also show a better understanding of packing would depend on how they represent the structure of number words. If syntax and packing are distinct processes, advantages in syntax should not necessarily extend to packing.

Methods

Materials and Procedures

To ensure that participants used their native language when thinking about numbers, we first asked them to complete a counting task in which they counted a picture of 10 bunnies. On the next page, they reported the language they had used. Only participants who spontaneously counted in their native language were allowed to proceed.

Next, participants were introduced to the number system of a newly discovered population called the “Numerons” (with a wink to Gelman & Gallistel, 1978). They learned that this system consisted of two components: digits and multipliers, both representing positive integers. Digits represented numbers up to the base value, analogous to the numbers *one* through *nine* in English. The instructions specified that *Fim* and *Miv* were the only two digits in the Numerons’ number language: *Fim* represented “*one*,” and *Miv* represented “*two*.” Multipliers were described as noun-like forms used in multiplicative contexts, similar to *thousand* or *billion* in English. *Gom* and *Lod* were introduced as multipliers, with *Gom* representing “*three*” and *Lod* representing “*nine*.” Therefore, the numerals form a base-3 number system. We explained that the Numerons’ number system formed complex numbers by combining these words. Participants’ task was to generate number words in the Numerons’ number system.

After the instructions, participants completed two comprehension checks to assess their knowledge of the novel words (e.g., identifying which two options were digit numbers in the Numerons’ number language). They received feedback on each question and could proceed to the next phase only if they passed both checks. All participants then proceeded to the practice trials, where the critical manipulation of the study occurred.

In the practice trials, participants used the four novel number words they had learned to produce number expressions in the Numerons’ language. The definitions of the categories and magnitudes of the novel number words *Fim, Miv, Gom*, and *Lod* were available to participants at all times (as in Figure 2). Participants in the baseline condition were asked to produce number expressions for 4, 6, 7, and 11. They received no feedback on their correctness. Participants in the small-number training condition produced the same numbers but received feedback on whether their responses were correct (conformed to both syntax and packing for the target number). They had up to three attempts on each trial. If they produced the correct expression within three attempts, they received confirmation of correctness. If they failed to produce the correct expression within these attempts, the correct form was displayed after the third attempt.

Participants in the large-number training condition received similar training, but on numbers 26, 33, 38, and 57. Crucially, this training exposed participants to all the structures (though not the exact items) required in the following test trials, whereas the small-number training did not. Participants in the both-training condition received training on both small and large numbers.

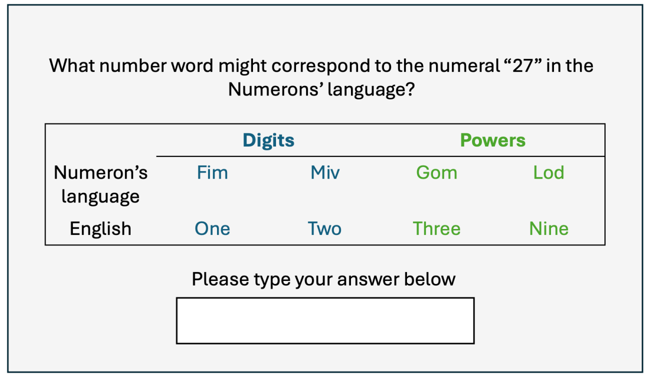

Figure 2: An example trial in the test phrase. Participants received an error message if their response did not include any of the four novel words. In the instructions, we used the term “powers” to refer to multipliers.

The test phase included 11 target numbers ranging from 12 to 56 (see Figure 2 for a sample trial). The test phase also included two catch trials in which participants were asked to write the expressions for 2 and 9 in the Numerons’ language. Trials were presented in a randomized order.

The initial expectation (as specified in the preregistration) for the catch trials was that participants would produce *Miv* for 2 and *Lod* for 9, as shown in the instruction. However, this criterion excluded more than a third of participants in the baseline condition. Upon reviewing their responses, we found that participants also produced alternatives such as *Fim Fim* and *Fim Miv* for 2, and *Gom Gom Gom* and *Fim Lod* for 9. The final exclusion criterion checked only for numeral 2, allowing *Miv Fim*, *Fim Miv*, *Fim Fim*, and *Miv* as acceptable expressions.

The experiment was implemented in Qualtrics and conducted in the participants’ native language. All instructions were first translated from English into Chinese and Turkish and then back-translated into English. The original and back-translated English versions were compared and discussed to ensure equivalence across languages.

Participants

The final sample included 100 L1 Chinese speakers (mean age = 32.3 years, SD = 7.8, range = 20–57), 109 L1 English speakers (mean age = 41.4 years, SD = 11.6, range = 19–71), and 102 L1 Turkish speakers (mean age = 32.4 years, SD = 7.5, range = 18–66). Participants were distributed across four training conditions (small, large, both, and no training), with approximately 20 participants per language in each training condition, and 40 participants in the baseline condition. Participants reported how many college-level mathematics courses they had taken. On average, Chinese speakers reported 3.7 courses (SD = 4.7), English speakers 3.0 courses (SD = 3.2), and Turkish speakers 4.9 courses (SD = 5.9). We excluded 11 participants who failed the catch trial. All participants were recruited through Prolific.

Analyses and Results

We coded responses, as pre-registered, into three categories: (1) expressions conforming to both syntax and packing, (2) expressions conforming to syntax only, and (3) expressions conforming to neither. We used a dynamic programming procedure that implemented either both syntax and packing or syntax only to generate all possible expressions consistent with categories (1) and (2), and we coded participants' responses using these lists. Only one expression was in category (1) for each numeral, but there could be multiple expressions in category (2). To constrain the solution space, generated expressions were limited to fewer than 16 elements; no participant responses exceeded this limit. Responses that did not fit either category were coded as category (3).

Full (Syntax+Packing) alignment. We first conducted the preregistered analysis examining participants' conformity to both syntax and packing. A response was scored as 1 in the full-alignment measure if it was coded as category (1), and 0 otherwise. Overall scores for language and training conditions appear as purple lines in Figure 3. We fit a mixed-effects logistic model predicting full alignment, with fixed effects of language contrasts, condition contrasts, and their interactions. For language, we specified two contrasts: a between-transparency contrast comparing Chinese speakers to the combined group of English and Turkish speakers, and a within-transparency contrast comparing English and Turkish speakers. For condition, we included three contrasts: (1) baseline versus all training conditions, (2) training on small numbers versus large numbers, and (3) training on either small or large numbers versus both. All predictors were sum-coded. Random effects included random intercepts for participants and trials.

As predicted, Chinese speakers were more likely than English and Turkish speakers to produce fully aligned expressions ($\beta = 2.68$, $SE = 0.45$, $z = 5.93$, $p < .001$). Turkish speakers showed higher alignment than English speakers ($\beta = -2.70$, $SE = 0.52$, $z = -5.18$, $p < .001$). Across language groups, training predicted higher alignment ($\beta = 3.80$, $SE = 0.47$, $z = 8.17$, $p < .001$), though the magnitude of numbers used in the training did not make a significant difference (small vs. large: $\beta = 0.95$, $SE = 0.61$, $z = 1.55$, $p = .121$; small-or-large vs. both: $\beta = 0.70$, $SE = 0.53$, $z = 1.32$, $p = .188$). We also observed no significant interactions between training and language (all *ps* > .34).

When math course was added as a covariate, the pattern of results remained unchanged, and math course did not significantly predict alignment ($\beta = 0.26$, $SE = 0.20$, $z = 1.30$, $p = .193$).

Syntax alignment[1]. We next examined the effect of language and training on the likelihood of producing syntax-aligned expressions. A response was coded as 1 in the syntax-alignment measure if it was classified as either category (1) or (2), and 0 otherwise. These scores appear as green lines in Figure 3. We fit a mixed-effects logistic model identical to the one predicting full alignment, replacing only the dependent variable.

Consistent with the pattern observed for full alignment, Chinese speakers were more likely than English or Turkish speakers to produce syntax-aligned expressions ($\beta = 3.90$, $SE = 0.63$, $z = 6.22$, $p < .001$). Turkish speakers again showed higher alignment than English speakers ($\beta = -3.32$, $SE = 0.65$, $z = -5.09$, $p < .001$). Training significantly predicted alignment across language groups ($\beta = 5.21$, $SE = 0.61$, $z = 8.55$, $p < .001$), and the magnitude of the training examples did not significantly affect alignment (small vs. large: $\beta = 0.50$, $SE = 0.80$, $z = 0.62$, $p = .536$; small-or-large vs. both: $\beta = 0.21$, $SE = 0.68$, $z = 0.31$, $p = .754$).

We noticed a significant interaction between Turkish (vs. English) and the small-or-large vs. both training contrast ($\beta = 3.18$, $SE = 1.61$, $z = 1.97$, $p = .049$), suggesting that training on both small and large numbers (vs. either alone) was more beneficial for English speakers than for Turkish speakers. However, when math background was added as a covariate, this interaction became marginally significant ($\beta = 3.09$, $SE = 1.61$, $z = 1.92$, $p = .054$). Math background was also marginally significant ($\beta = 0.41$, $SE = 0.25$, $z = 1.65$, $p = .098$). Other interactions between language and training were not significant (all *ps* > .10).

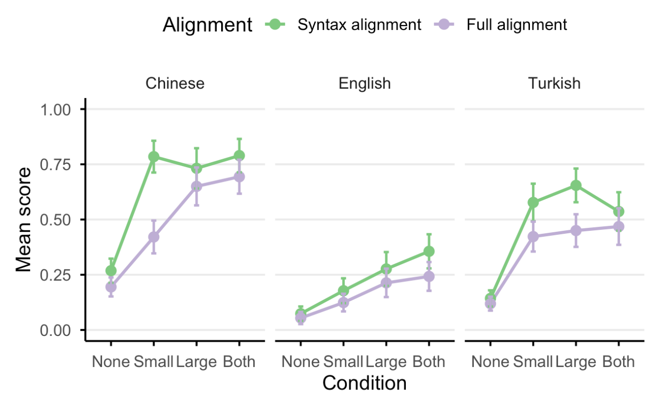

Figure 3. The y-axis represents the mean alignment score for syntax and full alignment. Error bars represent SE.

Packing alignment. To isolate packing from syntax, we examined packing alignment conditional on syntax alignment—that is, the probability of producing fully aligned expressions among responses that were already syntactically valid. Accordingly, we restricted the dataset to trials that aligned with syntax and coded responses in category (1) as 1 and all others as 0. Figure 4 plots these data.

In contrast to the language differences observed for full and syntax alignment, Chinese speakers did not show an

[1] We modified the preregistered error-type analysis. Rather than coding errors as either syntax or packing—which would incorrectly treat the two as opposites—we instead analyzed syntax alignment and packing alignment conditional on syntax alignment, which more directly captures their theoretical relation.

Table 2. Syntax-aligned expressions across number structures in the training group

Response	Example	None (%)	Small (%)	Large (%)	Both (%)
DMDm	1 × 9 + 1 × 3	88.0	88.1	83.0	91.5
DmDm	2 × 3 + 2 × 3	7.2	8.1	6.8	4.3
DmDM	1 × 3 + 1 × 9	—	3.7	8.2	3.5
DmM	1 × 3 × 9	75.0	48.4	84.1	83.8
DMDM	2 × 9 + 1 × 9	10.4	21.0	10.1	4.1
DMDMDM	1 × 9 + 1 × 9 + 1 × 9	8.3	19.4	4.3	—
DmMD	1 × 3 × 9 + 1	73.7	51.4	85.1	85.2
DMDMD	2 × 9 + 1 × 9 + 1	9.2	21.1	7.5	5.2
DMDMDMD	1 × 9 + 1 × 9 + 1 × 9 + 1	10.5	18.3	5.2	4.4
DMDM	2 × 9 + 2 × 9	76.9	60.7	23.5	29.6
DmDM	(1 × 3 + 1) × 9	15.4	21.4	26.5	22.2
DmMDM	1 × 3 × 9 + 1 × 9	—	17.9	41.2	44.4

Note. D denotes digits (Fim, Miv); m denotes the smaller multiplier (Gom); and *M* denotes the larger multiplier (Lod). Expressions in bold conform to both syntactic and packing principles. Unbolded expressions conform to syntax only. Examples are shown in Arabic numerals for clarity, although participants produced expressions using novel words.

advantage in applying the packing strategy (Chinese vs. English and Turkish: $\beta = 0.31$, $SE = 0.41$, $z = 0.76$, $p = .45$). We also found no effect of training (vs. baseline), $\beta = -0.12$, $SE = 0.54$, $z = -0.22$, $p = .83$.

Several unexpected patterns emerged. Participants who received training on large numbers performed better in applying packing than those trained on small numbers ($\beta = 1.06$, $SE = 0.56$, $z = 1.90$, $p < .01$). This effect was stronger for Chinese speakers (Interaction: $\beta = 3.03$, $SE = 1.12$, $z = 2.71$, $p < .01$).

Turkish speakers showed a marginal advantage in packing alignment relative to English speakers ($\beta = -0.95$, $SE = 0.55$, $z = -1.73$, $p = .085$) and benefited more from training on both small and large numbers (vs. either alone) than English speakers (Interaction: $\beta = -2.78$, $SE = 1.22$, $z = -2.27$, $p < .05$).

When math background was included as a covariate, the pattern of results remained unchanged, and math background was not a predictor of packing alignment ($p = 0.26$).

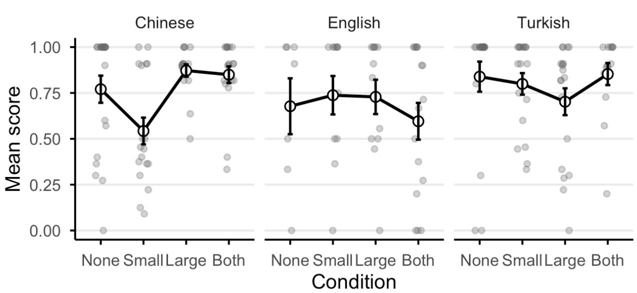

Figure 4. The y-axis represents the mean score for packing alignment. Error bars represent SE.

Taken together, we found that the language and training effects observed for full alignment and syntax alignment do not extend to packing alignment. This provides evidence for the prediction that people represent syntax and packing as two separate processes. In exploratory analyses, we further examined the conditions under which participants were more likely to fail to apply packing.

Exploratory analysis 1: Effect of number magnitude

A reasonable intuition is that participants are less likely to apply packing to larger numbers, as doing so requires manipulating greater numerical magnitudes. To test this possibility, we included standardized number magnitude as a covariate in the packing alignment model. The target numbers in the test ranged from 12 to 56. Results suggested that number magnitude did not significantly affect packing alignment ($\beta = -0.40$, $SE = 0.45$, $z = -0.89$, $p = .37$).

Exploratory analysis 2: Effect of number structure

We next tested a related hypothesis: that participants are less likely to apply packing when the target expression has a more deeply embedded structure, as this would require greater cognitive effort. The target numbers in the test included four different structures: Digit–Multiplier–Digit–multiplier (DMDm; e.g., 12: *Fim Lod Fim Gom*, 1×9 + 1×3), Digit–multiplier–Multiplier (DmM; e.g., 27: *Fim Gom Lod*, 1×3×9), Digit–multiplier–Multiplier–Digit (DmMD; e.g., 28: *Fim Gom Lod Fim*, 1×3×9 + 1), and Digit–multiplier–Digit–Multiplier (DmDM; 36: *Fim Gom Fim Lod*, (1×3 + 1)×9). Among them, the number with the DmDM structure has the most embedded structure. Because the number of items in each structure was uneven, we provide a qualitative analysis of number structure rather than a statistical test.

We summarize the three most frequently generated expressions for each number structure. As shown in Table 2, participants across all conditions did not consistently apply packing to syntax-aligned expressions (percentage < 100%). This was especially evident for DmDM structures, which were the only cases in which the correct expression was not the most frequent response. Moreover, participants in the training conditions did not show a clear advantage in packing alignment relative to the baseline (no training) condition, a pattern that was particularly apparent for DmDM structures.

A closer examination of responses that conformed only to syntax revealed a strong tendency to combine elements into local digit–multiplier units when forming larger expressions. For example, for 36 (a DmDM structure), participants often produced *Miv Lod Miv Lod* ("*two nine two nine*") rather than the expression that conformed to both syntax and packing, *Fim Gom Fim Lod* ("*one three one nine*," i.e., [[[*one*×*three*]+*one*]×*nine*]). A similar pattern appeared in other structures: for instance, the second most frequent response for 27 was *Miv Lod Fim Lod* ("*two nine one nine*"), rather than the target *Fim Gom Lod* ("*one three nine*").

Taken together, these patterns suggest that participants did not reliably apply packing when generating number expressions, particularly for DmDM structures. This is consistent with our hypothesis that participants are less likely to apply packing to more embedded structures. An examination of syntax-only expressions further suggests that, across number structures, participants tend to produce repeated digit–multiplier units to form larger numbers.

Discussion

This study examined people's mental representations of the structure of number words. Using a training paradigm, we tested whether syntax and packing—two principles proposed by Hurford—are represented by people as distinct cognitive processes. We compared native speakers of Chinese, which closely aligns with the number syntax, with speakers of English and Turkish, which contain irregularities in the syntax. We found that Chinese speakers were more likely than English or Turkish speakers to generate expressions that conformed to syntax. However, they did not show a stronger tendency to apply the packing strategy once syntax was satisfied. Moreover, training on expressions that aligned with both syntax and packing was associated with a higher likelihood of producing syntactically correct expressions (with or without packing), but did not improve the application of packing. Together, these findings suggest a dissociation between syntax and packing, supporting the hypothesis that the structure of number words is represented through two distinct cognitive processes rather than a single unified system.

The advantage of Chinese speakers in syntax, relative to English and Turkish speakers, is consistent with prior work showing that the transparency of number syntax facilitates number learning (e.g., Schneider et al., 2020; Srinivasan, 2024; Topaloğlu, 2026). Our results extend this literature by suggesting that this advantage may persist in how readily adult speakers access number-word structure. This may be because Chinese speakers encounter more consistent and transparent examples of number syntax. Interestingly, although both Turkish and English contain irregularities in number syntax, Turkish speakers showed some advantage over English speakers in producing syntax-aligned expressions. This suggests that syntactic transparency in number language is not the only factor shaping people's representations of number-word structure. Although we included math background as a covariate, it did not account for these differences, leaving open the possibility that other factors contribute to the cross-linguistic variation between Turkish and English speakers.

Crucially, the advantage observed for Chinese speakers in syntax did not extend to packing. This is consistent with the broader pattern that training was associated with a greater sensitivity to syntax but not to packing. Together, these findings extend Hurford's framework, suggesting that syntax and packing are distinct aspects in the mental assessment of number-word structure.

In the exploratory analysis, we found that participants were less likely to apply packing to numbers with a DmDM structure, which is more embedded than other structures. Instead of expressing 36 as *Fim Gom Fim Lod* $((1\times3 + 1) \times 9)$, participants were more likely to use a flatter structure, *Fim Lod Fim Lod* $(2\times9 + 2\times9)$. This repeated use of digit–multiplier units was also observed in other syntax-only expressions across numbers. If numbers can be expressed through repeated digit–multiplier units, however, why do number systems exhibit packing at all? In other words, why does packing exist?

Hurford (2007) proposed that packing emerges from everyday counting practices rather than from linguistic structure itself. On this account, packing is guided by efficiency-based principles such as "go as far as you can," whereby representations maximize the use of the largest available units (e.g., representing 14 as ten plus four rather than seven plus seven), and "minimize the number of entities," whereby representations favor fewer, larger groupings (e.g., 300 bags of 1,000 apples rather than 1,000 bags of 300 apples). This proposal aligns with broader accounts suggesting that pressures for efficiency shape language evolution (e.g., Fedzechkina et al., 2012; Gibson et al., 2019; Kirby et al., 2015; Mollica et al., 2020; Zaslavsky et al., 2018). If packing is indeed grounded in counting practices, then training that emphasizes efficient counting should increase people's tendency to apply packing. We plan to test this hypothesis in future studies.

Conclusion

We found evidence supporting the hypothesis that number syntax and the packing constraint are represented as two distinct processes. An important next step is to understand where these components come from and how they are acquired. A clearer account of how people represent number-word structure can not only inform the developmental literature on children's acquisition of number words, but also contribute to broader work in anthropology and historical linguistics on the forces that shape the evolution of numeral systems.

AsPredicted pre-registration:
https://aspredicted.org/bj3qi4.pdf

Acknowledgements

Special thanks to Sebastian Holt for reviewing earlier drafts. We also thank the members of the Higher-Level Cognition Lab for their insightful feedback during lab meetings and their support throughout the project.

References

Chrisomalis, S. (2004). A cognitive typology for numerical notation. *Cambridge Archaeological Journal*, *14*(1), 37-52.

Chrisomalis, S. (2010). *Numerical notation: A comparative history*. Cambridge University Press.

Chrisomalis, S. (2020). *Reckonings: Numerals, cognition, and history*. MIT Press.

Chu, J., Cheung, P., Schneider, R. M., Sullivan, J., & Barner, D. (2020). Counting to infinity: Does learning the syntax of the count list predict knowledge that numbers are infinite?. *Cognitive Science*, *44*(8), e12875.

Comrie, B. (2011). Typology of numeral systems. *Numeral types and changes worldwide. Trends in Linguistics. Studies and monographs*, *118*.

Comrie, B. (2022). The arithmetic of natural language: toward a typology of numeral systems. *Macrolinguistics*, *10*(1), 1-35.

Denić, M., & Szymanik, J. (2024). Recursive numeral systems optimize the trade-off between lexicon size and average morphosyntactic complexity. *Cognitive Science*, *48*(3), e13424.

Derzhanski, I., & Veneva, M. (2018). Linguistic problems on number names. In *Proceedings of the Third International Conference Computational Linguistics in Bulgaria* (pp. 169-176).

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

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*(44), 17897-17902.

Gelman, R., & Gallistel, C. R. (1978). The Child's Understanding of Number. Cambridge, MA: Harvard University Press.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389-407.

Guerrero, D., Hwang, J., Boutin, B., Roeper, T., & Park, J. (2020). Is thirty-two three tens and two ones? The embedded structure of cardinal numbers. *Cognition*, *203*, 104331.

Hammarström, H. (2010). Rarities in numeral systems. *Rethinking universals: How rarities affect linguistic theory*, *45*, 11-53.

Ho, C. S. H., & Fuson, K. C. (1998). Children's knowledge of teen quantities as tens and ones: Comparisons of Chinese, British, and American kindergartners. *Journal of Educational Psychology*, *90*(3), 536.

Holt, S., & Barner, D. (2023). How does the syntax of counting affect learnability? Evidence from artificial language learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Holt, S., & Barner, D. (2025a). Learning a novel number system: The role of compositional rules and counting procedures. *Cognitive Science*, *49*(6), e70071.

Holt, S., & Barner, D. (2025b). Can preschoolers learn the syntax of number? Using rules to combine familiar and novel number words. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *380*(1937).

Hurford, J. R. (1975/2011). *The linguistic theory of numerals*. Cambridge University Press.

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

Hurford, J. R. (2003). *The interaction between numerals and nouns*. https://doi.org/10.1515/9783110197075.4.561

Hurford, J. R. (2007). A performed practice explains a linguistic universal: Counting gives the Packing Strategy. *Lingua*, *117*(5), 773-783.

Ionin, T., & Matushansky, O. (2006). The composition of complex cardinals. *Journal of semantics*, *23*(4), 315-360.

Ionin, T., & Matushansky, O. (2018). *Cardinals: The syntax and semantics of cardinal-containing expressions* (Vol. 79). MIT press.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87-102.

Liu, E., & Xu, Y. (2019). Rapid information gain explains cross-linguistic tendencies in numeral ordering. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 41).

Miller, K. F., & Stigler, J. W. (1987). Counting in Chinese: Cultural variation in a basic cognitive skill. *Cognitive development*, *2*(3), 279-305.

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

Miura, I. T., & Okamoto, Y. (2013). Language supports for mathematics understanding and performance. In *The development of arithmetic concepts and skills* (pp. 229-242). Routledge.

Mollica, F., Bacon, G., Zaslavsky, N., Xu, Y., Regier, T., & Kemp, C. (2021). The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, *118*(49), e2025993118.

Park, J., Guerrero, D., Hwang, J., & McCollum, R. (2025). The Effect of Numerical Syntax on Number Comparison in Children. *Journal of Cognitive Science*, *26*(1).

Prasertsom, P., Silvi, A., Culbertson, J., Dubhashi, D., Johansson, M., & Smith, K. (2026, March). Recursive numeral systems are highly regular and easy to process. In *Proceedings of the 19th Conference of the European*

Chapter of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 4873-4885).
Schneider, R. M., Sullivan, J., Marušič, F., Biswas, P., Mišmaš, P., Plesničar, V., & Barner, D. (2020). Do children use language structure to discover the recursive rules of counting?. *Cognitive Psychology*, *117*, 101263.
Silvi, A., Thomas, J., Carlsson, E., Dubhashi, D., & Johansson, M. (2024). Learning Efficient Recursive Numeral Systems via Reinforcement Learning. *arXiv preprint arXiv:2409.07170*.
Srinivasan, A. (2024). *Children's Learning of the Compositional Number System*. Harvard University.
Topaloğlu, S. (2026). *Counting on Language: Morphological Transparency and Children's Acquisition of Number Words and Numerical Concepts.* Harvard University.
Xu, Y., Liu, E., & Regier, T. (2020). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind*, *4*, 57-70.
Yang, D., & Regier, T. (2025). Re-examining the tradeoff between lexicon size and average morphosyntactic complexity in recursive numeral systems. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 47).
Yuan, L., Prather, R. W., Mix, K. S., & Smith, L. B. (2019). Preschoolers and multi-digit numbers: A path to mathematics through the symbols themselves. *Cognition*, *189*, 89-104.
Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937-7942.