US-China Foreign Language, May 2025, Vol. 23, No. 5, 169-175
doi:10.17265/1539-8080/2025.05.002

A Corpus-Based Study on the Semantic Functions of Large Numerals in Mandarin-Speaking Children

YANG Meiling, DU Ailin
Northeastern University, Shenyang, China

The usage of large numerals in child language reflects not only cognitive development, but also sociocultural influences. To investigate this phenomenon, we analyze the semantic functions of large numerals "hundred", "thousand", and "ten-thousand" in Mandarin-speaking children by comparing them with adult data based on a self-constructed corpus. Our results showed that both groups predominantly use large numerals for non-cardinal functions, particularly for expressing abstract concepts and culture-specific meanings rather than for exact numerical quantification. Notably however, while adults tend to utilize large numbers primarily for metaphorical or rhetorical purposes, children demonstrate stronger preference for employing large numerals in counting routines and symbolic representations. These findings illuminate the multifaceted semantic properties of large numerical expressions in Mandarin Chinese.

Keywords: large numerals, child language, semantic functions, corpus-based study

Introduction

Numeral systems are not merely tools for counting but complex linguistic constructs that reflect and shape human cognition (Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004). Across languages, the structure of number naming systems varies dramatically, raising fundamental questions about how these differences influence children's acquisition of abstract mathematical concepts (Xu & Regier, 2014). Among these, the Chinese and English number naming systems stand out for their contrasting systems for forming number names. The Chinese number word system strictly follows the principle of decimal system, prioritizing precision, detail, and ordinal regularity. Through various combinations of coefficients (e.g., "三" in "三百") and place values (e.g., "百"), it can accurately express values ranging from single digits to astronomical numbers (Miller, Smith, Zhu, & Zhang, 1995). In contrast, the English numeral system prioritizes hierarchical progression and organizes numbers into discrete tiers through specialized magnitude words, such as "thousand" and "ten thousand" (Ngan & Rao, 2010). These fundamental differences profoundly manifest language-specific cognitive architectures and cultural numeracy traditions, underscoring the need to investigate acquisition of numerals in children of different linguistic backgrounds.

Acknowledgments: This study is funded by Liaoning Provincial Social Science Planning Fund Project "The Development of Children's Large-Number Representation and the Role of Parental Linguistic Input: A Cross-Linguistic Study of English and Chinese" (Project Fund No. L22CYY003) [本文系辽宁省社会科学规划基金项目"英汉对比视域下儿童大数量表征发展及父母语言输入的影响研究"(基金编号：L22CYY003)的阶段性研究成果].

YANG Meiling, Ph.D., lecturer, Foreign Studies College, Northeastern University, Shenyang, China.
DU Ailin, postgraduate, Foreign Studies College, Northeastern University, Shenyang, China.

Despite these cross-linguistic differences, most research on numeral acquisition has focused narrowly on English-speaking children and small numerals (e.g., “one”, “two”, and “three”), leaving large numerals understudied across languages. Compared to small numerals, large numerals like “hundred” (百), “thousand” (千), and “ten thousand” (万) involve more quantity units and complex numerical relationships. This structural complexity makes them difficult to perceive intuitively through concrete objects, thus constituting more abstract and complex mathematical concepts. When acquiring large numerals, children face challenges not only in understanding their relative sizes, but also in mastering the rich and variable meanings of these terms across different contexts.

Given this complexity, recent work has begun to explore how large numerals are used semantically. For example, Cheung and Daniel (2023) investigated how American children acquire large numerals by examining potential influences from adult usage patterns. In their study, the researchers coded the semantic functions of large numerals, such as “hundred”, “thousand”, “million”, “billion”, and “trillion”, identifying categories, such as nominal use (e.g., a bus number), adjectival use (e.g., a hundred-watt bulb), time/year references (e.g., thousands of years), numerical symbols (e.g., mathematical or statistical notations), cardinal use (e.g., a million books), and counting routine. This study revealed that adults disproportionately employ large numerals in daily communication to quantify abstract concepts (e.g., time, money, and weight) rather than to describe discrete physical entities. This finding suggests that children’s acquisition of large numerals may be significantly influenced by these adult usage patterns. However, the research did not explore how children acquire and use these terms across different contexts, particularly in their extended semantic functions beyond literal quantification.

In terms of large numerals acquisition, some Chinese researchers have investigated the acquisition of these large numerals in Chinese-speaking children, identifying four developmental stages: (a) the emergent stage (acquiring “one”), (b) the initial developmental stage (acquiring “two, three”), (c) the explosive stage (acquiring “four” through “ten”), and (d) the higher-order stage (acquiring “hundred,” “thousand,” and “ten-thousand”) (Yan & Su, 2023, p. 130). While this study has significantly enhanced our understanding of the cognitive progression in Chinese children’s numerical development, it offers limited insight into how children’s semantic usage of these terms diverges from that of adults.

Given this, the present study aims to systematically examine the semantic functions of large numerals, i.e., “hundred” (百), “thousand” (千), and “ten-thousand” (万) in Chinese-speaking children, contrasting their usage with adult patterns. This study mainly answers the following research questions:

1. What are the similarities and differences between Chinese-speaking children and adults in the semantic functions of the large numeral terms?

2. What factors might account for the observed differences between children and adults in using these numerical terms?

Method

Data Collection

The data for this study are primarily derived from the Mandarin Chinese child database in CHILDES. This corpus contains conversational records between Mandarin-speaking children and adults in natural contexts (Wen & Hu, 2001).

To capture the usage patterns of large numerals in different contexts among children and adults, we

constructed a quantifiable Chinese large-number word corpus based on established research methodologies. Using these numeral terms "hundred" ("百"), "thousand" ("千"), and "ten thousand" ("万") as search keywords, we extracted naturalistic speech samples from children aged 3-6 years and their adult caregivers. In this process, we excluded sub-datasets with missing critical information (e.g., children's ages or the number of participants). The final corpus comprised 79 child-caregiver dyads, with a mean child age of 59.82 months (*SD* = 14.65).

As search keywords, we extracted naturalistic speech samples from children aged 3-6 years and their adult caregivers. In this process, we excluded sub-datasets with missing critical information (e.g., children's ages or participant numbers). The final corpus comprised 79 child-caregiver dyads, with a mean child age of 59.82 months (*SD* = 14.65).

Data Coding

For analysis, following the framework established by Cheung and Daniel (2023), we categorized large-number word usage into two primary types: cardinal and non-cardinal uses. "Cardinal usage" refers to "the fundamental quantitative function of these words in denoting exact quantities". The non-cardinal category encompassed eight distinct functional subtypes: (a) enumeration (sequential counting, e.g., "131, 132, 133"); (b) measurement (quantifying continuous quantities, e.g., "300 kilometers"); (c) symbolic reference (numeral naming/metalinguistic discussion, e.g., "write the number 1,000"); (d) nominal usage (naming conventions for locations/objects, e.g., "130th Street"); (e) temporal reference (time-related expressions, e.g., "3,000 years"); (f) monetary value (currency denominations, e.g., "100 yuan"); (g) cultural-semantic usage (lexicalized expressions with figurative meaning, e.g., "passion fruit" ("百香果"); and (h) miscellaneous (unclassifiable instances).

To ensure the validity and reliability of the coding process, two trained coders performed independent classification coding. Inter-coder discrepancies were resolved through iterative discussions until consensus was achieved on all classification outcomes.

Results and Discussion

We quantitatively analyzed the usage patterns of large numerals ("hundred", "thousand", and "ten thousand") separately for children and adults, with the results presented in Figure 1 and 2.

According to Figure 1 and 2, one of the most notable research findings is that both children and adults consistently prefer employing large numerals for non-cardinal functions rather than cardinal meanings. This finding partly contrasts with Cheung et al., which found that American adults predominantly use cardinal meanings over non-cardinal meanings. The observed cross-linguistic difference suggests that semantic richness of numerals in Mandarin Chinese significantly surpasses that observed in English, particularly in their non-cardinal applications. The subsequent analyses will elaborate on this point.

In terms of cardinal uses, data from Figure 1 and 2 reveal that children and adults are able to use the large numerals "hundred" (百), "thousand" (千), and "ten thousand" (万) in cardinal expressions. These findings are consistent with previous research by Cheung and Daniel (2023), in that large numerals in both languages can be used for describing specific quantities of objects. However, children exhibit a significantly higher frequency of using these words in cardinal contexts than adults. This difference may reflect key characteristics of cognitive development and language acquisition. In the early stages of mathematical concept formation, children may rely on precise numerical expressions to reinforce their concrete understanding of quantity (Wynn, 1990). In contrast, adults, due to the maturation of the approximate number system (ANS), can process approximate quantification

more efficiently (Halberda & Feigenson, 2008). Additionally, this divergence may also arise from pragmatic constraints. Mandarin-speaking adults may prioritize communicative efficiency in daily interactions, thus favoring vague quantifiers (e.g., "many", "a lot", or "a few") (Zhu & Wu, 2012, p. 196), whereas young children are still establishing language-quantity mappings and therefore depend more on explicit numerical expressions.

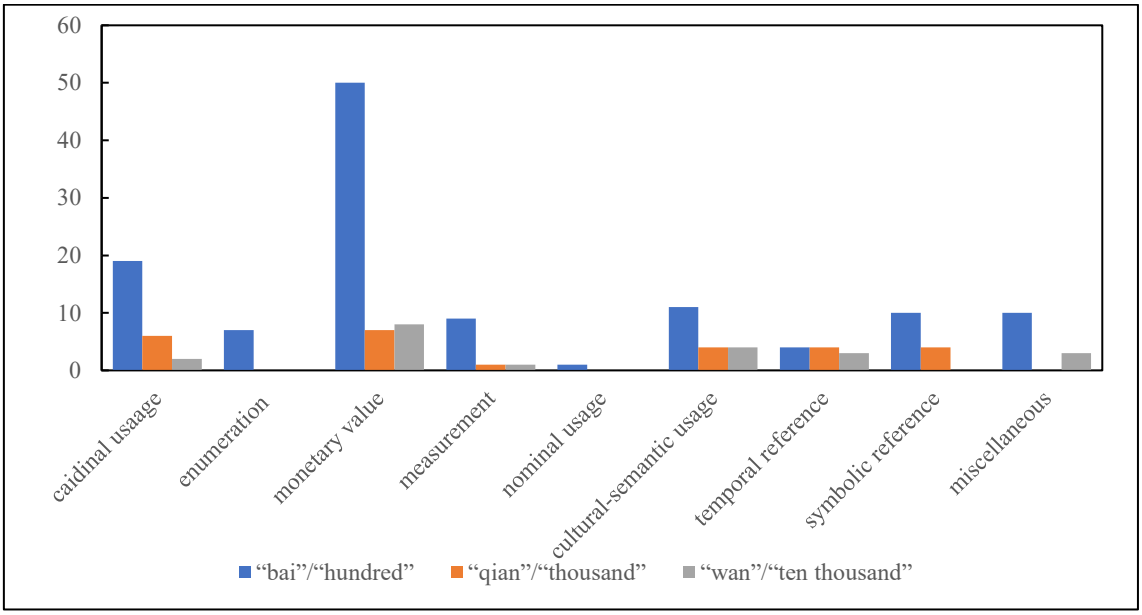

Figure 1. Frequency distribution of semantic categories in children's use of large numerals "hundred", "thousand", and "ten thousand".

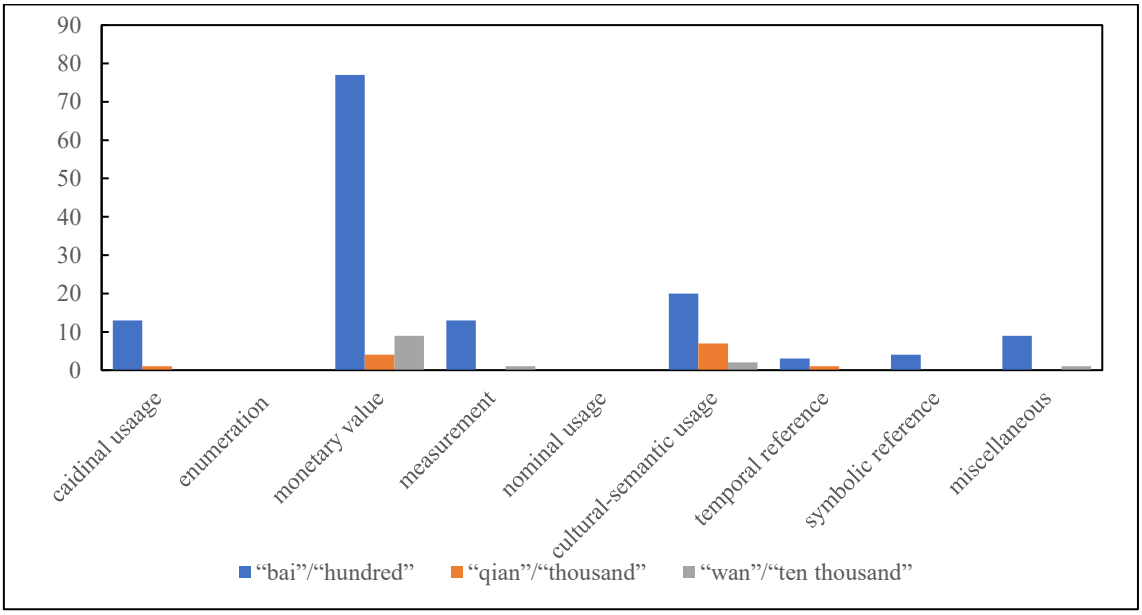

Figure 2. Frequency distribution of semantic categories in adults' use of large numerals "hundred", "thousand", and "ten thousand".

In terms of non-cardinal uses, children and adults demonstrate substantial similarities.

Firstly, both children and adults most frequently use large numerals to denote monetary amounts, highlighting their core significance in economic exchanges. This finding contrasts with the results of Cheung and

Daniel, who found that nominal use, such as “a two-thousand dollars loan” rarely occurred among American adults. These contrasting patterns between Mandarin and English speakers may reflect deeper cultural-linguistic differences in numerical cognition. In Mandarin, monosyllabic number words, such as “千” for “thousand” are readily incorporated into fixed monetary expressions. In contrast, basic English number terms like “thousand” and “hundred” contain multiple syllables. For example, the Chinese phrase “三千元” (“three thousand yuan”) requires only three syllables, while its English equivalent consumes five syllables (“three thousand yuan”). This reduction in syllabic load significantly decreases cognitive effort during financial transactions, making precise monetary expressions more economical for Mandarin speakers. An alternative explanation may lie in the cross-linguistic differences in morphological binding. Mandarin numerals form tight syntactic units with currency terms, as seen in “五万块” (“fifty thousand yuan”), which might function as a single lexical chunk. In contrast, English requires prepositional phrases (“a sum of fifty thousand yuan”), adding grammatical complexity. This structural advantage explains why Chinese speakers use exact monetary more frequently in daily communication. However, further research is needed to provide more robust empirical evidence to confirm this hypothesis.

Secondly, both adults and children use “hundred”(百), “thousand”(千), and “ten thousand”(万) to construct linguistic imagery and cultural metaphors. For example, “baibao xiang” (“百宝箱”, “hundred-treasure box”) uses “hundred” metaphorically to signify an abundance of valuable items rather than a literal hundred treasures, while “qianli mu” (“千里目”, “thousand-mile vision”) employs “thousand” to symbolize both far-reaching sight and profound wisdom. This parallel usage pattern between children and adults suggests that children’s acquisition of numerical expressions may be strongly influenced by adult language input, as they emulate the conventionalized usage patterns prevalent in their linguistic environment (Lieven, Behrens, Speares, & Tomasello, 2003). Moreover, these findings indicate that large numerals in Mandarin exhibit greater semantic diversity, encompassing both exact numerical meanings and hyperbolic or figurative usage. While this richness enhances expressive power, it might also introduce cognitive complexity for children as they learn to disentangle precise quantitative meanings from culturally embedded metaphors. This might contribute to a potential delay in Mandarin-speaking children’s acquisition of large number words compared to learners of languages where numerical expressions are more semantically constrained. Nevertheless, future research is needed to further investigate this issue.

Thirdly, in consistent to Cheung and Daniel (2023), we also found that large numerals are often used to serve to quantify abstract concepts like temperature (“100 degrees”), distance (“300 kilometers”), and time (“100 days”). These results further substantiate the remarkable adaptability and expressive power of these numerals when measuring non-material, perceptually intangible abstract domains including temperature, distance, and temporal duration (Yin & Li, 2002). This may reflect a shared cognition among humans. In other words, the association between large numerals and abstract concepts is established through the cognitive architecture of the human mind. When individuals encounter statements like “100 degrees of high temperature”, the brain engages in a process of conceptual mapping, where the concrete numeral serves as a reference point to conceptualize and understand the abstract notion of temperature. This process is rooted in the way human minds organize and interpret information, leveraging the tangible nature of numerals to lend structure and meaning to intangible concepts. Moreover, the use of large numerals to quantify abstract concepts may also reflect deeper socio-cultural and linguistic conventions that shape how numerical cognition develops in Mandarin-speaking populations. Cross-linguistic research suggests that the frequency and conventionalization of numerical expressions in a given language influence how readily children and adults associate numbers with abstract domains. In Mandarin, the

pervasive use of numerals like “hundred” (百) and “ten thousand” (万) in both concrete and metaphorical contexts may reinforce cognitive schemas that link numerical magnitude with abstract intensity or scale. This differs from languages where large numbers are less frequently employed in idiomatic or hyperbolic speech, potentially leading to variations in how numerical abstraction is acquired cross-culturally.

Notably, children and adults differ in specific aspects of non-cardinal usage. For example, children show a marked preference for the symbolic use of these words, possibly due to their exploratory learning phase in mastering large-number concepts. Since large-number words cannot be perceived through daily experience, children may need to rely on symbolic exercises when learning these terms. In daily life, parents or adults often increase children’s exposure to symbolic practices of large-number words (such as writing and reading exercises), helping them establish a connection between abstract concepts and symbolic representations through repeated manipulation. This pedagogical approach serves to scaffold children’s construction and internalization of quantitative concepts. In contrast, adults exhibit automated numerical processing capabilities, enabling efficient implicit quantification. Consequently, their linguistic production shows significantly reduced symbolic usage of large numerals, representing a fundamental developmental divergence.

Furthermore, it is worth delving into the impact of these differences in language input on children’s acquisition of non-cardinal uses of numbers. Previous studies have indicated that when adults employ child-directed speech, they tend to simplify numerical expressions. While this simplification might seem to make communication easier for children to comprehend initially, it could potentially lead to a delay in children’s grasp of non-cardinal usages. Non-cardinal uses of numbers, such as using “hundred” in a symbolic or metaphorical sense (e.g., “I have a hundred reasons to be happy”), require a deeper understanding of the number’s abstract meaning beyond its basic counting function. If children are predominantly exposed to simplified numerical expressions, they may not have sufficient opportunities to encounter and learn these more complex, non-cardinal applications. These findings carry significant implications for parental education practices at home. While parents naturally focus on teaching children the cardinal meanings of number words through daily activities (e.g., counting objects or reciting numbers), they should also consciously expose children to non-simplified and diverse numerical expressions. For instance, during storytelling or shopping, parents can intentionally incorporate large numbers in symbolic or metaphorical ways. For instance, when reading picture books or describing nature, parents can highlight culturally rich idioms like “万紫千红” (literally “ten thousand purples, thousand reds,” meaning vibrant colors in bloom) and explain how such expressions use large numbers metaphorically to convey abundance and beauty. Such intentional input helps children recognize that numbers function not just as quantitative tools, but also as rich linguistic devices for expressing abstract concepts. By balancing explicit cardinal instruction with natural exposure to non-cardinal usage, parents can support a more comprehensive understanding of numerical language.

Conclusions

This study is the first attempt to analyze the semantic functions of large numerals “hundred”, “thousand”, and “ten-thousand” in Mandarin-speaking children by comparing them with adult data based on a self-constructed corpus. There are three major findings. First, compared to cardinal uses, children and adults tend to employ these large numerals more frequently for non-cardinal functions. Second, significant similarities in non-cardinal usage patterns are observed between children and adults. For example, both children and adults demonstrate a strong tendency to employ large numerals for expressing abstract and complex concepts-including monetary values,

spatial distances, temporal durations, and culturally-specific meanings. The observed similarity in distribution across metaphorical, abstract, and monetary contexts implies that children are not simply learning these terms as mathematical concepts, but are acquiring culturally embedded linguistic practices through social interaction. Third, adults and children exhibit distinct preferences in certain non-cardinal uses. For example, children exhibit significantly higher frequencies of counting and symbolic uses compared to adults. This divergence in usage habits directly reflects the stage-specific characteristics of children's cognitive development.

Altogether, the present findings contribute to our understanding of numerical language development in Mandarin-speaking children and particularly emphasize the semantic richness of numeral terms in the Chinese language. Building on these insights, future research could further compare cross-cultural differences in the use of large-number terms, thereby advancing a more comprehensive understanding of language acquisition and cross-cultural numerical cognition.

References

Cheung, P., & Daniel, A. (2023). A million is more than a thousand: Children's acquisition of very large number words. *Developmental Science, 26*(1), e13246.

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

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307-314

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457-1465.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30*(2), 333-370

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.

Ngan, S. S., & Rao, N. (2010). Chinese number words, culture, and mathematics learning. *Review of Educational Research, 80*(2), 180-206.

Wen, Z. J., & Hu, G. L. (2001). Developing and utilizing the world's largest child language corpus—CHILDES. *Foreign Language Teaching and Research, 45*(5), 374-377.

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

Xu, Y., & Regier, T. (2014). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind, 4*, 57-70.

Yan, X. X., & Su, X. S. (2023). A study on the development of cardinal numerals in Chinese-speaking children. *Journal of Capital Normal University (Social Sciences Edition), 51*(1), 121-130.

Yin, J. Q., & Li, W. (2002). On the "Qian A Wan B" type of idioms. *Studies in Language and Linguistics, 22*(S1), 140-143.

Zhu, K., & Wu, J. (2012). A cultural perspective on the treatment of numerical fuzziness in English and Chinese idioms. *Journal of Southwest Minzu University (Humanities and Social Sciences Edition), 33*(5), 195-198.