Proceeding of ICIHCS 2025 Symposium: The Dialogue Between Tradition and Innovation in Language Learning
DOI: 10.54254/2753-7064/2025.HT26990

CHILDES-Based Investigation into the Development Path of Quantifier System Construction in Native English-Speaking Children

Yiwen Hua

School of English Studies, Xi'an International Studies University, Xi'an, China
18049475725@163.com

Abstract. In this paper, the situation of native English-speaking 4-8-year-old children's quantifier use is discussed, and the quantifiers' collocation accuracy and mistake types are analyzed. Based on the CHILDES corpus, this study examines 5,236 quantifier-containing utterances from 20 native English-speaking children aged 4-8. It analyzes collocation accuracy, error types, and developmental trends via quantitative statistics and corpus annotation. Results show total errors decrease from 16 (4-5 years) to 8 (7-8 years). Key findings: singular-plural errors dominate at 4-5 (15 cases) but vanish by 6-7; "the" misuse persists (6 cases at 5-6, 5 at 7-8); "a/an" confusion peaks at 6-7 (3 cases). Combining cognitive development theory and language input analysis, it systematically combs through the building process of children's quantifier use; its mistakes show an obvious aging tendency, with types changing from basic form confusion (singular-plural, most common among 4-5-year-olds) to context-dependent errors, especially "misusing 'the'" (at 7-8 years old). This provides scientific support for age-specific quantifier teaching practice and taking effective countermeasures to improve their quantifier use ability. These reflect cognitive-linguistic development, providing scientific support for age-specific quantifier teaching.

Keywords: CHILDES Corpus, Native English-Speaking Children, Quantifier Usage, Language Development, Cognitive Theory

1. Introduction

Quantifiers are essential grammatical elements that help specify the quantity and type of objects, contributing to precision and clarity in communication. For children, acquiring quantifier competence is a key aspect of early language development. The period between ages 4 and 8 is critical, as it coincides with rapid cognitive and linguistic development [1]. A child's ability to use quantifiers correctly not only reflects their level of thinking but also affects the quality of their daily exchanges. Existing studies often suffer from limitations such as small sample sizes, overreliance on controlled experiments, and a lack of focus on natural language use. Moreover, the 4-8 age range, a formative phase for linguistic and cognitive development, has received relatively little attention in terms of detailed, longitudinal analysis of quantifier usage patterns.

DOI: 10.54254/2753-7064/2025.HT26990

This study addresses these gaps by utilizing the CHILDES corpus, which contains rich natural language data from native English-speaking children [2]. The research adopts both categorical and quantitative analytical methods to examine the accuracy, collocational richness, and common error types associated with quantifier use. Through the lens of cognitive development theories and language input models, this study seeks to uncover developmental trajectories in quantifier acquisition and identify factors—such as parental input and frequency of exposure—that shape this process [3,4].

Specifically, the study addresses the following questions: (1) How does children's quantifier use change between the ages of 4 and 8? (2) What is the relationship between quantifier development and cognitive development levels? (3) How can it further prove that parental language input has an impact on this process? By answering these questions, the study aims to provide deeper theoretical insights for optimal child language education, allowing teachers and parents to offer effective intervention schemes aimed at promoting proficient language use.

2. Literature review

This chapter reviews existing studies on children's acquisition of quantifiers and the theoretical frameworks underlying this process, focusing especially on the cognitive and input factors that influence development in native English-speaking children aged 4 to 8.

Piaget's Principles of Genetic Epistemology provides a cognitive basis: his preoperational stage (4-7 years) explains why young children struggle with quantifiers. Their difficulty with abstract concepts like "singular-plural" or "countable-uncountable" leads to overgeneralizing rules (e.g., using "a" with plural nouns) [3]. The CHILDES corpus, introduced in MacWhinney's The CHILDES Project: Tools for Analyzing Talk, standardizes the analysis of naturalistic data, making it possible to track developmental trends [2].

Brown's A First Language: The Early Stages traced the complexity of quantifier use in 2-5-year-olds, noting improvements with age but only focusing on early childhood [4]. Li Hong added to this by linking 3-6-year-olds' quantifier use to cognitive growth (e.g., classificatory thinking) but did not look at how 6-8-year-olds transition to complex quantifiers (e.g., "few," "many") [5].

Cross-linguistic research provides more information. Zhou Jingyi found that English learners exhibited native language transfer when using Chinese quantifiers, and Hu Yuanlian revealed that Chinese learners showed similar transfer when using English quantifiers [6,7]. Language input was emphasized by Shaher Banu Vagh et al., who showed that quantifier errors in bilingual children were closely related to input diversity [4]. However, these studies did not pay sufficient attention to the spontaneous development of quantifiers in monolingual English children aged 6-8 years.

In short, existing research lays a foundation but lacks a dynamic analysis of 4-8-year-olds' quantifier systems, especially how cognition and input interact in late childhood. This study fills this gap by combining corpus data with cognitive and input analysis.

3. Research design

3.1. Corpus sifting

The target population includes children aged between 4 and 8 years, divided into specific age groups (4-5, 5-6, 6-7, and 7-8 years). All selected children were native English speakers living in multilingual environments, where they were regularly exposed to other languages, like family or community languages, but with special interest looking for interlingual effects. To control for

DOI: 10.54254/2753-7064/2025.HT26990

variables concerning daily language exposure (ratio English/other languages), it is encoded into covariates, allowing comparison of different subgroups among each other.

Sifting Criteria and Process:

The corpus sifting process began with identifying all transcripts within the relevant subcorpus of CHILDES that matched the specified age ranges. For each transcript, two key filtering criteria were applied:

Age verification: Check the accuracy of the child's age recording. Only transcripts with clear and verifiable age information—either directly documented in the metadata or inferable through contextual clues such as references to school grade—were included.

2) Context relevance: Retain transcripts that contain natural communicative contexts. It consists of daily interactions like family dinners, playtime conversations, and classroom discussions and excludes highly specialized or artificial contexts (e.g., scripted language tests without natural interactions) because they may not reflect typical use of quantifiers.

Simple filtering ensures that the data for further analysis is both representative of natural language use and focused on interesting linguistic elements.

3.2. Quantifier classification

Once the corpus was established, quantifiers were classified into three main types based on their semantic and syntactic functions, following developmental patterns observed in children's language [5,8]:

Individual quantifiers: These refer to quantifiers that modify single, countable entities. Examples include articles (a, an, the—though a and an are often studied in the context of quantifier acquisition due to their role in noun-phrase modification) and simple numerals (one, two). In the early stages of child language (e.g., 7-8 years), these are among the first quantifiers to emerge, as seen in the example "I see a octopus" where a is an individual quantifier (despite the article misusage).

Set quantifiers: These refer to groups or sets of entities, such as some, many, and few (e.g., some books, many children). Set quantifiers develop later as children gain a better understanding of aggregate concepts [5].

Measure quantifiers: These relate to units of measurement for uncountable or mass nouns, e.g., a lot of (a lot of water), much. They require an understanding of abstract quantity concepts and thus show a more protracted developmental course [9].

3.3. Annotation standards

In order to ensure the consistency and reliability of the analysis, the following annotation standards are established:

(1) Token identification: Each quantifier token in the corpus is marked, including its position in the sentence. For example, in "Give me two cookie," the quantifier two is identified as a modifier of the noun cookie.

(2) Error tagging: Errors are categorized based on type, such as "Misuse of articles" (e.g., a instead of an before vowel-starting nouns) and "Single and plural issues" (e.g., singular nouns after plural numerals) [5, 8]. Each error type is clearly defined.

(3) Semantic and cognitive labeling: In addition to syntactic and error tags, annotations include the semantic context of quantifiers (concrete vs. abstract entities) and their correlation with cognitive development stages [2, 10]. For example, the error in "I saw an elephant" (confusing a and an) is linked to preoperational-stage reliance on visual symbols rather than phonetic structures.

DOI: 10.54254/2753-7064/2025.HT26990

This study systematically analyzes how native English children acquire quantitative words, from corpus selection to detailed annotations, by adopting this research design, providing a solid basis for subsequent research results and analysis.

4. Research findings and analysis

4.1. Quantitative analysis of quantifiers development in native English-speaking children

Table 1. Types and frequency of quantitie-related grammar errors across different age groups

Error Type	Age 4-5	Age 5-6	Age 6-7	Age 7-8
Misuse of articles	0	2	3	1
Misuse of the definite article "the"	0	6	5	5
Single and plural issues	15	5	0	0
Wrong collocation of quantifiers	1	1	1	2
Misuse of mass quantifiers	0	0	1	0
Total	16	14	10	8

As shown in Table 1, quantifier-related grammar errors among native English-speaking children exhibit a clear developmental trend: the error rate declines steadily as they grow, indicating their gradual improvement in language skills. In general, children's error count decreases with age, reflecting that their understanding of quantifier rules strengthens alongside cognitive and linguistic development [2, 11].

A closer examination of error types across age groups reveals distinct developmental patterns:

(1) Single and plural issues are most frequent in early childhood (15 cases at 4-5 years) but disappear by 6-7 years, suggesting children rapidly master singular-plural agreement after age 5 [5, 8].

(2) Misuse of the definite article "the" is the most persistent error, with 6 cases at 5-6 years and 5 cases at 7-8 years, indicating that context-dependent use of "the" (e.g., distinguishing specific vs. general reference) remains a long-term challenge [1, 9].

(3) Misuse of articles (e.g., confusing "a" and "an") peaks at 6-7 years (3 cases) then decreases to 1 case at 7-8 years, reflecting improving integration of phonological perception and grammatical rules [8, 10].

4.2. Types and distribution of quantifier errors

Children's quantifier errors can be divided into five types, each with distinct age-related patterns:

(1) Single and plural issues

Definition: Incorrectly pairing singular articles (e.g., "a") with plural nouns (e.g., "a dogs").

Distribution: Most common in early childhood (15 cases at 4-5 years old), dropping to 5 cases at 5-6, and gone by 6-7 years old [5, 8].

(2) Misuse of articles

Definition: Confusing "a" and "an" (e.g., "an cat" for consonant-initial nouns).

Distribution: Starts at 5-6 years old (2 cases), peaks at 6-7 (3 cases), and decreases to 1 case at 7-8 years old [8, 10].

(3) Misuse of the definite article "the"

DOI: 10.54254/2753-7064/2025.HT26990

Definition: Overusing/underusing it in contexts like first mentions (“the book” for a new noun) or fixed phrases (“go to the school”).

Distribution: The most persistent error, with 6 cases at 5-6, 5 at 6-7, and 5 at 7-8 years old [1, 9].

(4) Wrong collocation of quantifiers

Definition: Incorrect pairing (e.g., “a Beijing” with proper nouns).

Distribution: Rare but consistent (1-2 cases across all ages) [7, 12].

(5) Misuse of mass quantifiers

Definition: Using “a” with uncountable nouns (e.g., “a bread”).

Distribution: Rare (1 case at 5-6 and 6-7 years old) [5, 9].

Overall, errors change from basic form issues (singular-plural) to context-dependent mistakes (e.g., “the” usage) as children grow older [2, 11].

4.3. Underlying causes of quantifier errors in native English-speaking children

The errors in children’s quantifier use are closely related to their cognitive development, language regulation learning, and context processing abilities, with age-specific features as follows:

(1) Cognitive limitations in early childhood (ages 4-5)

Children at this stage mainly rely on simple imitation and rule generalization, with weak skills to tell grammatical details apart.

As a result of overgeneralizing the “a + noun” structure to plural nouns, single and plural issues (e.g., “a dogs”) show children’s misunderstanding of “singular-plural agreement.” They know incompletely about the classification of nouns, so they concentrate on forms rather than grammatical features (e.g., treating all nouns as countable). They often forget articles (e.g., “There is bird”) or overuse “a” because of limited usage frequency, focusing on core nouns instead of grammatical markers.

(2) Rule integration difficulties in middle childhood (ages 5-6)

As children learn basic rules, they struggle to manage complex grammatical relationships, leading to conflicting errors.

Confusion between “a” and “an” (e.g., “an cat”) happens because they can’t well combine phonological perception (telling vowels/consonants apart) with grammatical rules, failing to link “vowel-initial pronunciation” to “an” usage.

Misuse of “the” in fixed phrases (e.g., “go to the school”) shows incomplete memory of phrase wholeness. They break down fixed collocations and apply literal rules, indicating an underdeveloped ability to recognize whole phrase structures.

(3)Contextual processing weaknesses in late childhood (ages 6-7)

Children make fewer basic errors but struggle with context-dependent quantifier use because their ability to adjust to situation changes flexibly is still developing.

Inappropriate “the” for first mentions (e.g., “The book is red”) shows limited tracking of “new/old information” in conversation, with judgments affected by long-term memory.

Overgeneralization of special rules (e.g., “play the football”) happens when they extend “the + instrument” rules (e.g., “play the piano”) to sports, showing rigid use of learned rules without considering context.

(4) Special rule mastery delays in early school age (ages 7-8)

Most errors are about specific context rules, as children haven’t fully learned less common or complex usages.

Missing “the” with superlatives (e.g., “He is best student”) and misuse with proper nouns (e.g., “a Beijing”) show that special rules (e.g., “the + superlative,” “zero article for proper nouns”) need

DOI: 10.54254/2753-7064/2025.HT26990

more exposure. Their classification of noun types (common vs. proper, countable vs. uncountable) is still unclear, leading to overgeneralization.

5. Conclusion

This paper employed the CHILDES corpus system to investigate the development change rules of English quantitative word use in children aged 4-8 years old. The findings revealed that there was a gradually increasing age rule, that is, the accuracy improved with age; children with less cognition and language experience will make more errors. Different types of mistakes vary by age level; younger children made more misuse and omission mistakes, while older children had more varied mistakes when understanding complex quantity words.

But at the same time, some deficiencies exist in this experiment, including the following: first, its main corpus comes from one aspect of the CHILDES corpus, so its sample diversity is narrow and cannot reflect the application situation of different culture areas in word quantity; Second, its error cause analysis involves internal cognitive development level, excluding external influences such as social interaction and education intervention. Finally, no detailed comparative study has been carried out on the error phenomenon of quantized words between modules like tense and pronoun. In the next stage of experiments, we should do more comparative experiments between cross-cultures and languages and verify its universality and speciality. Through long-term tracking, analyze how family language input and kindergarten teaching influence children's pathogenesis; the last one is to build a quantification word error model compared with other grammar modules and finally reveal the whole picture of children's language system pathogenesis process.

References

[1] Li, Y. (1995) The Development of Children’s Language. Central China Normal University Press.
[2] MacWhinney, B. (2000) The CHILDES Project: Tools for Analyzing Talk. Lawrence Erlbaum Associates.
[3] Piaget, J. (1981) Principles of Genetic Epistemology. The Commercial Press.
[4] Vagh, S.B., Pan, B.A., Mancilla-Martinez, J. (2009) Measuring Growth in Bilingual and Monolingual Children’s English Productive Vocabulary Development: The Utility of Combining Parent and Teacher Report. Child Development, 80(5): 1545-1563.
[5] Li, H. (2005) Developmental Characteristics and Influencing Factors of Quantifier Use in 3- to 6- Year-Old Children. Psychological Development and Education, (2): 32-36.
[6] Zhou, J.Y. (2017) Error Analysis of English Native Speakers’ Acquisition of Chinese Classifiers and Teaching Strategies. Higher Education Teaching, (20): 94-96.
[7] Hu, Y.L. (2020) Error Analysis of Chinese Borrowed Classifiers Acquisition by English-Speaking Students. Journals Related to Master’s Degree Theses in International Chinese Language Education (if formally published).
[8] Brown, R. (1973) A First Language: The Early Stages. Harvard University Press.
[9] Guo, X. (2000) A Cognitive Study of Chinese Classifiers. Language Teaching and Linguistic Studies, (3): 34-40.
[10] Shi, Y.Z. (2000) The Cognitive - Semantic Basis of Grammar. Jiangxi Education Press.
[11] Liu, Y. (2017) Children’s Language Development and Education. Education Science Press.
[12] Xing, F.Y. (2019) Chinese Grammar. The Commercial Press.