Modern Management Based on Big Data IV A.J. Tallón-Ballesteros (Ed.) © 2023 The authors and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/FAIA230213

Spoken Word List Development for Children's Second Language Study Based on Expression Category

Jingyi XIE¹

ICCS, Fujian Normal University, Fuzhou, China, 350117

Abstract. The study of spoken word lists is critical for the Chinese language learning by foreign children. At the same time, spoken corpora of native children's language provide an important basis for developing such word lists. Word list development is based on more than word frequency: other important factors must be considered. By analyzing a corpus of children's spoken language, this study proves that the aggregation of children's spoken vocabulary is related to expression categories, as well as to their native language. Furthermore, children's spoken vocabulary changes with age as follows: generalization gradually disappears; the number of words increases, and their meaning changes; a functional change in the use of sentences with common words occurs. The paper concludes by summarizing some core recommendations for children's spoken word lists: gathering words according to themes; dividing polysemy and collocation blocks according to age; adding and subtracting words according to different native language backgrounds.

Keywords. Spoken word list, children's language development, children's Chinese proficiency grading standards.

1. Introduction

1.1. Research Background

There has been an increase in young Chinese learners overseas in recent years. The average percentage of young learners is over 50%, with some countries reaching or exceeding 60% [1]. However, children's cognition and use of language are distinct from those of adults, making children's learning process of Chinese as a second language different and unsuitable to the standards applied to adults. Alternatively, the approach to teaching children a second language should be based on the same language expression system found among native speakers of the same age.

Research on children's brain and language development suggests that their first language begins to mature at three, and they establish a basic grammar system by age 5 [2]. The language and range of expression produced by children speaking their mother tongue at this age can be a reference for children learning that tongue as a second language [3].

¹ Corresponding Author: Jingyi Xie, International College of Chinese Study, Fujian Normal University, Fuzhou, China, 350117. Email: xjyfjnu@fjnu.edu.cn

This article addresses the corpora of children's spoken language to uncover the word distribution and development patterns. It discusses methods for developing children's second-language spoken word lists from the following perspectives: (1) the lexical distribution characteristics of children's spoken language; (2) the relationship between children's second and native languages; and (3) the development characteristics of children's languages.

1.2. Related Research

In this section, previous research findings regarding the development of children's spoken word lists are discussed in terms of the relationship between word lists and word frequency, lexical classification and the clustering of children's spoken words, and the development of meaning in children's spoken language.

The relationship between word list development and word frequency. Several language experiments have confirmed the effect of word frequency in the target language on second language learning [4]. There was general agreement that high-frequency words in the target language should be mastered first [5]. Dolch's "A Basic Sight Vocabulary", a formative text in the international community, was also based on word frequency. However, word lists based purely on word frequency have certain defects. Research regarding Chinese second-language teaching has also suggested that frequency statistics should not be limited to word frequency but should rather be expanded to encompass meaning [6].

The clustering of words in categories within children's spoken vocabulary. The linguistic expressions used by children aged 3 to 5 are closely related to the needs of daily life, and the vocabulary used to express these needs can be aggregated to form spoken language expression categories. Many children's word lists are also organized according to the clustering of their spoken vocabulary [7]. In best-selling books designed for children to learn English vocabulary, the vocabulary is divided into 28 categories, such as daily necessities, food, furniture, toys, clothing, kinship figures, places, natural phenomena, substances and materials, weather, seasons, plants, animals, time and date, body parts, orientation nouns, properties of things (colors and shapes), proper nouns, occupations, sports, thinking and abilities, transportation, and holidays and traditions. This is the basis for categorizing the vocabulary later on, as shown in studies [8-10].

The development of meaning systems in children's oral expressions.

Children's language development is intimately related to their cognitive development. Their semantic systems and the words they represent develop with age. Concerning children's acquisition of polysemic verbs, a positive correlation has been found between the time of acquisition and the frequency of use [11]. They will likewise use high-frequency tenses (generally prototypes) to refer to low-frequency tenses directly. For example, the acquisition of children's modal verbs shows a strong developmental trend that works from "root modality", especially "dynamic modality", to "cognitive modality" [12].

2. Study Background and Approach

The paper draws upon various corpora containing videos of daily spoken interactions between Chinese children and adults. After transcription, the combined corpus was divided into turns, and the number of words was counted. The results incorporate more than 20,000 turns and

280,000 words. The primary sources for the combined materials were the transcribed audio and video-recording-based corpus CHILDES (Child Language Data Exchange System).² The recordings of children whose native language was Chinese were selected. The study covered 14 children, aged 4.3 to 5.9 years old, from Beijing, Nanjing, Hangzhou, and other places in China. The corpus mainly contained conversations between children, parents, and researchers undertaking activities such as playing games, reading picture books, and telling stories. We used a total of about 30 hours of transcribed recordings. After the segmentation of the recordings, 16,960 turns were obtained, with a total of 121,609 words.

2.1. The clustering of high-frequency words in different categories of children's oral expressions

In the oral expressions children use in everyday life, there is often a connection of meaning and topic between high-frequency verbs and high-frequency nouns. An overall analysis of the corpus also showed that high-frequency words covered a wider range of the content than low-frequency words but that the trends relating to the coverage of the verbs and nouns differed. The coverage of high-frequency nouns grew slower than that of verbs. Concerning the general and word list-based corpora, the top 6.93% and 32.31% of most frequent nouns covered only 50% and 80% of the expressions, respectively. The nouns covered a total of 28 categories, among which the high-frequency nouns were clustered in categories relating to daily necessities, food, places, and natural phenomena, which are closely related to children's lives. These exceeded 50% of all nouns in the main corpus and 80% of all nouns in the word list-based corpus, so these comprise the main components of noun vocabulary in children's oral expressions.

2.2. Commonalities and differences in how vocabulary is used in different categories of children's oral expressions

There are core elements of human language that are universal. Word clusters that express basic conceptual categories, such as people, food, shapes, weather, etc., exist in almost all languages. Children all acquire language in the context of particular social circumstances. During this process, they are inevitably affected by their specific linguistic and socio-cultural environment, which will differ from place to place.

According to our analysis, there were 22 verbs representing actions and behaviors in the English children's spoken corpus. In comparison, there were just 14 verbs in the oral expressions of Chinese native-speaking children. Nonetheless, some notable overlaps existed between the Chinese and English high-frequency verbs. Another point worth noting is that food nouns such as "rice" (*mifan*), "noodles" (*miantiao*), and "soup" (*tang*) in Chinese are not present in Dolch's "Noun List".

² https://childes.talkbank.org/

3. Longitudinal Analysis of the Development of Word Meaning Systems in Children

3.1. Expansion in the number of high-frequency verbs

As the number of words in children's oral expressions increases with age, the number of words in children's oral word lists must also expand as they age. However, this expansion should involve increased quantity, concretization, and refinement of vocabulary content.

The performed analysis of the corpus revealed that the range of expressions covering different functions in children's spoken language reached 85% of that of adults by the time they were 5. Children's vocabulary growth was not just learning 31 more words: their expression categories exhibited a permanrent dynamic change. For example, "want" (*vao*) in the children's spoken language corpus could be matched with almost any noun, e.g., "baby wants red"; "little mouse wants mother"; "I want rice"; and so on. These expressions in adult spoken language correspond to such forms as "give me the red one", "little mouse misses its mother", and "I need a bowl of rice", respectively. Therefore, it can be seen that thwe verb "want" (*vao*), in younger children's oral expressions, "shares" meaning and functionality with such verbs as "give" (*gei*), "miss" (*xiang*), and "need" (*xuyao*).

3.2. Changes in the function of polysemic verbs in children's oral expressions

When children and adults use the same verbs, there are differences in the distribution of functions being expressed. The choice of a certain category of expression and its representative words needs to be modulated according to age [13]. For example, at the age of 3 years, there is no need to pay special attention to selecting categories related to guiding others to complete a message. This choice becomes more relevant for children aged five and above. At the same time, when guiding others, adult educators should be aware that their common vocabulary differs from children's. For example, adults like to give direct orders to others, e.g., "Go eat..." and "Go get...", while children may prefer to say "I want...", which is equal to "Help me to...".

3.3. Vocabulary growth and changes in the meaning categories of nouns

As a child's overall vocabulary increases, the range of words in different sense categories sometimes grows at different rates. We found four patterns in the growth of noun vocabulary. The proportion of nouns in the "daily necessities" category steadily declined (from 26.6 to 11.9%). The proportion of nouns in the category of "natural phenomena" skyrocketed dropped from 2.9 to 18.6%³. "Time and date" nouns increased rapidly between 3 and 4 years of age but remained more or less the same between 4 and 5 years (the usage frequencies were 2.9, 9.1, and 9.0%, respectively). The proportion of nouns in the "relatives and friends" category remained almost unchanged (the frequencies were 9.8, 9.4, and 9.3%, respectively). Thus, when screening high-frequency words for their use in word lists, it is important to consider what changes occur in children's attention span and cognition as they get older, the implications for changes in categories, and how they select corresponding words.

³ The increase in the frequency of "celestial nature" nouns here is partly because the words "heaven", "earth", and "star" are polysemous in Chinese, and children acquire more of their meanings as they grow up, leading to an increase in their usage frequency.

4. Principles for the Compilation of CSL Word Lists for L2 Children

The analysis of children's spoken language corpora in this paper shows that in developing CSL Word Lists for children, such parameters as age and cultural background need to be considered in addition to word frequency.

Based on the above findings, we propose the following recommendations regarding the design of word lists:

4.1. The development of word lists should be based on the age characteristics of children

Children's language has age-dependent characteristics, so child-oriented teaching, exam syllabuses, and word lists must be matched with their level of language development. For example, we have found some differences in the frequency distribution of verbs between adults and children. This suggests that when compiling a relevant syllabus for teaching children Chinese verbs, their level of language development must be considered, in addition to the survey results related to global word frequencies.

Concerning the syllabus and teaching content organization, it is impossible to arrange the teaching sequence according to word frequency directly. Most Chinese verbs have a polysemic structure, e.g., the auxiliary verbs and the structures associated with the verb "is" (*shi*). In teaching these verbs and related structures, their topic and function must be combined, and attention must be paid to the recurrence of high-frequency words with different meanings. A more appropriate teaching sequence needs to be developed that properly recognizes the functional preferences in children's spoken expressions.

4.2. The development of word lists should meet the needs of individualization and localization

This research confirms that children's common vocabulary revolves around their everyday sphere and has a distinctly individual character. Children only care about their own information at this stage of cognitive development. Therefore, to arouse their interest, starting with things around them is necessary. One can then move on to promoting cognitive development, improving cognitive ability, and ultimately improving language ability [14]. Therefore, when planning the instruction content for children and developing a list of bilingual words, one should consider the characteristics of their daily life. The geographical area where the children are located, the focus of their family life, the content of their daily play, etc., can be used as objects in the planning of their teaching. A reasonable next step should be developing various teaching strategies around "can-do" lists to form local vocabulary lists. Beyond this, in the actual use of vocabulary in various countries, the specific content must also be adjusted to reflect the customs and habits of the country where the children are located.

5. Conclusion

Given the findings of this paper, it is not advisable to use adult corpora to determine children's second language learning word lists. The performed analysis of children's native language corpora, the clustering of children's use of verbs and nouns, and differences in children's word choice in different native language contexts revealed significant differences between children's and adults' vocabulary use. The follow-up study envisages expanding the analysis of children's vocabulary and further exploring the development of comprehension, the expression of different functions, and the use of grammatical categories to compile more scientifically grounded word lists that would allow children to accomplish richer functional expressions with fewer words.

References

- Li Y.M. Reflections on the younger age of overseas Chinese learners. Chinese Teaching in the World (2018).
- [2] Liu L. Q., Zhang H. L. W., Effects of bilingual environments on children's early language development, Chinese Language Strategies, December 2022.
- [3] Bergen D., Coscia J. Brain Research and Childhood Education: Implications for Educators, Parents, and Society (Chinese Edition). Assn for Childhood Education Intl, 2001, ISBN-13: 978-0871731548
- [4] Li Y.M. The development of children's language. Wuhan: Huazhong Normal University Press, 1995.
- [5] Farrell L., Osenga T., Hunter M., Comparing the Dolch and Fry High-Frequency Word Lists, Readstters,
- [6] Li Y. M. Common destiny, common language. "Guangming Daily", 2020/08/22.
- [7] Zhang Y.Q., Zhou J. S., Fu J. Acquisition strategies of polysemy in early Chinese children: A case study of Beijing dialect children. Chinese Language and Literature (2010).
- [8] Zhang Y. Q., Liang Y.X. The early development of children's modal semantics and the significance of demonstration. Chinese Learning (2021).
- [9] Ellis N.C., Simpson-Vlach R., Raynard C., Formulaic language in native and second language speakers. Psycholinguistics, corpus linguistics, and TESOL (2008).
- [10] Krashen. S. and Terrell, T.D. The Natural Approach. Prentice Hall Europe (1998).
- [11] Widen S. C. & Russell J. A. A closer look at preschoolers' freely produced labels for facial expressions. Developmental Psychology (2003).
- [12] Cui Na. Discussion on classification and grading of polysemous morphemes in teaching Chinese as a foreign language. Education and Teaching Forum (2019).
- [13] Michael A. K. Halliday. Infant language(Chinese Edition). Peking University Press, Beijing (2015).
- [14] Yan Y.M. Introduction to Information Science. Wuhan University Press, Wuhan (1994).