

A Cognitive Research on the Lexicons in the Chinese Textbooks of Primary Schools

Meng Wang¹, Ge Xu² and Lulu Wang³

¹School of Humanity, Jiangnan University
No. 1800, Lihu Avenue, Wuxi 214122, P. R. China
wangmengly@163.com

²Department of Computer Science, MinJiang University
No.200, Xiyuangong Road, Fuzhou 350108, P. R. China
xuge@pku.edu.cn

³School of Literature, Communication University of China
No.1 Dingfuzhuang East Street, Chaoyang District, Beijing, 100024, P.R. China
lulu.wang@cuc.edu.cn

Received March 2015; revised March 2015

ABSTRACT. This paper investigates the lexical semantic distributions of the lexicons in the Chinese textbooks of primary schools. Based on the Chinese Semantic Dictionary, the nouns in the textbooks are mapped into 5 semantic categories: Entity, Abstraction, Process, Time and Space. The data shows that the semantic distributions of the lexicons in the textbooks are adapt to the cognitive development of the primary students.

Keywords: cognition; textbooks; lexicons; semantics

1. Introduction. Textbooks, one of the three important elements in the teaching structure, are the basic elements for teachers organizing teaching activities. And it is also an indispensable knowledge resources for the students. Textbooks, as an integration of teaching goals, contents and models, reveal the characters and features of education, thus becoming one of the paramount carriers of information in the process of teaching and learning.

The construction of textbooks focuses on the organization and presentation of information that is mainly conveyed by the language presented in textbooks. Consequently, the accuracy, order, popularity and profession of the language in textbooks will directly work up on the clarity and systematisms of the corresponding course and its contents ^[1]. As

a specific phenomenon of language and an important part of language resources, textbook language has become a subject of research and an increasing concern to the government and the academia¹. SU Xinchun (2007) defines the conception of the textbook language, and makes clear of the subject orientation, nature, characteristics and significance of the textbook language for research^[2]. He also points out that the composition of textbooks is inseparable from the application of language and the study of textbook language is therefore inseparable from imparting information to students.

The mode of “one syllabus and multi-versions” is implemented in the composition of Chinese textbooks for primary students in China and nearly 50 different versions are officially published at the moment. Standards of Compulsory Education for Chinese Course (2011 edition) stated that textbooks should be consistent with the students’ physical and mental characteristics and adapt to their cognitive levels. What's more, textbooks should also be in close connection with students’ experience and imagination that stimulate the studying interest and the innovation spirits. Since lexicons in the textbook are the carriers of information and knowledge of a course, the selection of the lexicons should be in consistence with the students’ cognitive development. In this work, we focus on the lexicons in the Chinese textbooks of primary schools published by the Jiangsu Education Publishing House (hereinafter referred to as Jiangsu Edition). We collect the semantic information of the nouns in the textbooks of Jiangsu Edition and analyze the features in choosing these lexicons in terms of the cognitive perspectives. Hopefully this paper will provide some useful advice for the compiling of Chinese textbooks. Furthermore, the quantitative analysis from real data can benefit those researchers who are concerned of this area.

2. Review of the Research on the lexicons in the Chinese Textbooks. Driven by the development of the computer technology, the construction of corpus and corpus linguistics flourish in recent years. As a result, obtaining massive electronic language resource is no longer a problem. If a collection of all textbooks is considered as a corpus, then a variety of quantitative analysis methods could be applied in textbooks to meet different needs in research. Therefore, the construction of textbook corpus for multi-applications has become the most fundamental and important research field. In the respect of textbook corpus construction, National Language Resources Monitoring and Research Center Education and Teaching Material Language Branch has built a series of corpus including Chinese textbooks, all subjects textbooks, Chinese as a second language textbooks and so on. All of them can make contribution to language education and teaching, textbook compiling and language information processing. Moreover, Chinese Information Processing Research Center of Ludong University has built the Chinese textbooks corpus, which is segmented and POS-tagged. Meanwhile, the syntactic and semantic information are also annotated to the corpus^[3]. In addition, the number of specialized textbook corpus for specific research

¹ In June 2005, the Department of Language Information Management of Chinese Ministry of Education and Xiamen University jointly established the National Language Resources Monitoring and Research Center Education and Teaching Material Language Branch.

purposes is growing every year [4][5].

Early studies on Chinese textbooks focused primarily on Chinese characters such as frequency statistic [6][7] and Chinese radicals statistic [8]. As natural language processing technologies such as word segmentation and POS-tagging get more and more advanced, they are gradually applied to textbook corpus. As a result, new findings and research results have been generated so far. For example, researchers can count word frequencies, extract common vocabularies, analyze the lexical constitution system of textbooks, and calculate new words repetition rates and so on [9].

3. The Semantics of the Lexicons from a Cognitive Perspective. For primary school students, the development of language ability has a direct relationship to their cognitive levels. It is essential that the contents of the Chinese textbooks should be based on students' cognitive characteristics and that the selection of lexicons should be coincided with students' cognitive capability. In this section, we will introduce the cognitive characteristics of primary students firstly and then the semantic taxonomy adopted in this research.

3.1 The Cognitive Characteristics of Primary School Students. Children's language capability is an important part of their cognitive development, advancing with cognition simultaneously. Piaget, a Swiss psychologist, did a profound research on the relation between cognition and language development. He finally put forward with an influential theory of cognitive development. And he found that children's cognitive process can be mapped into four phases: Sensorimotor Stage (birth to a 2 years), Preoperation Stage (2 to 6/7 years), Concrete Operational Stage (6/7 to 11/12 years), and Formal Operation Stage(11/12 to 14/15 years). In china, the primary school students range from age 6 to 12, which corresponds to the stage of Concrete Operation in Piaget's cognitive theory. Children in the Concrete Operations Stage are able to take in another's point of view and take into account more than one perspective simultaneously. They can also represent transformations as well as static situations. Although they can understand concrete problems, Piaget would argue that they cannot yet perform on abstract problems, and that they do not consider all of the logically possible outcomes. In this stage, children have the ability to develop logical thought about an object, if they are able to manipulate it. By comparison, however, in the Formal Operations stage, the thoughts are able to be manipulated and the presence of the object is not necessary for the thought to take place. In a word, children in this stage begin thinking logically about concrete events, but have difficulty in understanding abstract or hypothetical concepts. As a result, their thinking is closely related to the foundation of sensible experience and reality.

As they grew older, primary students are gradually capable of recognizing objects and their main characteristics more accurately, along with the relations among different parts, getting rid of general and inaccurate cognition for objects. In terms of thinking and imagination, students at lower grades have a rich imagination, because they cannot clearly distinguish the difference between reality and imagination and always confuse them, while students at higher grades can already present real objects and convert their imagination into perception towards real world.

3.2 Semantic Taxonomy in the Chinese Semantic Dictionary. This paper adopts the semantic taxonomy of *Chinese Semantic Dictionary* (CSD), which is a large machine-readable knowledge base developed by the Institute of Computational Linguistics at Peking University (ICL/PKU). It provides a large amount of semantic information such as semantic hierarchy and collocation features for 66,539 Chinese words and their English counterparts. The descriptions of semantic attributes are fairly thorough, comprehensive and authoritative. Its semantic taxonomy represents the latest progress in Chinese linguistics and language engineering. CSD consists of one general database and six sub-databases. The general database contains shared attributes of all the 66,539 entries, while the sub-databases provide detailed descriptions of the distinctive semantic attributes associated with the parts of speech (POS). For example, the noun database has 15 attribute fields, and the verb database has 16 attribute fields. CSD offers a powerful support for many NLP applications, including machine translation, automatic abstraction, information retrieval, hypertext navigation, thematic analysis, and text processing..

CSD adopts a four-level hierarchical taxonomy for the semantics of nouns. In the top level, nouns are divided into 5 major categories: Entity, Abstraction, Process, Time and Space. The fine-grained sub-categories in the lower levels are also given in the work of [12]. Figure 1 illustrates the semantic taxonomy of nouns in CSD. Some sample words (in italics) in each category are also given in Figure 1. For example, words such as “建筑物 [jianzhuwu]building”, “创作物 [chuangzuowu]works”, “食物 [shiwu]food”, “衣物 [yiwu]cloth” and “票据[piaoju]bill” all belong to the artifact category.

In the following, we will exemplify the 5 major categories in detail:

(1) **Entity.** Entity includes human, animal, plant, microbe, artifact, natural objects, body-part and object-part and so on. Words like “教师 [jiaoshi](teacher)”, “母亲 [muqin](mother)”, “老虎[laohu](tiger)”, “医院[yiyuan](hospital)”, “牡丹[mudan](peony)”, “面包 [mianbao](bread)”, “月亮 [yueliang](moon)”, “帽子 [maozi](hat)”, “河流 [heliu](river)” and “耳朵[eduo](ear)” belong to the Entity category.

(2) **Abstraction.** Abstraction includes attribute, information, field, rule, physiological state, psychological feature and motivation. Words such as “质量[zhiliang](quality)”, “体积 [tiji](volume)”, “勇气 [yongqi](courage)”, “状态 [zhuangtai](state)”, “声明 [shengming](declaration)”, “艺术 [yishu](art)”, “合同 [hetong](contract)” and “态度 [taidu](attitude)” belong to the Abstraction category.

(3) **Process.** Process consists of event and natural phenomenon. Words such as “早餐 [zaocan](breakfast)”, “战争 [zhanzheng](warfare)”, “风暴 [fengbao](storm)”, “雷 [lei](thunder)” and “闪电[shandian](lightning)” belong to the Process category.

(4) **Time.** Time involves specific and relative time. Words such as “宋朝 [songchao](Song Dynasty)” and “清代[qingdai](Qing Dynasty)” are specific time, while “昨天[zuotian](yesterday)”, “今天[jintian](today)” and “当代[dangdai](contemporary)” belong to the relative time category.

(5) **Space**

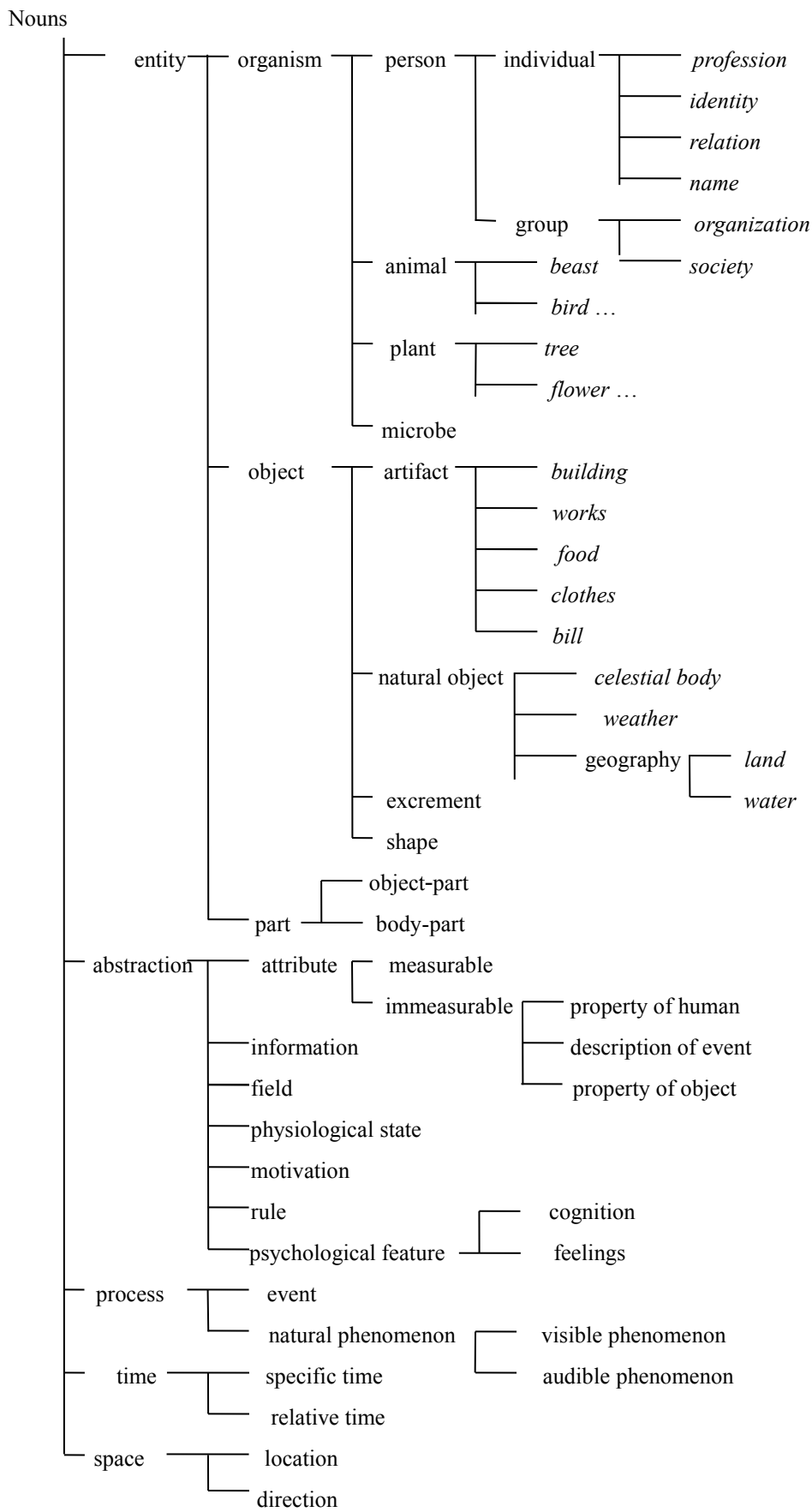


FIGURE 1. Semantic taxonomy of Nouns in CSD

4. Data Analysis of the textbooks

4.1 Size of the Vocabulary. The corpus we used in this research is provided by the Education and Teaching Material Language Branch of the National Language Resources Monitoring and Research Center. There are altogether 12 volumes that are segmented and POS-tagged. The vocabulary size is an important indicator of the language ability of students. Table 1 gives the number of lexicons in all the textbooks. The data shows that the vocabulary size enlarges with the rise of grade. There is an obvious increase in the second semester of the first grade and it reaches the maximum in the sixth grade.

TABLE 1. VOCABULARY SIZE

Volume	Number of Types	Number of Tokens
1	378	809
2	898	2456
3	1019	2545
4	1389	3715
5	1854	5017
6	2221	6369
7	2674	6687
8	2553	7520
9	3106	9007
10	3728	10858
11	3916	11901
12	3292	9929
Total	27028	76813

4.2 Semantic Distribution of Nouns in the Textbooks. As an important part of the content words, nouns have always been an attractive topic in the field of linguistic studies. The nouns can refer to the objects or abstractions like people, substances, issues, time, space, concepts and so on. Children cognize the world by interpreting what the nouns refer to. Among all the lexicons in the textbooks, the nouns take the largest proportion, accounting for 26.57%. Thus the focus of this paper is on the semantic distribution of nouns of Chinese textbooks in primary school. Based on the *CSD*, the nouns are mapped into 5 semantic categories: Entity, Abstraction, Process, Time and Space. The result is presented in Table 2.

TABLE 2. THE SEMANTIC DISTRIBUTION OF NOUNS (THE NUMBER AND PROPORTION ARE SEPARATED BY “/”)

Volume	Types of Nouns	Entity	Abstraction	Process	Time	Space
1	197	165/83.76	10/5.08	6/3.05	6/3.05	10/5.08
2	395	312/85.48	40/10.96	14/3.84	6/1.64	23/6.30
3	482	404/83.82	33/6.85	8/1.66	13/2.70	24/4.98
4	763	591/77.46	71/9.31	18/2.36	41/5.37	42/5.50
5	1131	861/76.13	124/10.96	41/3.63	34/3.01	71/6.28
6	1299	996/76.67	134/10.32	40/3.08	48/3.70	81/6.24
7	1462	1015/69.43	178/12.18	44/3.01	68/4.65	157/10.74
8	1568	1186/75.64	127/8.10	60/3.83	71/4.53	124/7.91
9	1840	1306/70.98	264/14.35	47/2.55	100/5.43	123/6.68
10	2427	1752/72.19	341/14.05	28/1.15	113/4.66	193/7.95
11	2310	1599/69.22	366/15.84	28/1.21	94/4.07	223/9.65
12	2026	1429/70.53	318/15.70	55/2.71	65/3.21	159/7.85
Total	15870	11616/73.19	2006/12.64	389/2.45	659/4.15	1230/7.75

From Table 2, we can see that:

(1) Among the five categories, the Entity takes the largest proportion in each of the volumes. But it decreases from 83% to 70% with the rise of grade.

(2) The Abstraction comes to the second in proportion and its number increases with the rise of grade, thus its proportion tends to climb too.

(3) With the growth of age, the students can use more and more nouns of Time and Space to express their intention more exactly and simply.

(4) The proportion of the Process increases first and decreases afterwards. It reaches the peak in the fourth grade. But the variation is not big.

The lexicon acquisition of primary students is closely related to their development of conception. The students of the first and second grade are more easily to accept the concepts with intuitive feelings like “父亲 [fuqin]father”, “母亲 [muqin]mother”, “草 [cao]grass”, “花 [hua]flower”, “河流 [heliu]river” and so on, because these are the concepts they have already known in daily life. With aging, their cognition enriches with their experience and they can gradually understand complex and abstract concepts. In the textbooks of the fifth and sixth grade, more and more lexicons of Abstraction are included. For example, words like “秘密 [mimi]secret”, “成就 [chengjiu]achievement”, “贡献 [gongxian]contribution” and “和平 [heping]peace” are added in. About the nouns of Time, the students at junior grades usually use some specific periods of time like “早晨 [zaochen]morning”, “晚上 [wanshang]evening” and “生日 [shengri]birthday”, while students at senior grades can use some abstract time concepts with wide range like “年轻 [nianqing]youth”, “时代 [shidai]time”, “世纪 [shiji]century”, “历程 [licheng]course” and so on. About the nouns of Space, most of the lexicons appear in the textbooks of the first and

second grades are intuitive and concrete concepts like “公园[gongyuan]park”, “家乡[jiexiang]hometown”, “地方[difang]place” and so on. With the rise of grade, the students can use more nouns including some abstract space concepts. Thus words like “边缘[bianyuan]edge”, “海域[haiyu]sea area”, “空隙[kongxi]interspace” and “天宫[tiangong]welkin” appear later in the textbooks of fifth and sixth grade.

5. Conclusion. This paper investigates the lexical semantic distributions of Chinese textbooks in primary schools. The results show: 1. The proportion of abstract nouns increases with the rise of grade. 2. Entity nouns decrease gradually. In general, the arrangements of the lexicons in the textbooks are in line with the development of students' cognition development. Lexicons are not only one of the most important basis of compiling textbooks, but also the resources crucial to most teachers during the teaching activities. We hope that our research could arouse more valuable research interests on Chinese textbooks in primary schools and course design.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (No.61300152, 61300156). We would like to thank the National Language Resources Monitoring and Research Center Education and Teaching Material Language Branch which provides the high quality textbook corpus for our research. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

- [1] YU Hong, Research on the Uncertainty of Chinese Curriculum's Content from Language of Teaching Material, Journal of Tianjin Normal University (Elementary Education Edition), Vol. 12 No. 3, pp.33-37, 2011.
- [2] SU Xinchun , DU Jingjing , GUAN Junhong and ZHENG Shuhua, The Nature, Characteristics and Significance of Research of Textbook Language, Applied Linguistics, No.4, pp.86-91, 2007.
- [3] KANG Shiyong, LIU Jinfeng, LV Lingna, Study on the Annotation of Chinese Textbooks in Primary and Middle Schools, The first Symposium on the Language of Education and Teaching Materials, Xiamen, 2006.
- [4] YU Long and Tao Benyi, Problems and Coping Strategies of Characters Teaching: A Corpus-based Study of the Usage of Chinese Characters in Primary School Textbooks, Applied Linguistics, No.1, pp.114-124, 2010.
- [5] GUO Shulun, Contrasting the Lexical Profiles of Large-scale and Small-scale Chinese Corpora, Applied Linguistics, No.2, pp.130-141, 2009.
- [6] LI Tang, Distribution of characters and words in the texts of Chinese courses at primary and middle schools: statistics and its applicable value, Applied Linguistics, No.3, pp.21-24, 2000.
- [7] ZHOU Meiling and HAN Jie, The Statistics and Analysis of Characters of the Basic Education New Curriculum in the Bnup Editions—Taking the Chinese teaching materials as an example, Journal of Jiaying University(Philosophy & Social Sciences), Vol. 31, No. 9, pp.76-80, 2013.

- [8] XING Hongbing and SHU Hua, A Statistical Study on the Basic Components of Chinese Characters in Chinese Textbooks for Primary Schools, *Applied Linguistics*, No.3, pp.72-80, 2008.
- [9] CHEN Bo, On the Lexical Constitutions of Chinese Textbooks in Primary School, *Journal of Xiangfan University*, Vol. 27 No. 6, pp.81-86, 2006.
- [10] SU Xinchun, Comparative Study of GOTCFL and the Glossaries in Two Textbooks, *Applied Linguistics*, No.2, pp.103-111, 2006.
- [11] SU Xinchun and Gu Jiangping, The Distribution State of “Making a Thin Pancake” of the Vocabulary in Chinese Textbooks, *Journal of Jiangxi Science & Technology Normal University*, No.4, pp.64-71,2009.
- [12] WANG Hui, ZHAN Weidong and YU Shiwen, Structure and Application of The Semantic Knowledge base of Modern Chinese, *Applied Linguistics*, No.1, pp.134-141, 2006.
- [13] ZHOU Xiaobing and LIU Yali, Research into Vocabularies of Chinese Elementary Intensive Reading Textbooks, *Language Teaching and Linguistic Studies*,No.5, pp.26-33,2012.
- [14] DU Ying, Acquisition of Children’s First Words and Their Semantic Analysis, *Language Teaching and Linguistic Studies*, No.3, pp.34-39, 2003.
- [15] Jean Piaget, *The Principles of Genetic Epistemology*, The Commercial Press, 1997.
- [16] LI Yuming, *The Development of Children’s Language*, Huazhong Normal University Press, 1995.