

Assessing Difficulty Levels of Chinese Idioms for Chinese L2 Learners

Jiangying Yu*

Yunnan Open University, Kunming, Yunnan, 650223, China

*yujiangying2008@qq.com

Keywords: Difficulty Levels, Chinese Idioms, Chinese L2 Learners

Abstract: The most important forms of Chinese idioms, chengyus (CYs), have a fixed length of four Chinese characters and have long attracted the attention of linguists, lexicographers, translators, language instructors. The work presented here work attempts to provide a quantitative perspective of assessing difficulty levels of Chinese idioms for Chinese L2 learners. The approach is based on three dimensions. One is the semantic parameters, which attributes mainly from the view of the relation of the literal meaning and deeper meaning. The second is the syntax parameters, which is mainly divided into syntactic structure and syntactic function. Third is the pragmatic parameters, which takes sentimental color, cultural color and special restriction and collocation of Chinese idioms into consideration. Finally, a general approach to assess the difficulty levels of Chinese idioms for Chinese L2 learners is proposed.

1. Introduction

The most important forms of Chinese idioms, chengyus (CYs), have a fixed length of four Chinese characters and have long attracted the attention of linguists, lexicographers, translators, language instructors. Some CYs ,for instance,“遍地开花”[biàn dì kāi huā](blossom everywhere) can be understood through the constituent words, but most CYs ,for example,“斩钉截铁”[zhǎn dīng jié tiě](speak or act with determination and courage) seems to have nothing to do with the conventional meanings of the words of which they are composed. That is to say, most CYs are characterized by full or partial semantic non-compositionality. On the other hand, CYs are largely fixed both lexically and syntactically. Consequently, It is the low semantic transparency that leads to the difficult problem of Chinese as a second language teaching and learning, as well as a focus and difficulties that natural language processing faces, and even a tricky task for the compilation of Chinese CYs dictionary faced to Chinese L2 learners.

So this put forward questions to us, which CYs are less-well understandable, and which are more easily understandable? How to judge an idiom is difficult or easy for Chinese L2 learners? These questions above converge at one point that difficulty levels of CYs should be assessed for Chinese L2 learners. We believe that research as such will have benefit on teaching Chinese as a second language and lexicography, Chinese linguistics study. On the other hand, our work may also contribute to NLP tasks like automatic metaphor recognition and processing, semantic role labeling etc.

2. Related Work

In the last two decades, the study of CYs for Chinese L2 learners has focused on three aspects: firstly, the importance of CYs in TCSL (Hong Bo,2003; Pan Xianjun, 2006; Zhang Wenyi, 2006; Shi Lin,2013); the second are CYs errors and reasons why learners made such errors, in the hope of providing practical assistance to the Chinese learning of L2 learners(Zhang Yongfang, 1999;Li Yanling, 2006; Shi Lin,2013, etc.); And third is compilation Chinese idioms dictionary for Chinese L2 learners(Wang Ruojiang, 2001;Zhang Shuai,2009; Yang Yuling, 2011).

Although it is generally agreed by the linguistics that the compilation of a suitable Chinese idioms dictionary for Chinese L2 learners is at all imminent, unfortunately, the dictionary has not been

published yet. The biggest challenge of lexicography is that it may be hard to determine definitions and annotations, which offer different levels of details and difficulty of each idiom. At present, the research of difficulty coefficient of Chinese idioms targeted for Chinese L2 learners is relatively few. This is not only the main difficulty of our work, but also the focus. We tried to start from the learner's errors(all errors examples were collected from our Chinese teaching process):

*(1)他没有主见，囫囵吞枣地听信别人的说话[tā méi yǒu zhǔ jiàn, hú lún tūn zǎo dì tīng xìn bié rén de shuō huà]。(不明特殊限制。“囫囵吞枣”多指读书不认真，不加分析思考，含贬义。)

He had no definite view, galloped listen to what others say. {Unknown special pragmatic restrictions:“囫囵吞枣”[hú lún tūn zǎo](swallow a date whole) means to read carelessly, without analyzing and thinking, including derogatory meaning. }

*(2)上课时不要轻举妄动[shàng kè shí bú yào qīng jǔ wàng dòng]。(语义偏误，按成语字面理解，指不经慎重考虑，轻率地采取行动，含贬义。)

Make no move without careful thought in class. {Semantic error: “轻举妄动”[qīng jǔ wàng dòng](act rashly and blindly) means to act thoughtlessly, including derogatory meaning. }

*(3)为了赚很多钱，他东奔西走了各地[wèile zhuàn hěnduō qián,tā dōng bēn xī zǒu le gè dì]。(不明成语的语法功能。“东奔西走”不带宾语。)

He rushes around every day to make a lot of money. {Unknown special syntactic function:“东奔西走”[dōng bēn xī zǒu](going in all directions for sth) cannot be followed by objects. }

*(4)我们班同学相敬如宾，关系十分友好[wǒmen bān tóng xué xiāng jìng rú bīn, guān xì shí fēn yǒu hǎo]。(不明限制，使用对象错误。“相敬如宾”仅限制在夫妻间。)

Our classmates are respectful and friendly. {Unknown special pragmatic restrictions: “相敬如宾”[xiāng jìng rú bīn](respect each other like guests)is restricted to couples only. }

*(5)他父亲虽然希望孩子自己要独立，但还是操心，就去学校探头探脑地看孩子[tā fùqīn suīrán xīwàng háizi zìjǐ yào dúlì, dàn háishi cāoxīn, jiù qù xuéxiào tàntóutànnǎo dì kàn háizi]。(不辨成语的感情色彩，“探头探脑”含贬义。)

Though he wanted his child to be independent, his father went to the school to pop his head in and look about. {Unknown sentimental color of the idiom:“探头探脑”[tàntóutànnǎo](crane one's neck to peer) has negative connotations. }

*(6)他七嘴八舌地说话，毫无根据，不要理他[tā qīzuǐbāshé de shuōhuà, háowúgēnjù, bú yào lǐ tā]。(不明“七嘴八舌”的主语一般是复数，不能是单数。)

He speaks without foundation, without understanding. {Unknown special pragmatic restrictions: the subject of “七嘴八舌”[qīzuǐbāshé](all sorts of gossip)usually plural and cannot be singular. }

*(7)他每天都来上课，今天突然逃课，大家都大惊小怪[tā měitiān dōu lái shàngkè, jīntiān tūrán táokè, dàjiā dōu dàjīngxiǎoguài]。(语义偏误，按字面义理解，也不明特殊限制。“大惊小怪”常用在否定结构里或含不以为然的语气。)

He comes to class every day, and all of us make a fuss when he skip class today. { Semantic error: “大惊小怪”[dàjīngxiǎoguài](get excited over a little thing)is often used in the negative structure or in a disapproving tone. }

*(8)老师在众目睽睽的时候，进了会场[lǎoshī zài zhòngmùkuíkuí de shíhòu, jìnle huichǎng]。(不明特殊搭配。该成语多与“在……下”搭配使用)。

The teacher came into the meeting in full view of the public. {Unknown special pragmatic collocation: “众目睽睽”[zhòngmùkuíkuí](under the watchful eyes of the people) is often used in collocation with“zài……xià”. }

*(9)他做事很小心翼翼[tā zuòshì hěn xiǎoxīnyìyì]。(不明成语本身已含有程度之义，不能再受程度副词的修饰。)

He works very carefully. {Unknown special pragmatic collocation: “小心翼翼”[xiǎoxīnyìyì](be very scrupulous) itself has a degree of meaning and can no longer be modified by the degree adverb. }

*(10)我在学习，他在我旁边熙熙攘攘不让我学[wǒ zài xuéxí, tā zài wǒ pángbiān, xīxīrǎngrǎng

bù ràng wǒ xué]。(语义不明。“熙熙攘攘”一般用于人多而喧闹。)

I am studying, and he is bustling beside me. {Semantic error: “熙熙攘攘”[xīxīrǎngrǎng](hustle and bustle)usually means people are crowded and noisy. }

While teaching Chinese to foreign students, we find the foreign students at middle and advanced level exists many errors in Chinese idioms and these errors often out of our expectation. Our students often made serious errors when using very simple Chi-nese idioms, which indicates that the difficulty of idioms is very different from our understanding of Chinese second language learners. Due to cultural difference, Chinese L2 learners have different understanding of our Chinese idioms. Therefore, we believe that it is of great theoretical and practical significance to assess the difficulty level of idioms from the perspective of Chinese L2 learners.

3. Difficulty Parameter Setting.

Setting difficulty parameter of Chinese idioms takes three steps to accomplish: One is to make a horizontal composition analysis of the Chinese idiom, and set the difficult parameter of the idiom; The second is to make longitudinal analysis of the Chinese idiom and make the difficult parameter values of the idiom. Thirdly, we quantitatively analyze idioms one by one according to parameter values, and then categorize them into three levels: easy, more difficult, and most difficult.

In order to close to the learners' learning experience, We selected 900 Chinese four-character idioms from four sets of teaching materials named Boya Chinese(Advanced ,3 volumes), Developing Chinese(Advanced, 2 volumes), XIANDAI HANYU GAOJI JIAOCHENG(3 volumes) ,SHIYONG HANYU GAOJI JIAOCHENG(4 volumes) and two sets of Chinese Syllabus , named National Syllabus of Graded Words and Characters for Chinese Proficiency, Chinese Syllabus for Foreign Students in Institutions of Higher Learning, as research and then we made a difficulty assessing study.

Table 1The proportion of CYs in textbooks and syllabuses

	Boya Chinese(Advanced ,3 volumes)	Developing Chinese(Advanced ,2 volumes)	XIANDAI HANYU GAOJI JIAOCHENG(3 volumes)	SHIYONG HANYU GAOJI JIAOCHENG(4 volumes)	National Syllabus of Graded Words and Characters for Chinese Proficiency	Chinese Syllabus for Foreign Students in Institutions of Higher Learning
words	2725	2315	2230	3213	8822	8024
CYs	277	351	369	504	120	249
proportion of CYs	10.16%	15.16%	16.54%	15.68%	1.36%	3.10%

Setting difficulty parameter of Chinese idioms, a key step in this study, needs to break up the structure of the idiom from the horizontal dimension. Therefore, this research sets up three idioms difficulty parameter items: semantic, syntactic and pragmatic parameters.

3.1 Semantic Parameters

We study idioms semantic parameters attributes mainly from the view of the relation of the literal meaning and deeper meaning. There are three kinds of relationship between the literal meaning and the deeper meaning, namely the original, extended and metaphor. Since most of the literal meaning of idioms is not the deeper meaning, semantic parameter is the major cause of the difficulty of a Chinese idiom. Mu Chunling (1998) confirmed that about 90% of Chinese idioms have implicit meaning, by carrying on sample gathering and statistical analysis about the implicit meaning of Chinese idioms. It accounted for more than 60% in 400 idioms errors we collected from learners due to not understanding of Chinese idioms deeper meaning, which is similar to the result Shi Lin (2008) studied idiom errors based on inter-language corpus. In addition, the character of idiom is hard to read, write and understand, the ancient Chinese semantic, syntactic, rhetorical characteristics retained

in the idiom, are also main reasons that cause learners to make errors, therefore, semantic parameters also should be combined with the content.

Table 2 Semantic Parameters and Values

Parameter Item	the Relationship between Literal Meaning & Deeper Semantic			Difficult Morpheme	
	original	extend	metaphor	Error prone words	Ancient Chinese characteristics
parameter values	1	2	3	2	3

3.2 Syntactic Parameters.

Syntax parameters mainly divided into syntactic structure and syntactic function two aspects. Chinese idioms are generally made up of four morpheme, idioms can be divided into the symmetry and asymmetry idiom structurally. Symmetry structure idioms are in the majority, the largest joint structure idioms, Chen Jie (2010) found that symmetry structure idioms accounted for 78% in A new dictionary of idioms. Previous studies (Huang Xiting, Chen Chuanfeng, 2000; Xia Xiuwen, 2009) suggest that it is faster for Chinese L1 learners to identify and recognize symmetry idioms than asymmetric idioms, so we focus syntactic structure parameters item mainly on symmetric structure and asymmetric structure.

Most Chinese idioms can serve as some parts in the sentence, which can be divided into six types: predicative idiom, attributive idiom, adverbial idiom, accusative idiom, complement idiom and paren idiom, Of which, predicative idioms are the most, followed by the attributive idioms and adverbial idioms. Xia Xiuwen (2009) classified syntactic functions of 120 Chinese idioms from National Syllabus of Graded Words and Characters for Chinese Proficiency based on about 150 million modern Chinese corpus of Peking University linguistics research center, which provides available thought and method for the study.

Table 3 Syntactic Parameters and Values

Parameter Item	Syntactic Structure		Syntactic Function			
	Symmetric	Asymmetric	Predicativity	Attributive	Adverbial	Accusative
parameter values	1	2	3	2	2	2

3.3 Pragmatic Parameters.

We take sentimental color, cultural color and special restriction and collocation of Chinese idioms into consideration for pragmatic parameters. In addition to the original meaning, some idioms also have a strong sentimental color, which has appreciatory and derogatory color, as well as a neutral one. Some idioms contains the rich Chinese cultural back knowledge, in which ancient fables, myths and legends, or historical story concentrated, are also the obstacles for Chinese L2 learners to understand. There are a lot of special restrictions and collocations, so we also take the fixed collocation into account for the parameter Setting.

Table 4 Pragmatic Parameters and Values

Parameter Item	Sentimental Color			Cultural Color		Special Collocation & Limitation	
	Appreciatory	Derogatory	Neutral	Yes	No	Yes	No
parameter values	2	2	0	3	0	2	0

In this research, we still need to set idiom difficulty parameter value to determine the degree of the difficulty. Synthesizing each kind of situation, we set three levels(easy, more difficult, most difficult) to use Numbers 1, 2 and 3 respectively.

Table 5 Difficulty Level of CYs

Difficulty Level	Degree of Difficulty	Difficult Value
Level1	Easy CYs	DV≤8
Level2	More Difficult CYs	9≥DV≤13
Level3	Most Difficult CYs	DV≥14

4. Steps of Assessing Difficulty Levels of an Idiom

The first step is to analyze the parameters item of the idiom and calculate the difficulty parameters value of the idiom. The second step is to assess the difficulty level of the idiom.

According to the difficulty parameters value is calculated, theoretically, the highest difficulty value is 19, the lowest is 4, most of the difficulty values are between 4 and 17, according to the scope of this value, the idiom of the difficulty level is set for three block, namely, easy, more difficult and the most difficult of idioms, the difficulty value difference between each grade is 5. For instance, according to the three-levels difficult assessing way, the difficulty value of the idiom“成千上万”[chéng qiān shàng wàn](tens of thousands of) is 4, while the idiom“唉声叹气”[āi shēng tàn qì](sigh in despair) is 7, both of them belong to level 1, the easy level. The difficulty value of the idiom“不远万里”[bù yuǎn wàn lǐ](make light of travelling from afar) is 9, while the idiom“众目睽睽”[zhòng mù kuī kuī](a crowd of people staring) is 11, and the idiom“相敬如宾”[xiāng jìng rú bīn](to be always courteous to each other like "guests") is 13, all of them belong to level 2, the more difficult level. The difficulty value of the idiom“滥竽充数”[làn yú chōng shù](hold a post without qualifications) is 17, it belongs to level 3, the most difficult level.

Table 6 the Parameter Content and the Difficult Value of CYs

CYs	Structure/Vales	Semantic/Vales	Syntactic Function/Vales	Color(Sentimental or cultural)/Vales	Difficult Morpheme/Vales	Special Collocation/Vales	Parameters Values in Total
唉声叹气[āi shēng tàn qì](sigh in despair)	Symmetric/1	Original/1	Predicativity/3		唉(哀)[āi]/2		7
众目睽睽[zhòng mù kuī kuī](a crowd of people staring)	Asymmetric/2	Extend/2	Predicativity/3		睽睽[kuī kuī]/2	[zài...zhī xià]/2	11
相敬如宾[xiāng jìng rú bīn](to be always courteous to each other like "guests")	Asymmetric/2	Extend/2	Predicativity/3	Appreciatory/2	宾[bīn]:guests/2	(of husband and wife)/2	13
滥竽充数[làn yú chōng shù](hold a post without qualifications)	Asymmetric/2	Metaphor/3	Predicativity/3	Derogatory/Cultural/5	滥(烂)[làn]/竿(竿)/4		17
不远万里[bù yuǎn wàn lǐ](make light of travelling from afar)	Asymmetric/2	Original/1	Predicativity/3		远:Adjective is used as verb/3		9
成千上万[chéng qiān shàng wàn](tens of thousands of)	Symmetric/1	Original/1	Attributive/2				4

We apply the study results into advanced stage Chinese L2 teaching, which shows that errors from Chinese L2 learners decreased dramatically, according to statistics, 53 students all used idioms and all use correctly in around 800-words essay. Second, the ability of learners' memorizing idioms improved obviously, most learners can remember more than 75%, the best learner reached 94%. Three is that the initiative and confidence of using idioms has been enhanced in the written Chinese.

5. Conclusions

The study conducted above focusing on assessing the difficulty level of idioms is always an issue deserved to study in TCSL. The approach proposed above is not able to get the most accurate result, this paper is only a trial study with small sample sizes on assessing difficulty levels of Chinese idioms and as the results show, the difficulty levels assessing based on teaching materials has a good effect. It's important to note that the inadequacies of the study are mainly limitations regarding the research tools and sampling, limiting us from studying the subject further.

Acknowledgement

This work was supported by the National Language Committee of China (Grant No. YB125-170).

References

- [1] Hong Bo. Idiom Teaching in Teaching Chinese as a Second Language [J]. SUN YATSEN UNIVERSITY FORUM, 2003, 23(2):297-300.
- [2] Pan Xianjun, Discussing Idiom Teaching in Teaching Chinese as a Second Language [J]. Chinese Character Culture, 2006(1):54-57.
- [3] Zhang Wenyi. On the Acquisition and Teaching of Four-character Idioms of Foreigners at mid and Advanced Chinese Level, [D]. Jinan University, 2006.
- [4] Shi Lin. Research on Errors in Chinese Idioms Acquisition Made by Korean Students and Teaching Tactics [J]. Journal Beijing University of Chemical Technology (Social Sciences Edition), 2013(2):80-84.
- [5] Zhang Yongfang. How Do Foreign Students Use Chinese Idioms: An error analysis [J]. Applied Linguistics, 1999(3):25-30.
- [6] Li Yanling. A Study of Grammatical Function of Common Chinese Idioms [D]. Beijing Language and Culture University, 2006.
- [7] Wang Ruojiang. Foreign students Chinese Idiom Errors Cause Analysis: from the View of Lexicography [J]. Chinese Language Teaching and Research, 2001(3):28-35.
- [8] Zhang Shuai. The Study of Internal Chinese Idiom Dictionary and New Exploration of External Inspection Lexicography [D]. Hebei University, 2009.
- [9] Chinese Idiom Errors from Foreign students and the Compilation Multi-function Idiom Dictionary [J]. Lexicographical Studies, 2011(1):101-109.
- [10] Mu Chunling Xu Shenghuan. Meaning and idioms [J]. Foreign language and foreign language teaching, 1998 (3) : 54-55.
- [11] Shi lin. Idioms Use Errors Based on the Analysis of Inter-language Corpus [J]. Journal of Social Scientists, 2008 (2): 158-160.
- [12] Chen Jie. Conceptual Integration Mechanism Research of Chinese Structure Symmetry Four-character Idiom [J]. Journal of Guangxi social science, 2010 (10): 132-135.
- [13] Huang Xiting, Chen Chuanfeng, Yu Hua. The Symmetry Structure Chinese Idiom Recognition Study [J]. Psychological Science, 1999 (3):193-196.
- [14] Chen Chuanfeng, Huang Xiting, Yu Hua. Morpheme Structure Symmetry Effect: Chinese Idiom for the Further Study of the Cognitive Characteristics of the Structure Symmetry [J]. Psychological Science, 2000, 23(3):10-13+125.
- [15] Xia Xiuwen. Analysis of the Chinese Idioms Prototype Syntax Function Based on Corpus [J]. Chinese language teaching and research, 2009 (3): 33-39.