

THE L2 ACQUISITION OF CHINESE CLASSIFIERS:
COMPREHENSION AND PRODUCTION

By

Jie Liu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies—Doctor of Philosophy

2018

ABSTRACT

THE L2 ACQUISITION OF CHINESE CLASSIFIERS: COMPREHENSION AND PRODUCTION

By

Jie Liu

There is a long-standing discussion on whether new functional categories (e.g. inflection, complementizer, determiner) and their features (e.g. gender, tense, number) are acquirable by L2 learners, and what the source of non-nativelike L2 performance in production and online comprehension is when morphological marking is involved. The current study employed an elicited production task, a self-paced reading task, a lexical decision task, a classifier knowledge test, and a proficiency test, to investigate the process by which English-speaking learners of Chinese acquire a new functional category, Mandarin classifiers, with a focus on the source of the challenges L2 learners face in the process.

Thirty-four English-speaking learners of Chinese and 33 native speakers of Chinese participated in the study. Main findings include: (1) In production, compared to native speakers of Chinese, L2 learners over relied on the general classifier *ge*, and used less specific classifiers; (2) L2ers were not sensitive to classifier omission in online comprehension, but they showed sensitivity to inconsistent classifiers that conflicted with the semantic features of the nouns; (3) L2 learners' lexical knowledge and their lexical retrieval ability play a crucial role in their performance in classifier production and online comprehension.

The results suggest that establishing the new functional category, classifiers, in L2 syntax is not unattainable for English-speaking learners of Chinese. The real constraint

may lie at the lexical level. Sufficient lexical knowledge can contribute to native-like performance regarding Chinese classifiers. L2 learners of Chinese differ from native speakers in ability to access co-occurring information between classifiers and nouns in mental lexicon organization with regard to classifiers. The role of L2 proficiency in classifier acquisition is also discussed.

Copyright by
JIE LIU
2018

This dissertation is dedicated to my great-grand farther and my parents.

ACKNOWLEDGMENTS

This is a long journey, and is by no means an easy one. However, when I look back now, all the struggling, self-doubt, and uncertainty seems to be disappeared, leaving me with unlimited thankfulness to my guiders and companions along this way.

My deepest gratitude goes to Dr. Patti Spinner, who is a model of the type of scholar I long to become. Without her patience and support, I would never be able to make it. Patti, you are my mentor in academics, and far beyond.

I would also like to express my appreciation to my committee members: Dr. Aline Godfroid, Dr. Xiaoshi Li, Dr. Charlene Polio, and Dr. Bill VanPatten. Thank you for your precious comments and feedbacks, which provide insights not only on this work, but also on my future direction. From you I learn how to become a qualified researcher.

I am truly grateful to my husband, Mingzhe Zheng, who has accompanied me since long time ago. Maybe there will be more challenges awaiting us in the future, but nothing is unconquerable as long as we stand by each other.

I am forever indebted to my mother, father, and brother, who always believe in me and support me without any doubt. I become fearless with you being my unconditional backing.

Also many thanks to my cohorts, thank you for all the support and encouragement during this challenging journey.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	xii
CHAPTER 1	1
INTRODUCTION	1
Overview	1
The Mandarin Chinese classifier system	3
Introduction	3
Types of classifiers	4
Syntactic properties of classifiers	5
Semantics of classifiers	6
Classifiers versus grammatical gender	9
Source of morphosyntactic variability	13
The syntactic account	13
The lexical account: production and comprehension	14
The syntactic account versus the lexical account	17
Predictive processing versus error detection	20
Chinese classifier acquisition	22
L1 acquisition of Chinese classifiers	22
L2 acquisition of Chinese classifiers	25
The current study: research questions and predictions	34
CHAPTER 2	39
METHODOLOGY	39
Overview	39
Participants	39
Proficiency test	40
Procedure and materials	40
Lexical decision task	41
Procedure and materials	41
Elicited production task	42
Procedure and materials	42
Data coding	45
Self-paced reading task	45
Procedure and materials	45
Offline cloze task	47
Procedure and materials	47
CHAPTER 3	48
RESULTS	48
Proficiency test	48

Offline cloze task	49
Lexical decision task	53
Elicited production task	54
Online comprehension task	66
Data trimming and analysis	66
General results	69
Descriptive statistics	69
Statistical analyses	71
L2 proficiency and online comprehension of classifiers	78
Descriptive statistics	78
Statistical analyses	81
Knowledge of classifiers and online comprehension of classifiers	86
Descriptive statistics	86
Statistical analyses	89
Lexical retrieval and online comprehension of classifiers	98
Summary of results	108
CHAPTER 4	111
DISCUSSION AND CONCLUSION	111
Summary of results	111
Classifier production and online comprehension	111
The role of L2 proficiency, knowledge of classifiers, and lexical retrieval	112
Answers to research questions	113
Research Question 1: Syntactic acquisition vs. lexical acquisition	113
Further remarks on the lexical account: Gradual L1/L2 difference	122
Research Question 2: Noun categorization in mental lexicon	123
Research Question 3: L2 proficiency and classifier acquisition	125
Pedagogical implications	126
Concluding remarks	128
APPENDICES	131
APPENDIX A Background questionnaires	132
APPENDIX B Proficiency test	137
APPENDIX C Stimuli for lexical decision task	139
APPENDIX D Pictures used in the elicited production task	141
APPENDIX E Classifier-noun pairs in self-paced reading task	143
APPENDIX F Stimuli for self-paced reading task	146
APPENDIX G Stimuli for offline cloze task	146
REFERENCES	153

LIST OF TABLES

Table 1.1 Examples of Mandarin classifiers and nouns	7
Table 1.2 Predictions for the first research question	35
Table 2.1 Classifiers and nouns used in the elicited production task	44
Table 3.1 Mean scores on different parts of the proficiency test	48
Table 3.2 Accuracy of the overall learner group and two subgroups in the proficiency test	49
Table 3.3 Learners' mean accuracy of each classifier-noun pair	51
Table 3.4 Accuracy of the overall learner group and two subgroups in the offline cloze task on classifiers and their accuracy on the proficiency test	52
Table 3.5 Native speakers and learners' use of classifiers for each noun	58
Table 3.6 Mean reading times for the native group	70
Table 3.7 Mean reading times for the learner group	71
Table 3.8 AIC and BIC indices of model fit	72
Table 3.9 Model results of the native group on the critical region	73
Table 3.10 Model results of the native group on the first spill-over region	74
Table 3.11 Model results of the native group on the second spill-over region	75
Table 3.12 Model results of the learner group on the critical region	76
Table 3.13 Model results of the learner group on the first spill-over region	76
Table 3.14 Model results of the learner group on the second spill-over region	76
Table 3.15 Mean reading times for the higher-proficiency learner group	80
Table 3.16 Mean reading times for the lower-proficiency learner group	80
Table 3.17 Model results on the critical region with proficiency levels	82
Table 3.18 Full model results on the critical region with proficiency level	83

Table 3.19 Model results on first spill-over region with proficiency levels	84
Table 3.20 Model results on the second spill-over region with proficiency levels	85
Table 3.21 Mean reading times for the higher-performance learner group	88
Table 3.22 Mean reading times for the lower-performance learner group	89
Table 3.23 Model results on the critical region with cloze test levels	90
Table 3.24 Full model results on the critical region with cloze test levels	91
Table 3.25 Model results of the classifier-consistent group on the critical region	93
Table 3.26 Model results of the classifier-consistent group on the first spill-over region	94
Table 3.27 Model results of the classifier-consistent group on the second spill-over region	94
Table 3.28 Model results of the classifier-inconsistent group on the critical region	96
Table 3.29 Model results of the classifier-inconsistent group on the first spill-over region	96
Table 3.30 Model results of the classifier-inconsistent group on the second spill-over region	96
Table 3.31 Simplified model results of natives on the critical region with lexical retrieval time	99
Table 3.32 Full model results of natives on the critical region with lexical retrieval time	99
Table 3.33 Simplified model results of natives on the first spill-over region with lexical retrieval time	102
Table 3.34 Simplified model results of natives on the second spill-over region with lexical retrieval time	102
Table 3.35 Simplified model results of learners on the critical region with lexical retrieval time	103
Table 3.36 Simplified model results of learners on the first spill-over region with lexical retrieval time	104

Table 3.37 Simplified model results of learners on the second spill-over region with lexical retrieval time 105

Table 3.38 Full model results of learners on the second spillover region with lexical retrieval time 105

LIST OF FIGURES

Figure 3.1. Percentage of each type of classifier use by all participants in the production task	55
Figure 3.2. Percentage of each type of classifier use by learners in the production task	62
Figure 3.3. Relationship between the use of the general classifier and proficiency score	64
Figure 3.4. Relationship between the use of congruent specific classifiers and proficiency score	64
Figure 3.5. Histogram of learners' reading times in the critical region	68
Figure 3.6. Reading time for each region: the native group	69
Figure 3.7. Reading time for each region: the learner group	70
Figure 3.8. Reading for each region: the higher-proficiency learner group	79
Figure 3.9. Reading for each region: the lower-proficiency learner group	79
Figure 3.10. Reading for each region: the higher-performance learner group	87
Figure 3.11. Reading for each region: the lower-performance learner group	87
Figure 3.12. Native reading time over lexical retrieval time on the critical region	100
Figure 3.13. Learner reading time over lexical retrieval time on the second spill-over region	107

CHAPTER 1

INTRODUCTION

Overview

A number of studies on adult second language (L2) acquisition have demonstrated the existence of variability in production and non-nativeness in processing when morphological marking (e.g. gender, tense, number) is involved (e.g. Franceschina, 2001; Prévost & White, 2000; VanPatten, Keating, & Leiser, 2012; White, Valenzuela, Kozłowska-Macgregor, & Leung, 2004). However, no consensus has been reached regarding the source of such variability.

There are several possibilities. According to the syntactic account, the problem is centered on new functional categories and features, which may not be acquirable by L2 learners after the critical period if the first language lacks them because adult learners are not able to instantiate the new categories or features in their grammar and thus they have difficulty in both production and comprehension of the new structure (e.g. the representational deficit hypothesis of Hawkins & Chan, 1997; Hawkins & Franceschina, 2004). The missing surface inflection hypothesis (MSIH, Prévost & White, 2000), however, attributes persistent L2 difficulty to, rather than problems in the syntactic level, lexical retrieval difficulties in production. Importantly, this hypothesis predicts that new functional categories and features are ultimately acquirable, and the morphological issue is due to morphological failure under processing pressure. The lexical learning account extends MSIH to comprehension, and attributes the difficulty more specifically to weak lexical representation due to inadequate learning of individual items (the lexical gender learning hypothesis of Grüter, Lew-Williams, & Fernald, 2012; Hopp, 2013). The

discussion is drawing growing attention because it helps us to understand the nature of learners' difficulties, and ultimately possibly how to address them through pedagogical interventions. Along this line, the current study aims to further explore the source of L2 difficulty by investigating L2 online comprehension and production of a new functional category for English-speaking learners, the Chinese classifier.

A classifier is a morpheme that “denotes some salient perceived or imputed characteristic of the entity to which the associated noun refers” (Allan 1977, p. 285). In Chinese, a classifier is obligatory between a noun and a numeral (e.g. *yi* ‘one’, *san* ‘three’), or between a noun and a demonstrative determiner (e.g. *zhe* ‘this’, *na* ‘that’), or some quantifiers¹ (e.g. *mei* ‘every’). Classifiers are different from gender systems primarily because no agreement is involved in classifiers, while agreement is a prerequisite for gender (Corbett, 1991; Kramer, 2015). It is relevant in the current study that classifiers are separate forms from the noun; therefore, they are subject to omission if the syntactic category of classifier is not instantiated in the grammar.

Despite the difference, classifiers are similar to grammatical gender in that they both involve the categorization of nouns. Unlike gender, classifiers can serve as head of an independent functional projection in generative analyses², but they are found to be challenging to adult learners, particularly those whose first language lacks them. Despite their similarities with the extensively-investigated feature, grammatical gender, they remain highly understudied in both L1 and L2 acquisition despite their unique properties that could shed light on current discussion in acquisition of new functional categories and

¹ Not all quantifiers occur with classifiers, for example, *yixie* ‘some’.

² It is generally thought that gender does not head an independent functional projection, instead “gender is a feature realized on one of the existing syntactic heads of the noun phrase” (Ritter, 1993, p. 795).

features. In particular, some classifiers have rich semantic information; there is a much larger number of classifiers than genders; they are separate morphemes that do not fuse with nouns and thus are subject to omission. The properties of classifiers are introduced in detail in the following section.

The current study employs an elicited production task and a self-paced reading task, a lexical decision task targeting lexical retrieval, as well as an offline cloze test for classifiers, to investigate the source of difficulty for classifier acquisition.

The first section of this chapter introduces the major syntactic and semantic properties of Chinese classifiers, and compares classifiers with grammatical gender. The second section reviews previous studies on sources of morphosyntactic variability, which mainly focused on grammatical gender. The third section introduces main findings on Chinese classifier acquisition in previous studies. The last section introduces the research questions and predictions of the current study.

The Mandarin Chinese classifier system

Introduction

Chinese is a classifier language. Many Asian languages have classifiers, including Japanese, Korean, Thai, Vietnamese, and more. Classifier languages belong to language families such as the Malayo-Polynesian, the Austro-Asiatic, the Sino-Tibetan, the Altaic, the Dravidian and the Indo-Aryan families (Senft, 2000). The number of classifiers in each language varies from two to 500 (Dixon, 1982). In Mandarin Chinese, depending on how they are counted, there are between 75 (Erbaugh, 2004) or several hundred classifiers.

Types of classifiers

There are several types of classifiers, among which the numeral classifier, which is obligatory in numeral phrases and demonstrative phrases, is the most commonly recognized type (Allan, 1977)³. They are called numeral classifiers mainly because they are required in phrases expressing quantity, although they also appear after demonstratives and some quantifiers. Numeral classifiers can be further divided into sortal classifiers and mensural classifiers (Croft, 1994; Gebhardt, 2011; Lyons, 1977). In some studies, only sortal classifiers are treated as classifiers, while mensural classifiers are called measure words (Cheng & Sybesma, 1999; Tai & Wang, 1990). The major difference between sortal classifiers and mensural classifiers, or classifiers and measure words, is that sortal classifiers individualize associated nouns and classify them based on inherent properties of nouns, while mensural classifiers, or measure words denote the temporary state of nouns (e.g. *a box of, a pile of, a cup of, a bottle of*) (Tai, & Wang, 1990). While classifiers can only be used with a limited set of nouns which share inherent properties, measure words can be used with various unrelated nouns. (1) shows an example from Mandarin.

(1) Sortal classifier: yi ben shu (* yi shu)

one Cl book

‘a book’

Mensural classifier: yi xiang shu

one Cl-box book

³ Classifiers fall into four categories: numeral classifier; concordial classifier; predicate classifier; and intra-locative classifier. See Allan (1977).

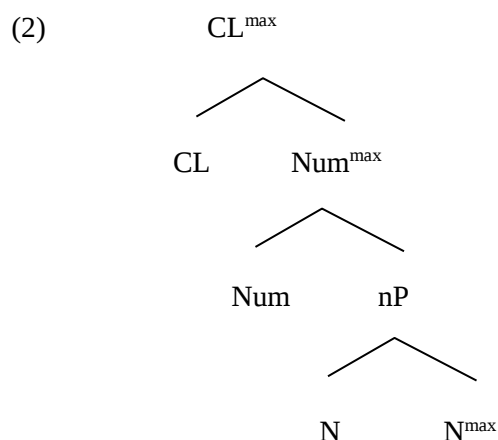
‘a box of books’

In the above example, the classifier *ben* is used for volumes; it can also apply to magazines, dictionaries, and more nouns denoting objects of this category with this sharing property. However, the measure word *xiang* (‘box’), can be used with any objects that can be put in a box, such as fruit, sand, and more. These objects do not need to belong to a specific category. Classifiers categorize nouns based on inherent properties of objects, such as animacy, shape, function, rigidity and orientation (Croft, 1994), while measure words do not function in this way. English has measure words, but not classifiers.

It is also argued that the nouns with which classifiers are used are count nouns; that is, classifiers designate the natural units of the objects (Cheng & Sybesma, 1999). Mass nouns, for instance, *water*, do not have such a natural unit; the temporary unit can be designated by measure words (e.g., *a bottle of water*, *a glass of water*, etc.). While in non-classifier languages such as English, the count-mass distinction is syntactically marked by the morphosyntactic marker of number, in Chinese, it is the classifier that marks the distinction. What is more relevant to the current study is the syntactic difference between classifiers and measure words. The syntactic properties are introduced in the following section.

Syntactic properties of classifiers

Gebhardt (2011) argued that only sortal classifiers are functional; that is, they can serve as head of an independent functional projection, as shown in (2) (Gebhardt, 2009, p. 18).



Li (2013) also showed the presence of an independent functional projection of classifier phrases, in which the classifier serves as the head, within a DP in Chinese. Therefore, the inner structure of a Chinese DP is [DP D [NumP Num [CLP CL [NP N]]]] under this analysis.

As mentioned, mensural classifiers, or measure words, are not included in the classifier system, as the term of *measure word* is used to contrast with classifier. In this study, the term *classifier* refers to sortal classifiers only. This study focuses on the acquisition of the new functional category for English-speaking learners; thus mensural classifiers are not included in the current study.

Semantics of classifiers

It has been suggested that nouns are classified by a limited set of semantic information across languages, among which are animacy, shape, function, and size (Allan, 1977; Croft, 1994). Animacy, shape, and function are among the most prominent semantic categories in the Chinese classifier system (Ken & Harrison, 1986). For instance, in Mandarin, there are classifiers for animate nouns, such as *zhi* (只) for small animals (e.g. *cat*, *dog*), and *tou* for larger animals (e.g. *elephant*, *cow*); classifiers denoting shapes, such as *zhang* for objects with a flat surface (e.g. *table*, *credit card*), and

tiao for long objects (e.g. *fish*, *pants*); and classifiers for different object functions, such as *liang* for vehicles (e.g. *car*, *bike*); and *ben* for volumes (e.g. *dictionary*, *textbook*).

More classifiers from these three semantic domains are listed in Table 1.1 below.

According to Craig (1986), the different types of semantic information that are used in classifiers form an implicational scale. In classifier languages, humanness and animacy are marked first, then shape, and then use or function.

Table 1.1

Examples of Mandarin classifiers and nouns

Semantic domain	Classifier	Noun
Animacy	<i>zhi</i> (small animals)	<i>mao</i> ('cat'), <i>xiaoniao</i> ('bird'), <i>ji</i> ('chicken')
Shape	<i>tiao</i> (long and slender objects)	<i>kuzi</i> ('pants'), <i>chuan</i> ('boat'), <i>xiaolu</i> ('road')
	<i>zhang</i> (flat surfaced objects)	<i>zhuozi</i> ('table'), <i>zhaopian</i> ('photo'), <i>zhi</i> ('paper'),
Function	<i>jian</i> (clothes)	<i>chenshan</i> ('shirt'), <i>maoyi</i> ('sweater'), <i>waitao</i> ('overcoat')
	<i>liang</i> (vehicles)	<i>chuzuche</i> ('taxi'), <i>zixingche</i> ('bicycle'), <i>motuoche</i> ('motorcycle')
	<i>ben</i>	<i>zidian</i> ('dictionary'), <i>keben</i> ('textbook'),
	(bound items)	<i>shu</i> ('book')

Table 1.1 (cont'd)

<i>tai</i> (machine)	<i>diannao</i> ('computer'), <i>dianshi</i> ('television'), <i>bingxiang</i> ('refrigerator')
----------------------	--

While semantic information plays an important role in deciding classifier membership, the information is not totally transparent. Some classifiers have more clear defining features compared to others (Gao, 1998). For instance, *liang* is for vehicles only; its members include *taxi*, *bus*, *truck*, *bicycle*, and so on. However, for many classifiers, which nouns they are associated with is not fully predictable without extensive knowledge of the language (Allan, 1977). For example, in Mandarin, the classifier *jian* can apply to clothes; however, *pants*, which also falls into the semantic category of clothes, takes another classifier, *tiao*, due to its shape, not its function. Similarly, although *tiao* classifies long and slender objects, objects such as *chopstick* and *pen* are denoted by another classifier, *zhi* (支), because of their cylindric shape (Srinivasan, 2010; Tai & Wang, 1990). There is no simple rule available to summarize which inherent property determines classifier membership.

It has been argued that classifiers other than the general classifier are learned through analogy, even for native speakers. For instance, after learning that *jian* can apply to *shirt*, speakers will apply it to *sweater* due to the similarity of these two objects, which is an example of correct analogy; however, when they extend *jian* to *pants*, it is not correct (for review, see Myers, 2000).

People organize concepts based on taxonomic relations (e.g., animals, such as cat, dog, sheep), but the semantic basis based on which classifiers categorize referents is not

fully consistent with conceptual categorizations (Gao, 1999; Saalbach & Imai, 2007). There is an argument that classifiers can be divided into two types, taxonomic-specific classifiers and shape classifiers. The former one categorizes referents based on taxonomy, while the latter one relies on shape (e.g. Downing, 1996; Sumiya, 2008). An example for taxonomy classifier is *liang*, which is used for vehicles including *car*, *taxi*, *bike*; *zhang*, which is used for flat-surfaced objects including *paper*, *credit card*, *bed*, is an example for shape classifiers.

Classifiers versus grammatical gender

Classifier languages such as Chinese categorize nouns into a number of classifier classes, while gender languages categorize nouns into different genders, for instance, in Spanish, nouns are either feminine or masculine. Therefore, in languages with grammatical gender, nouns are assigned to a gender class; gender features are associated with each noun. In many cases the lexical gender is transparent. For example, the majority of Spanish nouns ending with *-o* are masculine and ending with *-a* are feminine (Teschner & Russel, 1984). However, gender is not always transparent on nouns; there are also nouns without transparent markers. What can serve as another, less fallible cue to the gender assignment of each noun is the gender agreement marking on other elements that modify the noun, such as determiners, adjectives, and other elements. In generative analyses, the gender feature on the agreeing item is checked with the gender feature on the noun to make sure they agree (Carstens, 2000).

Lexical knowledge is involved in gender assignment while syntactic knowledge is involved in gender agreement. Different from gender, classifiers do not have agreement. Classifiers do not fuse with nouns but rather exist as separate morphemes. A lack of the

projection in syntax may lead to ungrammatical omission of classifiers; incorrect use of classifiers can be argued as equivalent to the assignment issue in gender because both suggest a wrong categorization of nouns.

In addition to the lack of agreement, classifiers also differ from gender in various ways, as was discussed in Dixon (1982). There are limited number of genders in a specific language, but the number of classifiers in a language is much larger; thus each classifier is associated with relatively fewer nouns. In addition, the selection of classifier is flexible to some extent, in that some nouns can be used with multiple classifiers; for instance, for the noun *qiche* ('car'), the vehicle classifier *liang* is usually used, but some speakers also tend to use the machine classifier *tai* with it. In the example above, there is no difference between the meaning of these two phrases. However, sometimes the choice of classifier differentiates between senses of a noun. For instance, for *ke* ('course/class'), when the classifier *men* is used, it means one course; when another classifier *jie* is used, it means one class session (Zhang, 2007).

Finally, there is often a general classifier that can be used for a variety of nouns. *Ge* is the general classifier in Chinese. Other individual classifiers are called specific classifiers (Li & Thompson, 1981). Specific classifiers are less frequent than the general classifier. Erbaugh (1986) found that only 22 specific classifiers were used in an 877-utterance sample of adult-adult conversations.⁴ Each specific classifier can be used with 5-20 nouns in Mandarin, and perhaps 40% of nouns can only take the general classifier (Erbaugh, 2004). The general classifier *ge* can be used with human beings (e.g. child, thief), large three-dimensional objects (e.g. watermelon, sun), abstractions (e.g. hope),

⁴ Among the seven target classifiers in the current study, only the machine classifier *tai* is not in this list.

and other nouns that do not require a specific classifier. It is the default classifier in that speakers use it when a specific classifier is not available for a noun (either the specific classifier does not exist or the speaker does not know it). Being the default form, the usage of the general classifier is complicated; even native speakers may replace almost every specific classifier with *ge* (Loke, 1996; Zeng & Hong, 2012). Both semantic and discourse factors influence the choice between the general classifier and specific classifiers. It is argued that speakers tend to replace a specific classifier with the general classifier for less prototypical members (e.g. *zhang* for *paper*, but *ge/zhang* for *sofa*); additionally, function-based classifiers are more likely to be replaced by the general classifier compared to shape and animacy-based classifiers (see Myers, 2000). The most frequent occurrence of specific classifiers is for first mention of new objects (Erbaugh, 2004). In summary, although the general classifier is more frequent, native speakers of Chinese tend to use specific classifiers for some nouns and in specific contexts. In addition, classifiers can serve to distinguish meanings in some cases (Zhang, 2007), therefore, the use of specific classifiers is not totally redundant despite the frequent use of the general classifier.

There is limited research on the frequency of specific classifiers in L2 language learners' input. In Erbaugh (1986), the two adult native speakers of Chinese who used specific classifiers the most in story-telling and conversation, were both Mandarin teachers. It provides evidence, although indirectly, that L2 learners of Chinese possibly are able to be exposed to specific classifiers in their Chinese language classes.

Classifiers are a new functional category for learners whose L1 lacks them. Therefore, the investigation of classifiers can contribute to the ongoing discussion of why

new functional categories/features are challenging to L2 learners. There is a possibility that the difficulty in the acquisition is due to lack of new features or functional categories in syntax, in which case persistent difficulty will be expected even with near-native L2 learners (FFFH) (Hawkins & Chan, 1997). Another possibility is that syntactically L2 learners can acquire the new features or functional categories even they are not instantiated in their L1, but learners have difficulty mapping those features with syntactic nodes and pronouncing the correct morphological forms in production because of the processing pressure (MSIH) (Prévost & White, 2000). It is also possible that difficulty in learning individual items (for instance, the classifier class for each noun) rather than an overall mapping issue is the source for the difficulty observed in L2 acquisition of classifiers, in which case lexical learning lies at the root of the difficulty and learners may be able to show improvement with adequate vocabulary learning.

The special characteristics of classifiers, including the way semantic information is involved, as well as the way classifiers and nouns are associated, may shed light on whether or how a new functional category or feature can be acquired. As noted, as separate morphemes classifiers would be subject to omission if the L2 grammar lacks the syntactic projection. In addition, the use of classifiers is similar to grammatical gender, and it is possible to examine whether lexical information plays a role in establishing connections between the syntactic node and lexical items. Therefore, the investigation of L2 acquisition of classifiers provides another way to tease apart the syntactic account and the lexical account of the source of difficulty in acquisition of new features or functional categories. The syntactic account and the lexical account are further discussed in the following section.

Source of morphosyntactic variability

The syntactic account

It has been widely accepted that variability in morphological marking exists even for advanced adult L2 learners. Much of the evidence is from the L2 acquisition of grammatical gender. For instance, Franceschina (2001) reported a case study in which the participant, Martin, was an English-speaking learner of Spanish who had lived in a Spanish-speaking environment for twenty-four years. In spontaneous conversation with the researcher, Martin still demonstrated variability with gender agreement between nouns and determiners/adjectives, overusing masculine markers on determiners and adjectives in feminine or neuter contexts. The results were interpreted as evidence for the representational deficit hypothesis (RDH), which argues that the source of the errors is representational and will exist permanently if the functional categories/features are not instantiated in the L1.

However, the interpretation of the variability was questionable from the perspective of Prévost and White (2000), who proposed the missing surface inflection hypothesis (MSIH), arguing that L2 deficits do not lie in representation; new features are represented at abstract level in L2, but learners have difficulty retrieving the lexical forms with which the features are expressed. Therefore, difficulty in production in particular can be anticipated, and learners tend to rely on default forms because of failures in lexical form retrieval under time pressure.

To tease apart representational deficits and retrieving difficulty, McCarthy (2008) investigated whether morphological variability in production extends to comprehension. It was found that English-speaking learners of Spanish showed qualitative similarity of

morphological variability across comprehension and production in gender and tense agreement. In both an elicited production task and a picture identification task in which communication pressure was not present, participants adopted masculine and singular defaults in gender and number agreement respectively. These results were argued to provide evidence for the representational deficit view.

More recently, several studies that have focused on online comprehension of gender agreement reported that L2 learners across L1s are able to show native-like sensitivity to violations in gender agreement in online comprehension tasks (Foote, 2011; Foucart & Frenck-Mestre, 2012; Sabourin & Stowe, 2008), providing further evidence against a syntactic deficit. At the same time sensitivity might be constrained by syntactic distance in that learners are only sensitive to local agreement violations (Keating, 2009), suggesting that processing difficulty plays a role in online L2 comprehension. These findings indicate that the difficulty may not lie at the representational level.

The lexical account: production and comprehension

As the MSIH argues, the root of L2 inflectional variability lies in lexical retrieval (the “mapping” issue), causing difficulty to produce consistent inflection in production. To further explore whether L2 difficulty is limited to the mapping issue in production, or it is a real-time processing issue spanning production and comprehension, Grüter, Lew-Williams and Fernald (2012) incorporated production, an offline comprehension task, and an online comprehension task examining predictive processing: specifically, the ability to predict upcoming nouns based on gender-marked articles. It was found that in the offline comprehension task, advanced English-speaking learners of Spanish performed at ceiling, indicating that advanced L2 learners may be able to establish abstract gender categories.

However, the learner group showed weakness with gender assignment in production. As discussed, lexical knowledge is involved in gender assignment. The asymmetry between gender agreement and assignment shown by learners in the production task suggested that lexical, rather than the syntactic aspect of gender underlies the persistent difficulty. In the predictive processing task, the learner group demonstrated limited ability to use the gender marking on determiners to predict the upcoming familiar nouns. The non-nativelike performance in the online comprehension task cannot be fully explained by the MSIH. As a result, Grüter et al. (2012) extended the MSIH from the domain of production to comprehension and proposed the lexical gender learning hypothesis. They argued that native speakers rely on computation of co-occurrence of determiners and nouns, which leads to tight associations between them. That is, in the development of L1 grammar, infants rely on the computation of co-occurrence of determiners and nouns to detect the gender of each noun, possibly because of the lack of sufficient phonological and semantic information on the noun to determine the gender category. Consequently, tight associations are formed. This hypothesis is supported by the fact that infants occasionally treat the determiner and noun as a whole chunk. L2 learners, on the other hand, use other cues including metalinguistic information and written form information, and consequently the strength of association between nouns and gender nodes are weaker in the L2 compared to in the L1.

Interestingly, however, the researchers found that L2 speakers were able to engage in predictive processing with novel nouns that they had been briefly trained on during the study. The authors argued that the learning process of the novel nouns was similar to the process of L1 learning, and co-occurrence relationship was the only cue available to

learner participants to learn the gender of each noun. This fact enabled the learners to engage in predictive processing. This idea places the source of morphosyntactic variability in the context of language learning.

Hopp (2013) also provided evidence for the lexical gender learning hypothesis. He examined how English-speaking learners of German performed on gender assignment in production and whether gender agreement can facilitate their online comprehension in a visual world eye tracking experiment targeting predictive processing. The L2 group showed variable performance on gender assignment in production. In the online comprehension task, only participants who showed consistent target-like gender assignment in production showed overall effects of predictive gender processing in all the three genders. The asymmetry indicated that strong lexical gender representation is essential for online predictive processing. The author argued that the weak link between lexical and abstract gender nodes in the L2 because of less frequent input and use, together with limited processing resources, underlies learners' difficulty in using gender as an informative cue in online predictive processing.

Hopp (2016a) further explored the relationship between lexical representation of gender and predictive agreement processing. Using the same visual-world eye-tracking paradigm as Hopp (2013), he found that intermediate English-speaking learners who received training on gender assignment on nouns showed native-like online processing of nouns. In the second experiment targeting native speakers of German, it was found that native speakers who received non-target information on gender assignment showed no predictive processing of gender. The results showed that while targetlike gender representation in the L2 lexicon facilitates predictive processing, non-targetlike

representation inhibits predictive processing, even for native speakers, because wrong predictions can be costly in online processing. Strong lexical representations of gender are a prerequisite for target-like processing of gender agreement in online comprehension.

The syntactic account versus the lexical account

Comparing different theoretical approaches with regard to the source of learners' difficulty with grammatical gender, the RDH argues that learners can never reach nativelikeness because they cannot acquire new categories and features. On the other hand, the MSIH argues that representation is intact, but production is problematic because learners have difficulty retrieving the specific lexical form under communication pressure. Both of these theories focused to a large extent on production. However, with increasing attention on L2 processing and the employment of processing methodologies, processing accounts have also taken hold, which suggests that to some extent, slower or less efficient processing is part of the problem. More recently, several researchers looking at predictive processing propose the lexical gender learning theory, arguing that weak links between lexical and gender nodes is the main reason why second language learners do not consistently behave like native speakers, and nativelike performance is attainable with sufficient lexical knowledge. However, it still remains unclear why learners do not show desirable learning results if lexical issues lie at the root.

If lexical learning of noun categories is at the root of learner difficulties with using classifiers and gender marking, it is likely that semantics, frequency, and other lexical information will all play a role. Indeed, the role of semantic information available in the L2 acquisition of a new functional category/feature has been investigated with regard to

gender acquisition. It was found that learners are sensitive to semantic information such as biological sex, humanness, and animacy in gender assignment and agreement (see Spinner & Thomas, 2014). Additionally, learners perform better on nouns with biological gender (Franceschina, 2005; Spinner & Juffs, 2008), indicating that semantic transparency of noun categorization facilitates the association of gender and noun. As for classifiers, it is possible that the rich semantic information encoded in classifiers can facilitate predictive processing of Chinese classifiers (Lau & Grüter, 2015), but the results are far from conclusive as to how accessible the semantic information is to learners and what kind of semantic information encoded in classifiers is more accessible.

One possibility is that, as there is an implicational scale of different types of semantic information that are used in classifiers (Craig, 1986), in that humanness and animacy are marked first, then shape, and then use, or function, L2 acquisition might follow the same scale, and learners might perform better on humanness and animacy classifiers than shape and function classifiers (see Spinner & Thomas, 2014, for a similar argument regarding grammatical gender). It is also possible that taxonomy-specific classifiers might be easier to acquire than shape classifiers, possibly because taxonomy specific classifiers are consistent with conceptual categorizations (Saalbach & Imai, 2007). Indeed, L1 speakers of Japanese acquire these classifiers first (Sumiya, 2008).

Another possibility is that semantic consistency instead of types of semantic information affects L2 acquisition of classifiers. It has been found that when L2 learners used the classifier to predict the upcoming nouns, they were distracted by objects that were consistent with the semantic feature encoded in the classifier. For instance, when learners heard *tiao*, they looked more to wrist watch, which was semantically consistent

but not grammatically consistent with the classifier, compared to the grammatically consistent target, *dog*. (Grüter, Lau, & Ling, 2018) (see more detailed discussion on this study in the following section). L2 learners' greater attention to semantic information on classifiers might suggest that it is more difficult for them to acquire classifier-noun combinations that are not semantically related in learners' perception. For instance, the long shape classifier *tiao* can be used for *dog*, usually large dogs; the small animal classifier *zhi* can be another option, usually for small dogs (although both classifiers can be used for all dogs). However, learners would use *zhi* more than *tiao* because *zhi* is semantically related to *dog*, while in learners' perception dog is not a prototype of long objects. Note that in this case, both the taxonomy classifier *zhi* and the shape classifier *tiao* are consistent with the noun *dog*, which pattern is not very frequent in the Mandarin classifier system. In addition, although in L2 learners' perception the long shape classifier *tiao* may not be semantically related to *dog*, actually no semantic conflict is involved in this combination, because the shape of dogs, especially those large ones, can be described as long.

In addition to the influence of semantic information, learners' ability to retrieve or access lexical items, and the relationship between their lexical access and online comprehension and production, could shed light on the lexical account (Hopp, 2017). According to the lexical account, gender assignment is challenging to L2 learners because learners have difficulty establishing associations between the gender feature and individual lexical items. Use of classifiers is similar to gender assignment in that both involve categorization of nouns. Therefore, presumably the ability to retrieve lexical

items would affect online comprehension and production of classifiers in a similar manner as grammatical gender assignment.

One way to measure speed of lexical retrieval is with a lexical decision task, in which test-takers decide whether the words they see are real words or not (Snellings, Van Geldern, & De Glopper, 2002). A widely-known form of such lexical task is LexTALE (Lexical Test for Advanced Learners of English) (Lemhöfer & Broersma, 2012), a test for lexical knowledge and language proficiency. In addition to the test score based on accuracy, reaction times in such a test could serve as a measurement for lexical retrieval speed. The current study adopts the LexTALE format and includes the nouns investigated in the production and comprehension task in the word list, to investigate whether participants' lexical knowledge, including their accuracy in the word decision task and their lexical retrieval speed, affects classifier production and comprehension.

Predictive processing versus error detection

As it has been discussed, several previous studies that provided evidence for the lexical learning account focused on predictive processing, that is, whether learners are able to use the pre-noun gender marking as a facilitative cue to predict the upcoming noun. Learners' sensitivity to violations provides different information than predicative processing, because of the involvement of backward checking of congruency between input and representation (Hopp, 2013). Sentence processing is incremental in the sense that the parser makes top-down expectations based on the linguistic features in previous segments before encountering bottom-up information (for review, see Hale, 2011). Backward checking or reanalysis occurs when an unexpected element is encountered (for instance, a feminine noun appears after a determiner with a masculine gender marker).

Self-paced reading tasks have been used extensively for investigations of grammatical violations. They have figured prominently in investigations of syntactic processing including establishment of filler-gap dependencies (e.g. Stowe, 1986; Traxler & Pickering, 1996), resolution of syntactic ambiguity (e.g., Hopp, 2006), and sensitivity to anomalies regarding morphological marking such as number and gender (e.g. Jiang, 2004; Jiang, Novokshanova, Masuda, & Wang, 2011; Sagarra & Herschensohn, 2011). In this task, elevated reading time can be treated as evidence for the parser's sensitivity to violations (Keating & Jegerski, 2015; Van Patten, et al., 2012). In addition, self-paced reading task has been used for different L2 proficiency groups, Marsden, Thompson, and Plonsky (2018) reviewed 68 journal articles using self-paced reading for investigation of L2 sentence processing, it was found that six studies included beginning learners of the target language, 18 studies included intermediate learners, and 69 studies included advanced, near-native or bilingual L2 group.

With manipulations of the stimuli, self-paced reading tasks can provide information on what type of information the parser can use in online processing. For instance, a number of studies looked at how adult L2 learners process sentences with violations on number or gender agreement in real time (e.g., Foote, 2011; Jiang, 2004; Sagarra, & Herschensohn, 2011). Whether learners are able to use the grammatical cues from agreement in online comprehension is reflected by whether they are sensitive to violations during the reading. In this sense, self-paced reading has something in common with the visual-world paradigm targeting predictive processing, in that both provide information on whether a specific type of information can be utilized in online comprehension. In addition, self-paced reading task allows manipulation of violations,

through which we could investigate to what type of violation the parser is sensitive or shows limited sensitivity. Consequently, it allows the investigation of what is difficult for the parser.

The current study focuses on error detection instead of predictive processing using a self-paced reading task. While predictive processing explores whether learner can use a specific type of information to predict the upcoming linguistic materials, grammatical sensitivity to errors reflects what learners do not accept, which can be achieved through manipulation of violation types. Both syntactic violations, which were manipulated by dropping the grammatically required classifier in a nominal expression, and lexical violations, which were manipulated by using an incongruent classifier with the noun, were included in target sentences, in order to track the root of difficulty learners have in classifier acquisition.

Chinese classifier acquisition

L1 acquisition of Chinese classifiers

In the field of L1 acquisition of Chinese classifiers, a large number of studies have centered on classifier development (Erbaugh, 1986; Hu, 1993; Tse, Li, & Leung, 2007), particularly the relationship between classifier systems and speakers' conceptual categorization (Bi, Yu, Geng, & Alario, 2010; Huetting, Chen, Boweman, & Majid, 2010; Saalbach & Imai, 2012).

In studies focusing on the developmental pathway of the classifier system in L1 children, general findings include:

- 1) Specific classifiers are rare in children's utterance; they appear late and develop slowly;

2) Children as young as three to four years seldom show errors of classifier omission. They tend to use the general classifier to hold the syntactic position, which is interpreted as evidence of successful acquisition of the syntactic properties of classifiers.

Erbaugh (1986) investigated young and adult L1 Chinese speakers' use of classifiers in four settings: adult-adult conversation; adult-child conversation; child-child conversation, and adult narratives, in which adults watched a speechless short video and told a story accordingly. Four findings emerged: adult speakers seldom omit classifiers; they use limited specific classifiers; the general classifier is much more frequent in their speech, even in contexts where a specific classifier is required; specific classifiers are more frequent in formal conversation. The choice of classifiers varies across individuals and across discourse contexts, which is not surprising as many nouns can be associated with multiple classifiers with or without changing the meaning. As for children, similarly to adults, they seldom drop classifiers in obligatory context. They use the same core set of specific classifiers as adults, although the frequency of the specific classifier is even lower in children's utterance compared to adults'; specific classifiers remain rare and develop slowly between age of 1.10 and 3.10. When children acquire a new classifier, they first used it with a prototype, for instance, *tiao* (for long objects) with *snake*; then they use the salient feature, in this case, shape, to generalize the application of this classifier to other referents, such as *boat* and *dragon*. The generalization in the example is acceptable, although in some cases children make false generalizations.

In studies focusing on the relationship between classifier systems and speakers' conceptual categorization (Bi, Yu, Geng, & Alario, 2010; Huetting, Chen, Bowman, & Majid, 2010; Saalbach & Imai, 2012; Zhang & Schmitt, 1998), it has been found that

classifier systems may affect the way nouns are categorized in speakers' mental lexicon, but how strong the effect is remains inconclusive. The classifier system may serve as another way to categorize objects in addition to taxonomy (Zhang & Schmitt, 1998); it is also possible that the effect of classifier on conceptual categorization is not comparable to taxonomic relationship (Huetting et al., 2010; Saalbach & Imai, 2012).

It is also worth mentioning another study on the L1 acquisition of classifiers. The target language is Japanese, which has similar classifier system as Chinese. Sumiya (2008) investigated children's (from age of three to five) performance on three taxonomy-specific classifiers and three shape classifiers. It was found that children perform better on taxonomy-specific classifiers, which indicates asymmetry in accessibility of these two types of semantic information. The comparison of learners' performance on these two types of classifiers may be able to provide insight on what type of semantic information is more accessible to learners in adult L2 acquisition of classifiers.

Semantics appear to play an important role in the detection of classifier and noun mismatches. In studies using event-related potentials (ERPs), it has been found that when nouns are used with an incongruent classifier (*zhang* (flat surfaced objects) / **tai* (machines) for *chair*), the mismatch elicited N400 on the noun, which is an indicator of semantic processing (Chou, Huang, Lee, & Lee, 2014; Zhang, Zhang, & Min, 2012; Zhou et al., 2010; Qian & Garnsey, 2016).

Chou, Lee, Hung, and Chen (2012) used functional magnetic resonance imaging (fMRI) to investigate brain activation in online comprehension of Chinese classifiers. They included two types of violation in the stimuli. One violation involved an

incongruent classifier with a noun, and the other was to use a word from other categories to take the place of the classifier (e.g. one CL- *pian* / *V-make leaf). They found greater activation in brain areas related to semantic processing when an incongruent classifier was used. When a word from another category was used, activation in brain areas related to semantic processing as well as syntactic processing was observed.

The results of the ERP studies confirmed that syntactic and semantic processing is involved in online comprehension of classifier-noun combinations, and native speakers are sensitive to both syntactic and semantic violations with regard to classifiers. It is important to investigate whether adult L2 learners of Chinese are sensitive to these violations too, which will improve our understanding of why classifiers are difficult for learners.

L2 acquisition of Chinese classifiers

Compared to L1 acquisition of Chinese classifiers, studies on L2 acquisition are more limited. Several studies have investigated how adult L2 learners use classifiers in oral or written production. Offline comprehension of classifiers has also been investigated. However, very few studies have focused on the online comprehension of Chinese classifiers.

In Polio (1994), 21 English and 21 Japanese-speaking learners of Chinese watched a short speechless video from Chafe (1980), then told the story in Chinese to a native speaker. The participants were divided into three proficiency levels based on class placement, native speakers' rating of their proficiency, and a proficiency test. Polio found that in oral production, adult L2 learners of Chinese tended to rely more on the general classifier *ge*; their use of specific classifiers was limited, and a few unacceptable

or questionable uses were observed. For instance, one participant used *tiao* for *tree*; although trees are long, they require another classifier *ke* (棵), for plants. Learners did not omit classifiers in obligatory environments, even at low proficiency levels. Ungrammatical uses of multiple classifiers were observed, as shown in (3) below. The author argued that the overuse of the general classifier relates to the fact that learners regard classifiers as bound to determiners or numerals, not to nouns. The chunking may have a negative effect on classifier use as it was observed in this study, but positive effect may also be possible in that L2 learners would not omit classifiers.

(3) Ungrammatical use of multiple classifiers

*Yizhi	kan	nage	sange	xiaohai.
Continue	look at	that-Cl	three-Cl	kid

‘(He) kept looking at the three children.’

In Gao (2010), participants were 30 Swedish-Chinese bilingual children whose ages ranged between 6 and 19, as well as 39 adult Swedish-speaking learners of Chinese. They were divided into three proficiency levels based on a proficiency test. All participants were asked to name 30 objects using numeral expressions. Adult learners took the same test three times at four-week intervals. It was found that bilingual children were more accurate than low and intermediate adult learners, while advanced learners had slightly higher accuracy than bilingual children. As for the influence of L2 proficiency, it was found that adult learners’ performance was correlated with their Chinese proficiency. Learners of all proficiency levels showed improvement on classifier production during the two-months’ study. Another interesting finding is that it seems frequent usage in daily communication facilitates classifier learning. For instance, the

accuracy for the noun *book* is 100%; all learners used the correct classifier *ben* (for bound volumes) with it. The high frequency of the noun in participants' life might have contributed to the correct use of the classifier. As for the error type, learners demonstrated incorrect use of specific classifiers, as well as classifier omission and overuse of the general classifier.

Zhang and Lu (2013) investigated the L2 development of classifiers in written production using a corpus. It was suggested by their results that learners used more general classifiers and fewer types of classifiers than native speakers; learners with higher L2 proficiency showed less classifier omission and more diversity in classifier production.

Liang (2009) investigated how native speakers of Korean and English used classifiers in an offline comprehension task. Participants were presented with pieces of clay in different shapes, and phrases with different shape classifiers, e.g., *tiao* (for long objects), *zhang* (for flat objects), and *tuan* (for rounded-shape objects). Participants chose the pieces of clay that they thought best matched each phrase, and then rated how sure they were on their choices on a 5-point scale. In this task, high-proficiency learners outperformed low-proficiency learners in general. Novice Korean-speaking learners outperformed their English-speaking counterparts, while intermediate English-speaking learners outperformed their Korean-speaking counterparts. There was also a production task in which participants looked at pictures and wrote down answers to questions 'How many XXX are there in the picture?'. In the written production task, there was a positive correlation between participants' performance and Chinese proficiency. Korean-speaking learners performed better than English-speaking learners. Both groups

performed better on animacy classifiers compared to function classifiers. From the results it can be seen that learners' performance on classifier improves with L2 proficiency, both in comprehension and written production. Another important finding is that some semantic information is more accessible to learners, making classifiers from a specific semantic domain easier to acquire than others (animacy classifiers were easier than function classifiers in this case, but shape classifiers were not included in the comparison). The nature of the production task in the study makes it possible for learners to utilize their explicit knowledge of classifiers, it is possible that semantic information is used differently in online comprehension and production with more processing pressure.

According to previous studies that have included L2 proficiency as a factor, it seems that the classifier system develops as proficiency grows, which is similar to the L2 acquisition of grammatical genders. It has also been argued that gender develops with proficiency; lower-level learners have difficulty with both agreement and assignment with gender, while upper level learners primarily have difficulty with assignment (Alarcón, 2011; Grüter et al., 2012; Kupisch, Akpınar, & Stöhr, 2013). The majority of previous studies focused on how the classifier system develops in L2 grammar in production (typically offline production) and offline comprehension. However, there has been little investigation regarding the ways in which the L2 classifier system develops with proficiency in online comprehension and production.

Although previous studies on the L2 acquisition on Chinese classifiers provided us with information what the classifier system looks like in L2 production and off-line comprehension, more work, especially investigation of online comprehension of classifiers, together with production, is needed to help us to locate the source of difficulty

in the L2 acquisition of classifiers, to determine where the difficulty lies, in lack of the new functional category in representation or lexical learning.

However, studies focusing on the L2 online comprehension of Chinese classifiers are quite limited. Lau and Grüter (2015) investigated whether English-speaking learners of Chinese could use classifiers as facilitative cues to predict upcoming nouns. Two classifiers, *tiao* (long and slender objects) and *zhang* (flat surfaced objects) were examined, and each was combined with four nouns (*boat, fish, pants* and *towel* for *tiao*; *bed, table, map* and *credit card* for *zhang*). They found that English-speaking learners showed a trend towards a facilitative effect; learners who were more proficient based on their performance on a cloze test showed a pattern more similar to the native group. In comparison with previous studies on Spanish gender, in which the facilitative effect was absent, the authors attributed the facilitative effect they observed to semantic information available in classifiers and the limited number of nouns associated with each classifier. That is, rich lexical information encoded in classifiers might facilitate acquisition, particularly in cases where an encounter with a particular classifier narrows down the possible selection of nouns to a very small number.

Similar results were found in Liu and Spinner (manuscript), in which four classifiers, *tiao, zhang, zhi* (small animals), and *jian* (clothes), each with six nouns were used as stimuli; it was found that learners with high accuracy in the multiple-choice task on classifiers showed a facilitative effect in online comprehension, which suggested that with sufficient lexical knowledge, classifiers can be acquirable.

Despite the findings that it is possible for adult L2 learners to use classifiers to predict upcoming nouns, it remains unclear what type of information was utilized by

learners in online comprehension: the syntactic information, the semantic information, or other lexical information. Grüter, Lau, and Ling (2018) tried to determine what information is available to learners by including three different competitor conditions in the visual world eye-tracking experiment: competitors from the same classifier class as the target noun, matching the semantic features of the class (for instance, when the target was *dog*, which could take the long shape classifier *tiao*, the competitor was *rope*); competitors from another classifier class as the target but semantically matched with the class (for instance, when the target was *dog*, the competitor was *wrist watch*, which has a long shape but takes a different classifier than *tiao*); competitors from another classifier class, not consistent with the semantic features of the target (e.g., when the target was *dog*, the competitor was *apple*). It was found that for native speakers of Chinese, the class-consistent competitors were more distracting than the class-inconsistent but semantics-consistent ones. As for L2 learners of Chinese, semantics-consistent competitors were equally distractive no matter whether they were from the same classifier class as the targets. When both the target and the competitor appear to be semantically inconsistent with the classifier (e.g., *dog* with the long shape classifier *tiao*, with *apple* as the competitor⁵), the classifier-class information could also serve as a facilitative cue to predict the upcoming nouns. The authors argued that L2 learners primarily rely on semantic information for online comprehension of classifier-noun combinations, because the competitors that were consistent with the semantic features of the classifier were distractive to learners, even though they were not from the classifier category. When semantic information is not informative in learners' perception, co-

⁵ Note that actually there is no semantic conflict between *dog* and *tiao*, but in learners' perception *dog* is not related to long objects.

occurrence relationship between classifiers and nouns could also be utilized. The findings indicate that semantic features of classifier were easier for learners to acquire, and meanwhile co-occurrence information between classifiers and nouns was also accessible to learners.

In the previous studies targeting online comprehension of Mandarin classifiers, the role of learners' lexical knowledge has been investigated by using an offline task to test whether participants could connect classifiers with nouns. However, a lexical test targeting lexical retrieval speed could provide further information on the quality of lexical representation (Hopp, 2017). To this end, the present study also includes a lexical decision task modeled on LexTALE, in which all the target nouns in the production and comprehension tasks were included in the stimuli list; participants' accuracy as well as reaction time in deciding whether the words they saw were real words were recorded. More information on the lexical test is introduced in the methodology section.

To summarize, classifiers are understudied in the field of L2 acquisition. The only conclusive finding seems to be that classifiers are challenging to adult learners, as demonstrated by variability in production and reduced ability to use it as an informative cue in predictive processing, at least for many learners. However, more evidence is needed to locate the source of the observed difficulty in the L2 acquisition of classifiers.

The current study will investigate the online comprehension of classifiers via error detection, which provides different information than predictive processing. Different from predictive processing, which explores whether learners can use a specific type of information to predict upcoming elements, grammatical sensitivity to errors reflects what

learners do not accept. Manipulation of violation types could help to track the root of variability in use of classifiers.

Syntactic violations with the classifier can be created by omitting the classifier between the numeral/determiner and the noun. Violations on the lexical level can be created by using an inconsistent classifier with the noun. Lexical information includes semantic features of classifier class, and also co-occurrence relationships between classifiers and nouns. The co-occurrence information is needed to know which semantic feature is picked to determine the classifier membership. For example, pants are clothes and they are long, their shape rather than function is picked by the classifier system, therefore a shape classifier rather than a taxonomy classifier is required. Semantic feature is not sufficient to determine which classifier is the correct one, whether a shape classifier or a taxonomy classifier is required. For example, *book* has a flat surface, but it takes the taxonomy classifier *ben* (for bound items) rather than a shape classifier.

A large number of classifiers are from three semantic categories or semantic domains, animacy, shape and, function. For the relationship between the semantic domain and the taxonomy-shape contrast in the classifier system, classifiers from the domains of animacy and function are taxonomy classifiers, and unsurprisingly classifiers from the semantic domain of shape are shape classifiers. In classifier-noun combinations, one kind of violation can be created when use a classifier from the same semantic domain as the correct classifier. For example, both *tiao* and *zhang* are shape classifiers, *pants* requires the long shape classifier *tiao*, and it does not match the semantic feature of the

flat shape classifier, *zhang*⁶. Another type of violation can be created by pairing a noun with an inconsistent classifier from a different semantic domain than the correct classifier (for instance, using the clothes classifier *jian* for *pants*). In this manipulation, I did not replace animacy classifiers with function classifiers or in the reverse direction, as classifiers from these two semantic domains are taxonomy classifiers. Instead, I made sure that taxonomy classifiers took the place of shape classifiers, or shape classifiers replaced taxonomy classifiers. For this type of violation, where a taxonomy classifier is used when a shape classifier is required, or vice versa, semantic information itself is not sufficient to rule out the incorrect classifier, as semantic confliction is not necessarily involved. Therefore, co-occurrence information in addition to semantic information is needed for detection of the violation.

Admittedly, whether L2 learners are exposed to certain classifiers may also affect classifier acquisition. The current study focused on specific classifiers that were introduced in the participants' textbooks in their Chinese classes, and were relatively frequently used in their classroom conversations. All the target classifiers, except for one (*tai* for machines), were also among the 22 most frequently used specific classifiers in Chinese adult-adult conversations in Erbaugh (1986).

In the current study, self-paced reading, a method that has been widely used to measure grammatical sensitivity to errors, was used. A large number of studies on Chinese linguistics have employed this method. Notice that Chinese words differ in length in that each word consists of one or more characters. Different from English and

⁶ Each semantic category/domain has a number of classifiers. For example, shape classifiers include *tiao*, *zhang*, *zhi* (支), *kuai*, and more. Usually classifiers from the same semantic domain denote different features. *Tiao* is usually used for long and soft objects, *zhi* instead is used for long, rigid, and cylindric objects.

other alphabetic languages, the Chinese script does not have space between words, hence word boundaries are not overly marked (Chen & Tang, 1998). The majority of studies on Chinese processing utilize word-by-word or phrase-by-phrase display window (sometimes the word boundaries are not clear-cut, so some studies included phrases as well as words as displaying regions, although they described it as word-by-word reading), instead of character-by-character reading (e.g. Chen, Ning, Bi, & Dunlap, 2008; Hsiao & MacDonald, 2016). In the current study, I followed this common practice to investigate word-by-word reading of sentences containing classifier-noun pairs.

In the self-paced reading task, I examined whether learners of Chinese show sensitivity to classifier omissions and uses of incongruent classifier. In searching for more information regarding the nature of difficulty in classifier acquisition, an elicited production task was also included to investigate the relationship between learners' performance on production and online comprehension.

The current study: research questions and predictions

This dissertation focuses on the L2 production and online comprehension of Chinese classifiers. An elicited production task, a self-paced reading task, a lexical decision task, an offline cloze task, and a proficiency test were employed. The research questions and predictions are as follows:

1) Although a few previous studies suggest that it is possible for English-speaking learners to use classifiers as facilitative cues in predictive processing, there is evidence that learners may also behave in a non-nativelike manner in production and comprehension. What is the source of these patterns of behaviors?

a) The source of difficulty may be representational at the syntactic level. That is, learners may not be able to acquire the category Cl. If this is the case, we expect that learners may demonstrate classifier omission in production, limited sensitivity to ungrammatical omission of classifiers, and incongruent classifiers in online comprehension, and non-targetlike performance in the offline task.

b) The source of difficulty may lie mainly at the lexical level, specifically in the association between nouns and their classifiers. If this is the case, we expect that learners may use incongruent or default classifiers in production; they may demonstrate sensitivity to classifier omission, but limited sensitivity to incongruent classifiers, and their performance in the offline task should also be non-target like.⁷

c) Finally, it is possible that the learners in this study will behave in a nativelike manner, indicating that they have fully acquired Cl and the lexical associations between nouns and classifiers. In this case, I expect they will show sensitivity to different types of classifier violation, as well as native-like performance in production and the offline task.

The predictions for this research question are summarized in Table 1.2 below.

Table 1.2

Predictions for the first research question

	Production	Online comprehension	Cloze task
Syntactic	classifier omission	× omission × incongruency	non-targetlike

⁷ As learners were forced to fill in the gap between the numeral and the noun with a specific classifier in the offline task, their responses were highly constrained. They might leave the gap blank or fill it with a random classifier no matter whether syntactic or lexical issue underlines their difficulty in acquisition.

Table 1.2 (cont'd)

Lexical	incongruent or default classifiers	√ omission × incongruency	non-targetlike
Fully acquired	native-like	√ omission √ incongruency	target-like

2) Do learners organize their mental lexicon similarly to native speakers with regard to classifiers? Specifically, do learners rely on semantic information of classifier categories only? Is co-occurrence relationship between classifiers and nouns also accessible to learners for noun categorization? To answer this question, I manipulated the type of incongruent classifiers used in the online comprehension task, which include use of classifiers from the same semantic domain as the target classifier but semantically inconsistent with the noun (e.g. a shape classifier *zhang* (flat-surfaced objects) was used for *pants*, while another shape classifier *tiao* (long and slender objects) is the correct one because pants are long), as well as classifiers from another semantic domain but matching the semantic feature of the nouns to some extent (e.g. a taxonomy classifier *jian* (for clothes) was used for *pants*, although the long shape classifier *tiao* was actually required), I expect that learners' sensitivity to these two types of incongruence may reflect how they organize the mental lexicon, whether they rely on semantic information only, or they are also able to use classifier-noun co-occurrence relationship to determine which semantic feature is picked by the classifier for noun categorization, whether a taxonomy classifier or a shape classifier is required.

a) If learners rely on semantic information only, they will only be sensitive to semantic conflict between incongruent classifiers and nouns. Specifically, in this study,

they will show sensitivity to incongruent classifiers from the same semantic domain (e.g., the flat-surfaced shape classifier *zhang* for *pants*, while the long shape classifier *tiao* is the correct one; *zhang* is not consistent with the shape of pants). When no semantic conflict is involved, their sensitivity will be limited (e.g., the clothes classifier for *pants*, while the long object classifier *tiao* is required), because the nouns match the semantic features of the classifiers, so the incorrect classifier cannot be ruled out based on semantic features only. Co-occurrence information is needed to determine that it is the long shape of pants that is used by the classifier system, not the function; in another word, a shape classifier rather than a taxonomy classifier is required.

b) If learners can use the classifier-noun co-occurrence information in addition to semantic information, they will also be sensitive to incorrect classifiers from another semantic domain (e.g., e.g., the clothes classifier for *pants*, while the long object classifier *tiao* is required), because they are able to learn which feature of the noun is picked by the classifier system based on the co-occurrence relation between classifiers and nouns, whether a shape classifier or a taxonomy classifier is required.

Additionally, I also looked at the production and offline tasks to see whether there is indication that semantic information affects the choice of classifiers in a nativelike way. Specifically, I focused on the types of error made by participants: whether they use inconsistent classifiers from the same semantic domain as the correct classifier or from a different semantic domain.

3) How does the classifier system develop overtime with L2 Chinese proficiency?
Do L2 learners of different Chinese proficiency levels use classifiers differently in

production and comprehension? Does their performance on classifiers improve with L2 proficiency?

CHAPTER 2

METHODOLOGY

Overview

The current study includes a lexical decision task, an elicited production task, a proficiency test, a self-paced reading task targeting participants' online comprehension of classifiers, and an offline cloze task. In the study, participants went through the consent form, then filled out a background questionnaire (Appendix A) on paper. After filling out the background questionnaire, participants took part in the lexical decision task, then the elicited production test. The proficiency test was given after the elicited production task and before the self-paced reading task. The cloze task was given last. The following sections introduce the participants of this study and the tasks.

Participants

Thirty-four English-speaking learners of Chinese and 33 native speakers of Chinese took part in the study. The L2 learners were students who had enrolled in at least three semesters' Chinese classes in a large Midwestern university in the United States. Their ages ranged from 18 to 36 ($M=20.82$, $SD=3.17$), and they started learning Chinese between the age of nine and 25 ($M=14.12$, $SD=4.21$). Among all the learners, seven went to a 10-week study abroad program before the data collection, and another ten lived in China for one to four years ($M=2.25$, $SD=1.27$). The learner participants were asked to self-report their proficiency on a 5-point scale with 1 being "poor" and 5 being "superior" on listening, reading, speaking and writing. The self-report scores were as follows: Reading ($M=3.18$, $SD=0.80$); Speaking ($M=2.94$, $SD=0.70$); Writing ($M=2.76$, $SD=1.00$); Listening ($M=3.03$, $SD=0.87$). Two of the L2 participants were excluded from

data because of their extremely low accuracy on the reading comprehension questions in the self-paced reading task, suggesting they were not proficient enough to complete all tasks in the current study (see details in the results section).

The native participants were recruited from the same large Midwestern university as well as a Midwestern liberal arts college. The majority of the native speakers recruited were full time students or visiting scholars in the university and the college ($N=32$). The native participants' ages ranged from 19 to 40 ($M=24.88$, $SD=5.62$), and their length of stay in the United States ranged from 2 months to 12 years ($M=3.29$ years, $SD=2.62$). All of them spoke Mandarin, and 23 of them spoke one or two dialects in addition to Mandarin.

Proficiency test

Procedure and materials

In addition to self-rated proficiency, a reading test was also included as a measure for Chinese proficiency. The reading test was adapted from the HSK (Chinese Proficiency Test), which is a standardized test that is widely used in China for assessing Chinese learners' ability to use Chinese in daily life and academic and professional lives. The HSK includes six levels. According to the administrator of the test, Hanban/Confucius Institute Headquarters, Levels 1-6 of HSK are equivalent to Levels A1, A2, B1, B2, C1, C2 of the *Common European Framework of Reference* (CEF) respectively.

The HSK consists of three parts: listening, reading, and writing. Speaking is tested separately in the HSK Speaking Test (HSKK). The proficiency measure used in this dissertation is adapted from the reading part of the mock tests available on the official

website of the HSK. It is composed of three sections with five items in each section. The first section, adapted from the mock HSK test for Level 3, involves choosing correct words to complete sentences (only one part of the test was used because of length restriction). The second section was adapted from the Level 4 test. The task is to choose correct words to complete short conversations. The third part, adapted from the Level 5 test, involves choosing correct words to complete a short story. Questions targeting classifiers were excluded. The proficiency test has 15 items in all. The test is included in Appendix B.

All participants took the test on paper. The time limit for the test was 15 minutes, and all participants finished the test within the time limit. In cases that participants from the L2 group reported that parts of the test were far beyond their level, they were allowed to skip the difficult section(s).

Lexical decision task

Procedure and materials

After filling out the background questionnaire, participants completed a lexical decision task. The test was modeled on LexTALE, and administered on a computer via *Praat* 6.0.30 (Boersma & Weenink, 2017), in a quiet lab. In the test, participants saw three practice items and 60 words on a computer screen. The words appeared in the middle of the screen, one word each time. Participants were instructed to decide whether the word they saw was a real Chinese word or not. If they thought the word is an existing Chinese word, they hit ‘Y’ on the key board, if not, they hit ‘N’. They hit ‘Y’ if they thought the word is a real word, even if they could not recall the meaning of the word; if they were not sure whether the word exists, they were asked to hit ‘N’. After hitting one

of the two keys, the next word appeared. There was no time limit to this task, but participants were instructed to respond as quickly as possible. All participants completed the task in no more than three minutes. The test score based on accuracy was calculated automatically. Reaction times were also recorded.

The test includes three practice items and 60 critical items, all of which were nouns. The word list consisted of 40 real words and 20 non-words. In the 40 real words, 32 were the target words investigated in the elicited production and self-paced reading tasks. Of the rest of the eight real words, four were likely to be unfamiliar to L2 learners, for instance, *tange* 探戈, ‘tango’; the other four were likely to be unfamiliar combinations of familiar words, for instance, *jiu* (‘wine’)*guan* (‘room’) 酒馆, ‘bar’.

Non-words were created in three ways. The first type of non-word was a non-existing combination of characters, for instance, **xuesan* 雪伞, ‘snow umbrella’. The second type was words with one character replaced by another character that has a similar meaning, pronunciation or shape, for instance, **weiqing* 味情, which was made up from the real word *weijing* 味精 ‘MSG’. The third type was made by reversing the word order in real words, for example, **jingyan* 睛眼, which was based on *yanjing* 眼睛, ‘eye’. The whole list of words is in Appendix C.

Elicited production task

Procedure and materials

The elicited production task was administered in a quiet laboratory space. Data collection was carried out individually. In the elicited production task, three pairs of pictures including a total of 32 target objects were used (See Appendix D). In these two

pictures within each pair, target objects differ in quantity; participants were instructed to spot and describe the differences. A sample description was provided, as shown in (4).

(4) Zher you yige beizi, nar you liangge.

Here have one-Cl cup there have two-Cl

‘There is one cup, there are two cups’

In the description, participants produced sentences similar to the sample sentence, or other sentence patterns such as ‘There is one more cat in this picture.’ Classifier-noun pairs were successfully produced in this task. In cases that participants failed to notice the difference or produce the classifier, the experimenter reminded them to use sentences such as ‘How about here?’; ‘How many?’ No classifiers were used in the prompts. When participants had difficulty naming the objects, which turned out to be frequent in the experiment, the experimenter reminded them of the words, and again use of classifiers was avoided. It took around 2-3 minutes for native speakers of Chinese to complete the task, and 3-7 minutes for learners.

The three pairs of pictures depict 32 noun pairings with classifiers from three semantic domains: animacy, shape, and function (seven classifiers, each with three to six nouns⁸), together with a few distractors in which some objects differ in color or size. All of the classifiers and nouns were selected from the Chinese textbooks the L2 group had used in their first three semester Chinese classes (Liu, Yao, Bi, Shi, & Ge, 2009). The target nouns and classifiers are shown in Table 2.1 below.

⁸ Due to the fact that some classifiers have more noun members than others, as well as learners’ limited vocabulary size, it is not very practical to come up with an even number of nouns with each classifier for stimuli.

Table 2.1

Classifiers and nouns used in the elicited production task

Semantic domain	Classifier	Noun
Animacy	<i>zhi</i> (small animals)	<i>mao</i> ('cat'), <i>xiaoniao</i> ('bird'), <i>ji</i> ('chicken'), <i>yazi</i> ('duck'), <i>yang</i> ('sheep/goat')
Shape	<i>tiao</i> (long and slender objects)	<i>kuzi</i> ('pants'), <i>qunzi</i> ('skirt'), <i>chuan</i> ('boat') <i>xiaolu</i> ('road'), <i>tanzi</i> ('blanket'), <i>yu</i> ('fish')
	<i>zhang</i> (flat surfaced objects)	<i>zhuozi</i> ('table'), <i>zhaopian</i> ('photo'), <i>ditu</i> ('map'), <i>xinyongka</i> ('credit card'), <i>zhi</i> ('paper'), <i>chuang</i> ('bed')
Function	<i>jian</i> (clothes)	<i>chenshan</i> ('shirt'), <i>maoyi</i> ('sweater'), <i>waitao</i> ('overcoat'), <i>jiake</i> ('jacket'), <i>yundongfu</i> ('sweatshirt'), T-xushan ('T-shirt')
	<i>liang</i> (vehicles)	<i>chuzuche</i> ('taxi'), <i>zixingche</i> ('bicycle'), <i>motuoche</i> ('motorcycle')
	<i>ben</i> (bounded items)	<i>zidian</i> ('dictionary'), <i>keben</i> ('textbook'), <i>shu</i> ('book')
	<i>tai</i> (machine)	<i>diannao</i> ('computer'), <i>dianshi</i> ('television'), <i>bingxiang</i> ('refrigerator')

Data coding

Production data was coded for accuracy of classifier use. Four categories were used in data coding: correct use of specific classifiers in required context; incorrect use of specific classifiers; ungrammatical omission of classifiers; and use of the default classifier (*ge*).

Self-paced reading task

Procedure and materials

A word-by-word non-cumulative moving window self-paced reading task was carried out after the production task, aiming to investigate participants' processing of classifier-noun combinations. The task was administered on a computer via *SuperLab* in a quiet laboratory room. Participants were instructed to read at a natural speed, and not to take breaks in the middle of reading sentences. Five practice items were presented to familiarize participants with the task. The task took approximately 15 minutes for native speakers of Chinese to complete and around 30 minutes for L2 learners.

In this task, the same seven classifiers and 32 nouns used in the elicited production task were included. There were four conditions for each sentence, which were the grammatical condition, the classifier omission condition, the incongruent condition in which a classifier from another semantic domain is used, which is also the semantic-consistent condition, because there is no semantic clash between the classifier and the noun, as well as another incongruent condition in which another classifier from the same semantic domain but denoting different properties is used, which is also the semantic-inconsistent condition, because there is semantic clash between the classifier and the noun. The sentences were divided into four lists using a Latin-square design. Each

participant was presented with one of the four lists composed of 32 target sentences (eight items in each condition), along with 64 fillers, so participants saw each sentence in only one condition. All of the target sentences and half of the fillers were followed by a true/false comprehension question to ensure participants processed the sentences. The use of classifiers was avoided in all the questions. A sample set of stimuli is shown below in (5) (display regions are divided by '/').

(5) Grammatical/classifier omission/cross semantic domain/within semantic domain condition.

小王/看到/一只/猫/在/桌子/上/睡觉。

Xiaowang kandao yizhi mao Zai zhuozi shang shuijiao

**yi*

**yitiao*

**yiwei*

Xiaowang See one-(Cl) **Cat** On table sleep

‘Xiaowang saw a cat sleeping on the table.’

Comprehension question:

True or false: *Mao zai chuangshang shuijiao*. (‘The cat was sleeping on bed.’)

The critical region was the noun after the classifier. The classifiers appeared after a numeral in 16 sentences, and after a demonstrative in 16 sentences; both types of context are very familiar to learners. In all sentences, the critical region was followed by a

preposition (*zai* ‘on/in/at’, *cong* ‘from’, or *gei* ‘to’) and a noun (two characters) to ensure consistent spill-over regions. Reading time on the following two regions (preposition + noun) was analyzed for spill-over effects. All the classifier-noun pairs and the classifiers used in the two types of incongruence condition are listed in Table E1 in Appendix E. The full list of sentences is included in Appendix F.

Offline cloze task

Procedure and materials

An offline paper-and-pencil cloze task was carried out after the elicited production task and the self-paced reading task. Materials were the 32 nouns used in the previous two tasks, appearing in numeral phrases. Participants were instructed to fill in a classifier to complete the phrase, and they were instructed not to use the general classifier *ge*. A sample item of the task is shown below in (6).

(6) 三 _____ 猫

three cat (‘three cats’)

There was no time limit for this task. It took less than five minutes for native speakers of Chinese, and less than ten minutes for learners.

CHAPTER 3

RESULTS

Proficiency test

The proficiency test consisted of 15 multiple choice items adapted from the Level 3, Level 4, and Level 5 reading test of the HSK, which are equivalent to Level B1, B2, and C1 of the *Common European Framework of Reference* (CEF) respectively. There were five items for each level. Thirty-three native speakers of Chinese and 34 English-speaking learners of Chinese participated in this task, and 32 L2ers were included in data analyses. The native speakers of Chinese performed at ceiling in the test, with an average score of 14.91 (out of 15) ($SD=3.02$). An unpaired t-test revealed a significant difference between the native group and the learner group ($p<.001$). The two participant groups' mean accuracy on the three parts of the proficiency test is shown in Table 3.1. The learner group performed better on Part 1 compared to Part 2, and they also performed better on Part 2 than Part 3, suggesting the test worked well in differentiating participants in terms of L2 proficiency.

Table 3.1

Mean scores on different parts of the proficiency test

	Natives mean accuracy (SD)	Learners mean accuracy (SD)
Part 1 (Level 3)	1 (0)	.88 (.21)
Part 2 (Level 4)	1 (0)	.54 (.31)
Part 3 (Level 5)	.98 (.06)	.37 (.27)

For further analysis of whether L2 proficiency affects Chinese classifier proficiency, the L2 group was divided into two groups based on proficiency test scores. Learners who had an accuracy of over 60% (scoring no less than 9) were categorized into the higher performance group, and learners who scored below 9 were categorized into the lower performance group. The descriptive results of the participants' performance are shown below in table 3.2. An unpaired t-test showed a significant difference between these two subgroups ($p < .001$).

Table 3.2

Accuracy of the overall learner group and two subgroups in the proficiency test

	All L2s N=32	Higher proficiency L2 group N=17	Lower proficiency L2 group N=15
Mean	.59	.75	.41
Range	.27-1	.60-1	.27-.53
SD	.20	.13	.09

Offline cloze task

In the offline task, participants were asked to fill in a classifier for each of the 32 nouns that were used in the online comprehension task and the production task, and they were instructed to not use the general classifier *ge*. For learners, they were encouraged to write characters, but the transliteration of Chinese, *Pinyin*, was also acceptable, and writing or spelling errors were ignored. Because it is possible that each noun can be paired with multiple classifiers, two native speakers of Chinese worked as raters for this task; only those responses that were rated as acceptable by both raters were marked as correct.

The same 33 native speakers of Chinese and 34 English-speaking learners of Chinese took part in this task. Two learners' results were excluded as stated. Unsurprisingly, all the classifiers provided by native speakers were acceptable. On the other hand, the task appeared to be challenging for learners of Chinese. Learners' accuracy ranged from 0.19 to 0.91, with a mean accuracy of 0.56 ($SD=0.22$). Participants were also encouraged to mark the nouns that share classifiers, even when they could not remember the classifiers. As a result, ten out of the 32 learners reported that they were aware of some semantic rules for classifiers. For example, they knew that there was a specific classifier for animals, or vehicles, but they could not remember them, so they were aware of the rule but not the form. When the classifier was presented to them after they completed all the tasks, they could recall it and read it out. It seems that retrieving the classifiers tended to be difficult for learners. A multiple choice format would make the offline task easier⁹, but it is subject to discussion which option is a better measurement of participants' knowledge of classifiers.

Seven classifiers from three semantic domains were included in the current study. The mean accuracy of the L2 group for each classifier and noun pair in different domains is shown in Table 3.3. It seems that there was no fundamental difference among semantic domains: the animacy classifier had an average accuracy of 0.55; the shape classifiers had an average accuracy of 0.50, and the function classifiers had an accuracy of 0.59. As for individual classifiers, the one with the highest accuracy was *ben*

⁹ In Liu and Spinner (manuscript), a multiple-choice task was used, and the learners performed much better on the offline task than the learner participants did on the offline task in the current study. Although the participants were of similar proficiency level, and most of the target classifiers and nouns in these two studies overlapped, the learner group had an average accuracy of 0.86 on the multiple-choice task in the previous study, while the average accuracy on the cloze task in the current study was only 0.56.

(bounded items) (0.79), followed by *tai* (machine) (0.65) and *zhang* (flat surfaced objects) (0.65).

Table 3.3

Learners' mean accuracy of each classifier-noun pair

Semantic domain	Classifier	Noun	Mean accuracy
Animacy	<i>zhi</i> (small animals)	<i>mao</i> ('cat'), <i>xiaoniao</i> ('bird'), <i>ji</i> ('chicken'), <i>yazi</i> ('duck'), <i>yang</i> ('sheep/goat')	.55
Shape	<i>tiao</i> (long and slender objects)	<i>kuzi</i> ('pants'), <i>qunzi</i> ('skirt'), <i>chuan</i> ('boat'), <i>xiaolu</i> ('road'), <i>tanzi</i> ('blanket'), <i>yu</i> ('fish')	.34
	<i>zhang</i> (flat surfaced objects)	<i>zhuozi</i> ('table'), <i>zhaopian</i> ('photo'), <i>ditu</i> ('map'), <i>xinyongka</i> ('credit card'), <i>zhi</i> ('paper'), <i>chuang</i> ('bed')	.65
Function	<i>jian</i> (clothes)	<i>chenshan</i> ('shirt'), <i>maoyi</i> ('sweater'), <i>waitao</i> ('overcoat'), <i>jiake</i> ('jacket'), <i>yundongfu</i> ('sweatshirt'), <i>T-xushan</i> ('T-shirt')	.49
	<i>liang</i> (vehicles)	<i>chuzuche</i> ('taxi'), <i>zixingche</i> ('bicycle'), <i>motuoche</i> ('motorcycle')	.43
	<i>ben</i> (bound items)	<i>zidian</i> ('dictionary'), <i>keben</i> ('textbook'), <i>shu</i> ('book')	.79

Table 3.3 (cont'd)

<i>tai</i> (machine)	<i>diannao</i> ('computer'), <i>dianshi</i> (‘television’), <i>bingxiang</i> (‘refrigerator’)	.65
----------------------	---	-----

For further investigation of the relationship between offline knowledge and online comprehension and production of classifiers, the L2 participants were divided into two sub-groups based on their accuracy on the offline cloze task. L2 participants whose mean accuracy was higher than the group average were categorized as the higher-performance group; learners whose mean accuracy was below the group average were categorized as the lower-performed group. The information of these two groups is shown in Table 3.4.

It is possible that learners' knowledge may grow with their L2 proficiency. However, only a weak correlation was found between L2 participants' accuracy on the offline cloze task and the proficiency test ($r=0.29$), suggesting that learners' knowledge of classifiers might not grow when they get more proficient in Chinese. This issue will be further explored in the Discussion.

Table 3.4

Accuracy of the overall learner group and two subgroups in the offline cloze task on classifiers and their accuracy on the proficiency test

	All L2ers N=32		Higher performance L2ers N=17		Lower performance L2ers N=15	
	Cloze	Proficiency	Cloze	Proficiency	Cloze	Proficiency
Mean	.56	.59	.73	.60	.36	.57
Range	.19-.91	.26-1	.59-.91	.26-1	.19-.56	.33-.87

Table 3.4 (cont'd)

SD	.22	.20	.10	.23	.11	.17
----	-----	-----	-----	-----	-----	-----

Lexical decision task

In the lexical decision task, participants were presented with all the 32 target nouns together with 28 fillers individually on computer, and they were instructed to decide whether the word they saw was a real word or not as quickly as possible. As noted above, the fillers included eight real words and 20 non-words. The non-words were created by combining real characters in non-existing ways. Participants' accuracy as well as reaction time to each word was recorded for analyses. The same 33 native speakers of Chinese and 34 English-speaking learners of Chinese took part in the task. Again, only 32 learners' data were included in the analyses. Extreme reaction times beyond 2.5 SD of the group mean were removed, which affected 1.52% of the native speakers' data and 1.07% of the learners' data.

For native speakers of Chinese, their accuracy on the target nouns was very high, with an average accuracy of 0.99 ($SD=0.05$). Learners had an average accuracy of 0.84 ($SD=0.36$).

As for reaction time, native speakers of Chinese had an average reaction time of 1.22 seconds for the target nouns ($SD=0.68$). Learners had a longer reaction time for the target nouns, which was 2.44 seconds ($SD=2.42$). An unpaired t-test confirmed that learners reacted to the nouns significantly more slowly than native speakers ($p<.001$). In the section of the comprehension task results, participants' reaction time will be analyzed with their reading time to investigate whether lexical retrieval speed influences participants' sensitivity to different types of violations regarding classifiers.

Elicited production task

In the production task, participants looked at three pairs of near-identical pictures and described the differences between each pair. The pictures depicted all the 32 target nouns with fillers, and the target nouns differed in number. Therefore participants had to use numeral phrases to describe the differences. The same 33 native speakers of Chinese and 34 English-speaking learners of Chinese took the task, but again two of the learners' data were excluded due to low performance on the self-paced reading task.

In the task, participants' responses were recorded, and they were coded by two native speakers of Chinese together. Participants described the differences using different types of sentences, such as “there are two cats” or “that picture has one more cat.” Numeral phrases with classifiers were used by all participants. Both the general classifier *ge* and various specific classifiers were used to complete the task.

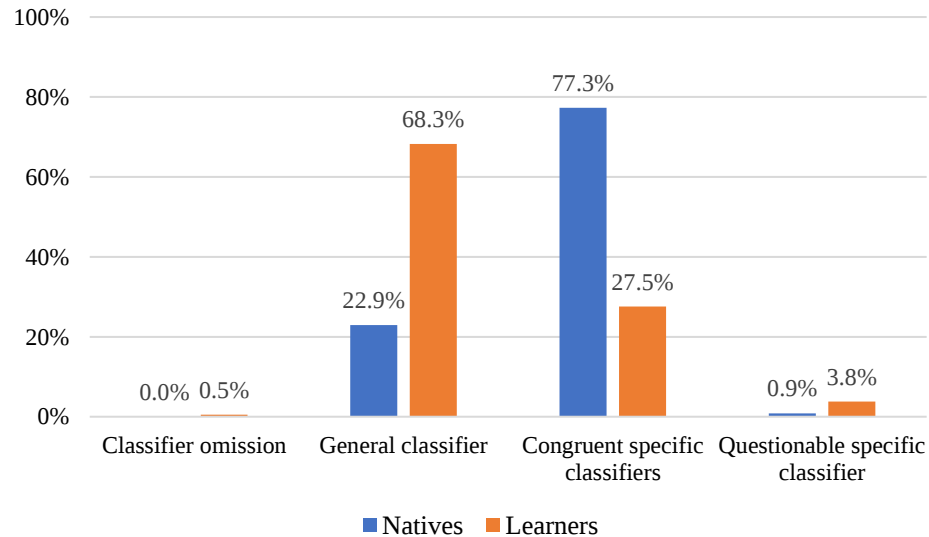
There were four types of classifier use: classifier omission, use of the general classifier *ge*, use of congruent specific classifiers, and use of questionable or incongruent specific classifiers (for instance, *shuang* ‘pair’ for pants). Each classifier use was judged as congruent or questionable by the two native raters together and differences of opinion were resolved through discussion. The percentage of each use is shown in Figure 3.1 below.

It can be seen from the figure that native speakers of Chinese never omitted the classifier, and similarly classifier omission was rare for learners of Chinese, only accounting for 0.5% of learners' responses ($SD=0.01$). Learners tended to overuse the general classifier *ge*, using it for 68.3% of all the responses ($SD=0.25$). Native speakers used the general classifier much less frequently than learners, for 22.9% of all responses

($SD=0.23$). On the other hand, native speakers of Chinese used specific classifiers more frequently than learners; although the native speakers did not use specific classifiers all the time, the overall percentage of congruent specific classifiers used was 77.3% ($SD=0.22$), while learners only used congruent specific classifiers in 27.5% of all the cases ($SD=0.24$). In rare cases, native speakers of Chinese used questionable classifiers (0.9%, $SD=0.01$), which might be due to the influence of dialects. For instance, one of the native participant used the classifier *jia* for *taxi*, while in Mandarin *jia* is usually used with *air plane*, the questionable use of classifier might be due to the influence of southern dialects. Learners of Chinese used more questionable specific classifiers than natives, which accounted for 3.8% of their overall classifier use ($SD=0.05$).

Figure 3.1.

Percentage of each type of classifier use by all participants in the production task



In Mandarin Chinese, some nouns can appear with multiple classifiers. Therefore unsurprisingly, various classifiers were used for some target nouns in the task. As this study focused on Mandarin classifiers, the classifiers that were used in dialects but not in Mandarin were also regarded as questionable. Table 3.5 shows all the classifiers used by the native group and the learner group for each target noun in the production task and their frequency.

Generally speaking, native speakers of Chinese used more specific classifiers than the general classifier, and they showed variation in classifier use for some nouns. They used specific animacy classifier (*zhi*) for small animals and function classifier (*ben*) for books more frequently than other classifiers; around 30 out of 33 native participants used the specific classifiers for each noun in these two categories. On the other hand, they used the function classifier for machines less frequently. Only around half of the native participants used *tai* for computer, television and refrigerator. The rest used the general classifier for them.

Most of the different choices of classifiers were related to size of the object. For instance, the classifier *zhi* is for small animals, and another animacy classifier, *tou*, is for large animals such as elephants, tigers. In the production task, four of the native participants used *tou* for sheep/goat, which was congruent, and their choice might have to do with their perception of the size of the animal. This was also the case for the photo and map in the task; for picture-like objects, *fu* instead of *zhang* is often used for large ones and it is also more formal. For boat, *tiao* or *zhi* is often used for small ones such as canoes, while for large ones such as cruises or ferries, *sou* is often used. The boat in the picture for the task looks like a ferry, so more participants used *sou* instead of *tiao* in their

production. Another noun that had multiple congruent classifiers was *qunzi* ‘skirt’. A few participants used the function classifier for clothes, *jian*, instead of the shape classifier *tiao*. In Chinese, *qunzi* can mean skirt as well as dress, and usually *jian* is used for dress, so several participants used *tiao* for the skirt in the picture probably because of the confusion. Both *tiao* and *zhang* were used for blankets by native speakers, which might be due to their different perception of the shape of the object. Use of questionable classifiers was rare, and they were probably related to the influence of dialects.

The learner group relied much more on the general classifier *ge*, and they used specific classifiers less frequently compared to native speakers of Chinese. The category of nouns that they used specific classifiers with the most was volumes; around half of the learner participants used the function classifier *ben* for *textbook*, *book* and *dictionary*. They also used the animacy classifier *zhi* for small animals relatively frequently, as well as the function classifier *jian* for clothes, and *tai* for machines. Their use of the function classifier *liang* for vehicles was limited; only five learners used it for *taxi*, two of them used it for *motorcycle*, and one of them used it for *bike*. As for the shape classifier *zhang* and *tiao*, the use varied depending on nouns. For some nouns such as *paper*, *pants*, and *bed*, over ten participants used the specific shape classifiers for each of them. As for *boat*, *blanket*, *fish*, *table*, and *credit card*, less than five learner participants used the shape classifiers for each noun. It is possible that learners used specific classifiers more for familiar nouns, such as *textbook*, *paper*, and *cat* (Gao, 2009), as well as prototypical members of the category, such as *bed* and *taxi* (Myers, 2000), while general classifier would be used more frequently for unfamiliar nouns and non-prototypical members.

Table 3.5

Native speakers and learners' use of classifiers for each noun

Noun	Frequency of classifiers used by native speakers (L1) and L2 learners (L2)									
	Null		<i>ge</i>		Specific classifiers (frequency)					
	L1	L2	L1	L2	L1				L2	
<i>mao</i> ('cat')	0	0	2	17	<i>zhi</i> (31)				<i>zhi</i> (14)	? <i>fen</i> (1)
<i>xiaoniao</i> ('bird')	0	0	7	23	<i>zhi</i> (26)				<i>zhi</i> (8)	? <i>fen</i> (1)
<i>ji</i> ('chicken')	0	0	3	23	<i>zhi</i> (30)				<i>zhi</i> (9)	
<i>yazi</i> ('duck')	0	0	1	25	<i>zhi</i> (32)				<i>zhi</i> (6)	? <i>fen</i> (1)
<i>yang</i> ('sheep/goat')	0	0	3	22	<i>zhi</i> (26)	<i>tou</i> (4)			<i>zhi</i> (10)	
<i>kuzi</i> ('pants')	0	0	6	13	<i>tiao</i> (27)				<i>tiao</i> (13)	? <i>shuang</i> (3) ? <i>zhang</i> (1)
									? <i>jian</i> (1)	? <i>zhi</i> (1)
<i>qunzi</i> ('skirt')	0	0	6	19	<i>tiao</i> (24)	<i>jian</i> (2)	? <i>fu</i> (1)	<i>tiao</i> (8)	<i>jian</i> (2)	? <i>zhang</i> (2)
									? <i>duan</i> (1)	

Table 3.5 (cont'd)

<i>chuan</i> ('boat')	0	1	10	29	<i>tiao</i> (3)	<i>sou</i> (15)	<i>zhi</i> (5)	<i>tiao</i> (0)	? <i>tai</i> (2)
<i>xiaolu</i> ('road')	0	0	1	22	<i>tiao</i> (32)			<i>tiao</i> (9)	<i>dao</i> (1)
<i>tanzi</i> ('blanket')	0	0	10	26	<i>tiao</i> (18)	<i>zhang</i> (5)		<i>tiao</i> (2)	<i>zhang</i> (3) ? <i>zhi</i> (1)
<i>yu</i> ('fish')	0	0	6	22	<i>tiao</i> (22)	? <i>zhi</i> (5)		<i>tiao</i> (5)	? <i>zhi</i> (4) ? <i>fen</i> (1)
<i>zhuozi</i> ('table')	0	1	18	25	<i>zhang</i> (13)	? <i>tai</i> (2)		<i>zhang</i> (5)	? <i>tai</i> (1)
<i>zhaopian</i> ('photo')	0	0	6	25	<i>zhang</i> (20)	<i>fu</i> (7)		<i>zhang</i> (7)	
<i>ditu</i> ('map')	0	0	8	23	<i>zhang</i> (12)	<i>fu</i> (113)		<i>zhang</i> (9)	
<i>zhi</i> ('paper')	0	1	11	13	<i>zhang</i> ()			<i>zhang</i> (17)	? <i>fen</i> (1)
<i>chuang</i> ('bed')	0	0	7	20	<i>zhang</i> ()			<i>zhang</i> (11)	? <i>tai</i> (1)
<i>xinyongka</i> ('credit card')	0	0	13	26	<i>zhang</i> ()			<i>zhang</i> (5)	? <i>fen</i> (1)
<i>chenshan</i> ('shirt')	0	0	8	23	Null (0)			<i>jian</i> (8)	? <i>tiao</i> (1)
<i>maoyi</i> ('sweater')	0	0	7	20	Null (0)			<i>jian</i> (10)	? <i>tiao</i> (2)
<i>waitao</i> ('overcoat')	0	0	6	22	Null (0)			<i>jian</i> (8)	? <i>tiao</i> (1) ? <i>tao</i> (1)

Table 3.5 (cont'd)

<i>jiake</i> ('jacket')	0	0	7	22	Null (0)		<i>jian</i> (9)	? <i>tiao</i> (1)	
<i>yundongfu</i> ('sweatshirt')	0	0	8	20	Null (0)		<i>jian</i> (11)	? <i>tiao</i> (1)	
<i>T-xushan</i> ('T-shirt')	0	0	7	24	Null (0)		<i>jian</i> (8)		
<i>chuzuche</i> ('taxi')	0	0	4	26	Null (0)	? <i>jia</i> (1)	<i>liang</i> (5)	<i>tai</i> (1)	
<i>motuoche</i> ('motorcycle')	0	0	5	28	Null (0)		<i>liang</i> (2)	<i>tai</i> (2)	? <i>zhi</i> (1)
<i>zixingche</i> ('bicycle')	0	1	9	28	Null (1)		<i>liang</i> (1)	<i>tai</i> (1)	? <i>zhi</i> (1)
<i>zidian</i> ('dictionary')	0	0	4	17	Null (0)		<i>ben</i> (15)		
<i>keben</i> ('textbook')	0	0	5	15	Null (0)		<i>ben</i> (17)		
<i>shu</i> ('book')	0	0	3	14	Null (0)		<i>ben</i> (17)	? <i>zhang</i> (1)	
<i>diannao</i> ('computer')	0	1	12	22	Null (1)		<i>tai</i> (8)	? <i>tiao</i> (1)	
<i>dianshi</i> ('television')	0	0	16	23	Null (0)		<i>tai</i> (8)	? <i>zhang</i> (1)	
<i>bingxiang</i> ('refrigerator')	0	0	23	22	Null (0)		<i>tai</i> (8)	? <i>zhang</i> (1)	? <i>jia</i> (1)

Note: '?' indicates questionable classifiers

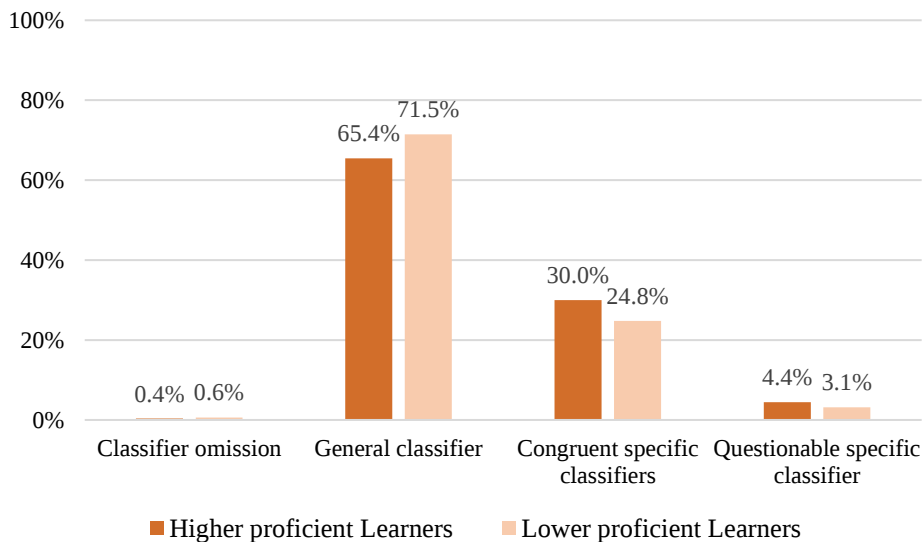
Learners of Chinese also showed more variation in classifier production and more ungrammatical use of specific classifiers than native speakers. The most typical type of ungrammaticality was to use a classifier from a different semantic domain than the correct classifier. For instance, some participants used the shape classifiers *tiao* or *zhang* for clothes such as *shirt* and *sweater*, and machines such as *computer*, *television* and *refrigerator*, where the function classifiers *jian* and *tai* are required; *jian* was also used for pants while the shape classifier *tiao* is required. This pattern suggests that learners rely more on semantics in classifiers use (Grüter et al., 2018), as there was no conflict of semantic features between the nouns and the classifiers they used, suggesting they are familiar with the semantic feature of the classifiers. They extended the classifier to nouns that match the features (e.g., use the clothes classifier *jian* for all objects that fall in this category), and the error reflects learners' knowledge of the semantic information of the classifier. Classifier-noun co-occurrence information in addition to semantics is needed to differentiate the semantically consistent options of classifiers and determine the grammatical one. The use of classifiers from the same semantic domain as the correct classifier was also present, but it was limited. For instance, *zhang*, which is for flat-surfaced objects, was used for *pants* and *skirt*, though the long object classifier *tiao* is the grammatical one. It seems to be difficult to locate the potential reasons for the misuse, because the nouns did not match the semantic features of the classifiers, and the classifier they used would not co-occur in their input. I speculate that such misuse might be due to learners' different perception of the shape, for instance, learner might think that pants and skirts are flat instead of long, and therefore a mismatch on semantic features was not

necessarily involved in this case, because their perception of the shape was consistent with the classifier they used, although it was not grammatical.

To investigate whether learners of different L2 Chinese proficiency showed different patterns in classifier use, I divided the learners group into a higher proficiency group (N=17) and a lower proficiency group (N=15) based on their performance on the proficiency test. Figure 3.2 shows the percentage of each of the four types of classifier use by the two proficiency groups.

Figure 3.2.

Percentage of each type of classifier use by learners in the production task



It seems that these two proficiency groups did not differ greatly in terms of classifier use. Both the higher proficiency group and the lower proficiency group rarely omitted classifiers. The higher proficiency group only dropped the classifier for 0.4% of all the responses, and the lower proficiency group dropped the classifier for 0.6% of all the

responses. Both groups tended to use the general classifier for nouns. The higher proficiency group used the general classifier for 71.5% of all the responses, and the lower proficiency group used it for 65.4% of all the responses. The higher proficiency group used slightly more congruent specific classifiers (30.0%) than the lower proficiency group (24.8%), and they also used slightly more questionable specific classifiers than the lower proficiency group (4.4% vs. 3.1%), which might be due to their more frequent attempts to use specific classifiers.

Since all participants rarely omitted classifiers or used questionable specific classifiers, I then focused on their use of the general classifier and congruent specific classifiers to investigate how the use of the general classifier and specific classifier changes with L2 proficiency. Figure 3.3 shows the relationship between the rate of the general classifier used by learners and their scores in the proficiency test; Figure 3.4 shows the relationship between the rate of specific classifiers and proficiency test scores. A Spearman's rho test showed that there was a weak negative correlation between learners' use of the general classifier and their scores on the Chinese proficiency test ($r = -.29, p = .11$), indicating that there was a trend that more proficient learners of Chinese relied less on the general classifier in production. As for the relationship between the grammatical use of specific classifiers and L2 proficiency, it was found that there was a weak positive correlation between the rate of using specific classifiers that were congruent with the nouns and the proficiency score ($r = .24, p = .18$).

Figure 3.3.

Relationship between the use of the general classifier and proficiency score

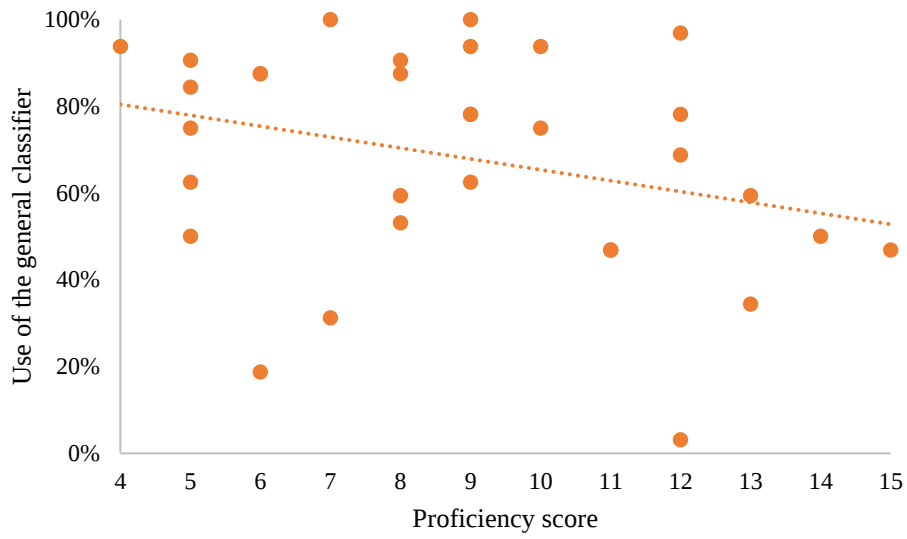
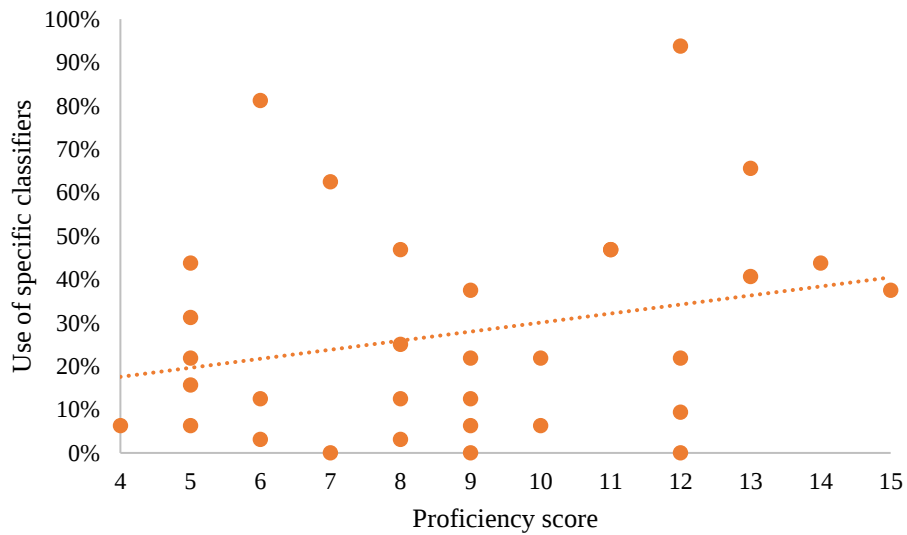


Figure 3.4.

Relationship between the use of congruent specific classifiers and proficiency score



Learners' knowledge of classifiers might also influence their performance on the production task, so Spearman's rho tests were carried to investigate the relationship between learners' use of the general classifier and the score on the classifier knowledge task, as well as the relationship between the rate of using correct specific classifiers and the classifier knowledge task score. A strong negative correlation was found between the rate of using the general classifier and the cloze task score ($r = -.61, p < .001$). Meanwhile, a strong positive correlation was detected between the rate of using correct specific classifiers and the cloze task score ($r = .56, p < .001$). The results suggested that more knowledge of classifiers enables learners to rely less on the general classifier in production, and use more specific classifiers.

To summarize, in the production task, English-speaking learners of Chinese showed a different pattern in use of classifiers compared to native speakers of Chinese. They overused the general classifier *ge* and used specific classifiers much less frequently than native speakers. They also used more ungrammatical specific classifiers. Native speakers never omitted classifiers. Learners of Chinese showed classifier omissions, but only in very rare cases. The results may suggest that learners of Chinese were able to acquire classifiers syntactically, an issue explored in the Discussion. On the other hand, classifiers were challenging in terms of semantics.

When the selection of a classifier was incorrect, learners typically used a classifier from a different semantic domain as the correct classifier in most cases, indicating they may organize nouns differently than native speakers, and they might rely more on semantic features to determine which classifier category a noun should belong to,

because the most frequent type of errors was to use an incorrect classifier with nouns that match the semantic features.

To turn to the relationship between classifier use and L2 Chinese proficiency, lower proficiency learners and higher proficiency learners of Chinese showed a similar pattern in classifier production. Both groups overused the general classifiers, and their use of specific classifiers was relatively limited. They both omitted classifiers in rare cases. Meanwhile, learners who performed better on the classifier knowledge task relied significantly less on the general classifier, and used more specific classifiers. There was a trend that with the growth of L2 proficiency, learners relied less on the general classifier and used more specific classifiers, but the relationship was not significant. It is possible that the relationship will be found to be stronger with learners with higher Chinese proficiency.

Online comprehension task

Data trimming and analysis

For the self-paced reading task, thirty-two sets of sentences were included as stimuli, each set consisting of four conditions: the grammatical condition, the classifier omission condition, the incongruent condition in which another classifier from a different semantic domain was used and there was no semantic clash between the classifier and the noun (incongruent, no semantic clash condition), as well as another incongruent condition in which another classifier from the same semantic domain was used and there was semantic clash between the classifier and the noun (incongruent, with semantic clash condition). The sentences were divided into four lists using a Latin-square design, so each participant saw each sentence in only one condition. Each participant read 32 target

sentences, (and also 64 fillers), therefore contributing 32 observations to the overall data set, eight observations in each condition.

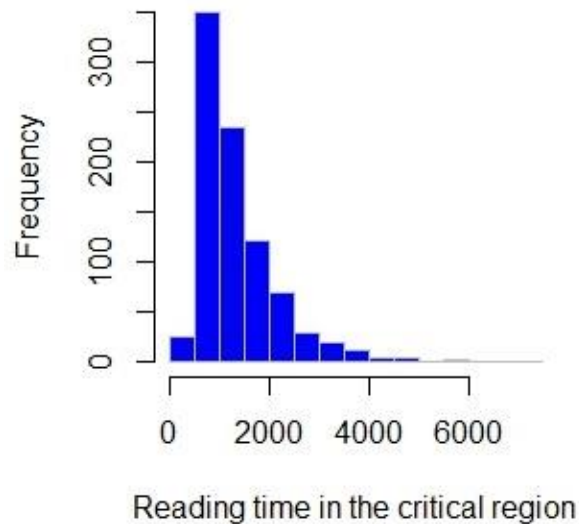
For the learner group, the mean accuracy for the comprehension questions after the sentences was 0.82 ($SD=0.12$). The native group had an average accuracy of 0.96 ($SD=0.03$). Participants whose accuracy was lower than 70% were excluded, resulting in the exclusion of two learners but no native speakers, and their data were also excluded from other tasks in the current study because of the concern they were not proficient enough for the study. After the exclusion, the data from 32 learners and 33 native speakers were included in the analysis. Reading times beyond 2.5 SD of the participant mean at each region were removed (Jegerski, 2014), which affected 3.6% of the learners' data, and 3.7% of the native speakers' data. In addition, for the learner group, if their response for a noun in the lexical decision task was incorrect, the corresponding item in the self-paced reading task was removed; this further affected 17.2% of the learners' data.

Many previous studies have used ANOVA to analyze reading time data (Jegerski, 2014; Keating & Jegerski, 2015). However, in such approach, mean reading times are used, which average across individual responses, and thus call for the need of both by-participant analysis and by-item analysis. In addition, reading times have an absolute bound; they cannot be less than zero; usually they are also not normally distributed, which violates the assumptions of ANOVA, which requires a normal distribution of data (Lo & Andrews, 2015). For instance, Figure 3.5 shows that the distribution of reading times of the learner group on the critical region was positively skewed in the current study.

The current study utilized generalized linear mixed-effect models (GLMM), instead of ANOVA, mainly for the following reasons: (1) this practice can avoid the limitations of using ANOVA in the analysis of reading time data; (2) the data in the current study were skewed to the right with an absolute bound; (3) GLMMs make it possible to analyze the data at a fine-grained level, as all individual responses are entered into the model in the analysis; (4) in the model participants and items can be set as random effects, which takes into account the individual differences of participants and items.

Figure 3.5.

Histogram of learners' reading times in the critical region



Reading times in the critical region (Region 4), the noun after the numeral phrase, and the two spill-over regions (Region 5 and 6), the preposition and the noun after the critical noun, were included in the analysis. The dependent variable is the reading time.

The native group and the learner group were analyzed separately. The following sections introduce the two groups' results on each region.

General results

Descriptive statistics. Figure 3.6 shows the mean reading time for each region for the native group after data trimming; Figure 3.7 shows the mean reading time for the learner group. Region 4 was the critical region, the noun; Region 5 and 6 were the two spill-over regions. It can be seen from the two figures that, for both native speakers and learners, the reading time for the classifier omission condition on Region 3 was shorter than the other three conditions. This pattern is not surprising because Region 3 was shorter in the classifier omission condition, which was a one-character numeral or demonstrative, while in the other three conditions, Region 3 was composed of a numeral or a demonstrative with a classifier.

Figure 3.6.

Reading time for each region: the native group

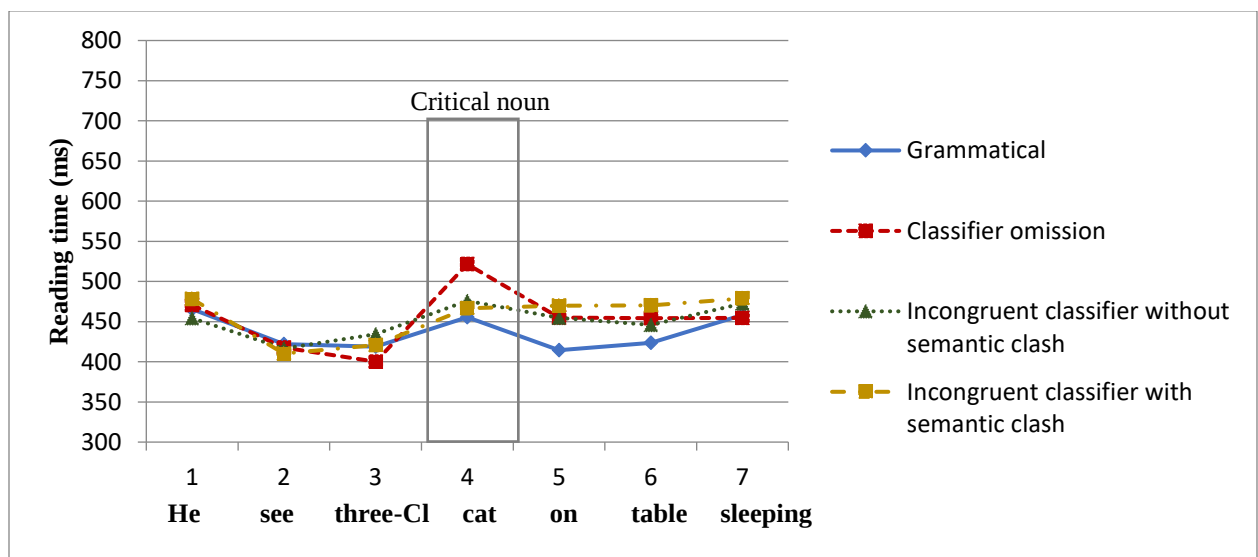
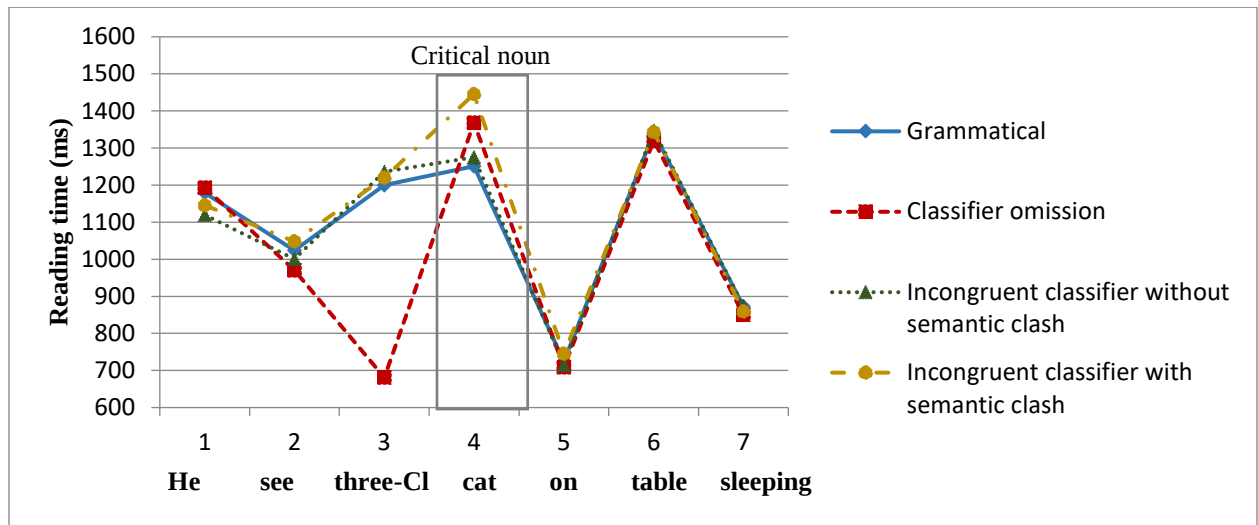


Figure 3.7.

Reading for each region: the learner group



The reading times of the critical regions, Region 4, and two spill-over regions, Region 5 and 6, were used for further analysis. Table 3.6 shows the mean reading times on these three regions across conditions, as well as standard deviation for the native group. Table 3.7 shows the same information for the learner group.

Table 3.6

Mean reading times for the native group

	Grammatical		Classifier omission		Incongruent classifier without semantic clash		Incongruent classifier with semantic clash	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Critical region	455.37	145.92	521.46	236.30	475.88	194.81	466.53	178.63

Table 3.6 (cont'd)

Spill-over region 1	414.41	108.12	455.05	114.33	454.75	141.48	469.67	154.49
Spill-over region 2	423.64	113.48	454.12	120.34	445.76	134.09	470.03	137.74

Table 3.7

Mean reading times for the learner group

	Grammatical		Classifier omission		Incongruent classifier without semantic clash		Incongruent classifier with semantic clash	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Critical region	1264.56	327.00	1374.88	490.58	1275.37	346.19	1445.45	540.53
Spill-over region 1	722.92	174.24	708.84	205.62	715.33	162.79	745.36	189.23
Spill-over region 2	1336.21	466.22	1316.59	456.18	1346.90	380.18	1342.85	491.70

Statistical analyses. To investigate whether native speakers of Chinese and English-speaking learners of Chinese are sensitive to different types of ungrammaticality regarding Chinese classifiers, the following model was used for analyses of natives' data and learners' data:

$$\text{Reading Time (RT)} \sim \text{Condition} + (1 \mid \text{Subject}) + (1 \mid \text{Item})$$

The independent variable, Condition, has four levels: (a) grammatical; (b) classifier omission; (c) incongruent classifiers without semantic clash; (d) incongruent classifiers

with semantic clash. The reference category was the grammatical condition. Random intercepts of Subject and Item were also included in the model.

The analyses were carried out using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2017). AIC and BIC were checked to decide the best-fitting model; lower AIC and BIC indicate better fit (Lo & Andrews, 2015). It was found that an Inverse Gaussian distribution with inverse link yields the lowest AIC and BIC. For instance, Table 3.8 shows the AIC and BIC for different models carried out on the critical region for the learner group.

Table 3.8

AIC and BIC indices of model fit

Distribution	Link function	AIC	BIC
Gamma	Identity	1364.1	1397.3
	Inverse	1334.0	1367.2
Inverse Gaussian	Identity	1287.0	1320.2
	Inverse	1263.1	1296.3

GLMM was carried out on the critical region, which was the noun after the classifier; the first spill-over region, which was the one-character preposition after the critical noun; and the second spill-over region, which was the two-character noun after the first spill-over region. The native group and the learner group were analyzed separately. For the independent variable, Condition, the grammatical condition (Condition a) was set as the reference category. The results of the native group for the critical region are shown in Table 3.9.

Table 3.9

Model results of the native group on the critical region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	2.42	0.14	17.16	<.001
Condition b	-0.23	0.06	-3.84	<.001
Condition c	-0.08	0.06	-1.37	.17
Condition d	-0.07	0.06	-1.22	.22

As an inverse link was used in the model, which reversed the +/- sign and hence the directionality of the effects. The native group's average reading time for the grammatical condition was 2.42 in reverse time units, which was 0.413s, or 413ms. Compared to the grammatical condition, for Condition b, the classifier omission condition, the average reading time elevated to 456ms ($1/(2.42-0.23)$); this reading time slowdown was significant ($t=-3.84, p<.001$). As for Condition c, in which an incongruent classifier from another semantic domain was used and there was no semantic clash between the classifier and the noun (for instance, the clothes classifier *jian* was used for pants, though a shape classifier is required), although the reading time elevated to 427ms compared to the grammatical condition, the slowdown did not reach a significant level ($t=-1.37, p=.17$). For condition d, in which an incongruent classifier from the same semantic domain was used and there was semantic clash between the classifier and the noun (for instance, the shape classifier for flat objects *zhang* was used for pants, though the long-shaped classifier *tiao* is required), the reading time slowdown was not significant either ($t=-1.22, p=.22$). The model results for the first spill-over region is shown in Table 3.10.

Table 3.10

Model results of the native group on the first spill-over region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	2.51	0.12	21.50	<.001
Condition b	-0.15	0.06	-2.60	.01
Condition c	-0.17	0.06	-3.13	.002
Condition d	-0.26	0.05	-4.66	<.001

On the first spill-over region, compared to the grammatical condition which had an average reading time of 398ms, the average reading time for Condition b, the classifier omission condition, was 424ms, which was significantly longer than the reading time in the grammatical condition ($t=-2.60$, $p=.01$). For Condition c, the incongruent classifier without semantic clash condition, the reading time was also significantly longer compared to the grammatical condition ($t=-3.13$, $p=.002$), with an average reading time of 427ms in this condition. Similar to Condition b and c, in Condition d, the incongruent classifier with semantic clash condition, a significant elevation in reading time was also present ($t=-4.66$, $p<.001$), the average reading time for this condition was 444ms. The model results for the second spill-over region are shown in Table 3.11.

On the second spill-over region, again all the ungrammatical conditions were compared with the grammatical condition. Compared to the grammatical condition, the native group had a longer reading time for the classifier omission condition, Condition b, and the reading time difference between these two conditions (405ms vs. 427ms) reached the significant level ($t=-2.39$, $p=.02$). For condition c, the reading time was not significantly longer ($t=-1.07$, $p=.28$). The reading time for Condition d was significantly longer compared to the grammatical condition in this region ($t=-3.96$, $p<.001$).

Table 3.11

Model results of the native group on the second spill-over region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	2.47	0.13	19.54	<.001
Condition b	-0.13	0.05	-2.39	.02
Condition c	-0.06	0.05	-1.07	.28
Condition d	-0.21	0.05	-3.96	<.001

To summarize, the native speaker group were sensitive to different types of ungrammaticality with regard to classifiers, which was suggested by the significantly longer reading times compared to the grammatical condition. The reading time difference was observed in either the critical region, or the spill-over regions, or both. In online processing of sentences with classifiers, when the classifier was omitted, or an incongruent classifier was used, no matter whether there was semantic clash between the classifier and the noun or not, it took the native speakers of Chinese longer to read the sentence components because of the unexpected structure.

GLMMs with an Inverse Gaussian distribution and an inverse link were also applied to the learner group's data. Again, the grammatical condition was set as the reference category. Model results on the critical region, the first spill-over region, and the second spill-over region were listed in Table 3.12, Table 3.13 and Table 3.14 respectively.

Table 3.12

Model results of the learner group on the critical region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.91	0.07	12.12	<.001
Condition b	-0.05	0.03	-1.48	.14
Condition c	-0.04	0.03	-1.16	.25
Condition d	-0.10	0.03	-3.25	.001

Table 3.13

Model results of the learner group on the first spill-over region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	1.48	0.07	19.90	<.001
Condition b	-0.01	0.05	-0.27	.79
Condition c	-0.03	0.05	-0.57	.57
Condition d	-0.05	0.05	-1.17	.24

Table 3.14

Model results of the learner group on the second spill-over region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.85	0.07	12.94	<.001
Condition b	-0.005	0.03	-0.15	.88
Condition c	0.02	0.03	0.73	.46
Condition d	0.006	0.03	0.20	.84

On the critical region, the region after the classifier position, the learners' average reading time for the grammatical condition was 0.91 in reverse time units, which was 1.099 seconds, or 1099 milliseconds (ms). For Condition b, when the classifier was

omitted, the average reading time elevated to 1163ms ($1/(0.91-0.05)$), but there was no significant difference between the reading speed of the grammatical condition and the classifier omission condition ($t=-1.48, p=.14$). For Condition c, when an incongruent classifier from another semantic domain was used and there was no semantic clash between the classifier and the noun, the average reading time was 1149ms, but again, the reading time difference between this condition and the grammatical condition was not significant ($t=-1.16, p=.25$). For Condition d, when an incongruent classifier from the same semantic domain was used and there was semantic clash between the classifier and the noun, the learner group showed a significantly longer reading time compared to the grammatical condition ($t=-3.25, p=.001$), with an average reading time of 1235ms.

On the first spill-over region, the learner group had an average reading time of 1.48 in reverse time units, which was 676ms for Condition a, the grammatical condition. When the classifier was omitted, which was the case for Condition b, there was no significant difference in reading times compared to the grammatical condition ($t=-0.27, p=.79$). When an incongruent classifier from a different semantic domain was used and there was no semantic clash between the classifier and the noun, which was the case for Condition c, there was no significant difference in reading time between this condition and the grammatical condition ($t=-0.57, p=.57$). In Condition d, where an incongruent classifier from the same semantic domain was used and there was semantic clash between the classifier and the noun, the reading time was not significantly different from the grammatical condition ($t=-1.17, p=.24$).

The results on the second spill-over region showed a similar pattern to those on the first spill-over region. The average reading time for the grammatical condition was 0.85

in reverse time units, which was 1176ms. Compared to the grammatical condition, there was no significant difference on reading time when the classifier was missing ($t=-0.15$, $p=.88$), when an incongruent classifier without semantic clash with the noun was used ($t=0.37$, $p=.46$), or when an incongruent classifier with semantic clash with the noun was used ($t=0.20$, $p=.84$).

In summary, English-speaking learners of Chinese showed sensitivity to specific types of violations in classifier use, which was reflected by the elevated reading time on the critical noun when an incongruent classifier that had semantic clash with the noun was used. However, when the classifier was omitted, or an incongruent classifier without semantic clash with the noun was used, the learners did not show significantly longer reading times, which suggested a lack of sensitivity to the violations. Therefore, the learner group showed different patterns in online processing of sentences containing classifiers compared to native speakers of Chinese, who were sensitive to all types of ungrammaticality with classifiers.

L2 proficiency and online comprehension of classifiers

Descriptive statistics. To investigate whether learners of different Chinese proficiency levels perform differently or not in their online comprehension of classifiers, I separately examined the performance of the groups based on their scores on the proficiency test and the offline classifier knowledge test. The subgroups based on proficiency were a higher-proficiency group ($N=17$) and a lower-proficiency group ($N=15$). Figure 3.8 shows the mean reading time for each region of the higher proficiency group, and Figure 3.9 shows the mean reading time of the lower proficiency group.

Figure 3.8.

Reading for each region: the higher-proficiency learner group

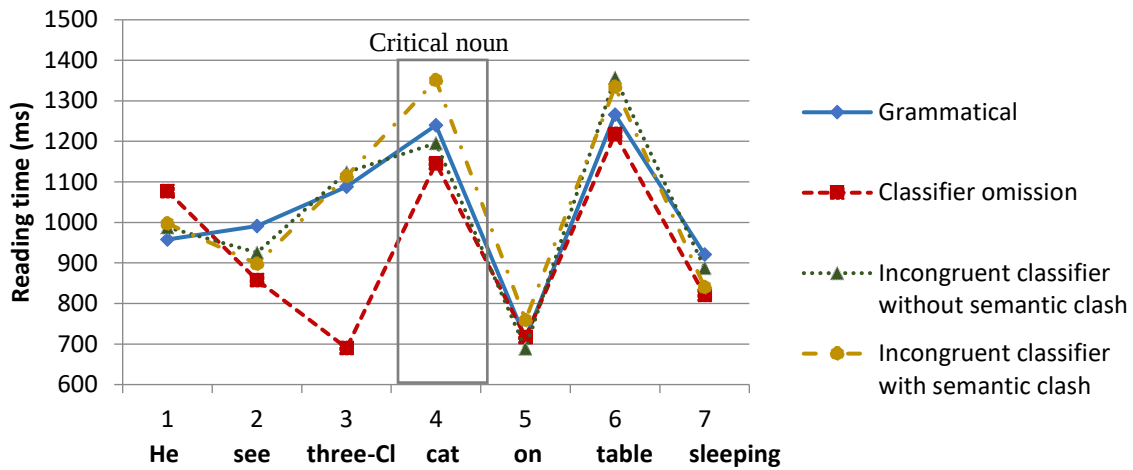
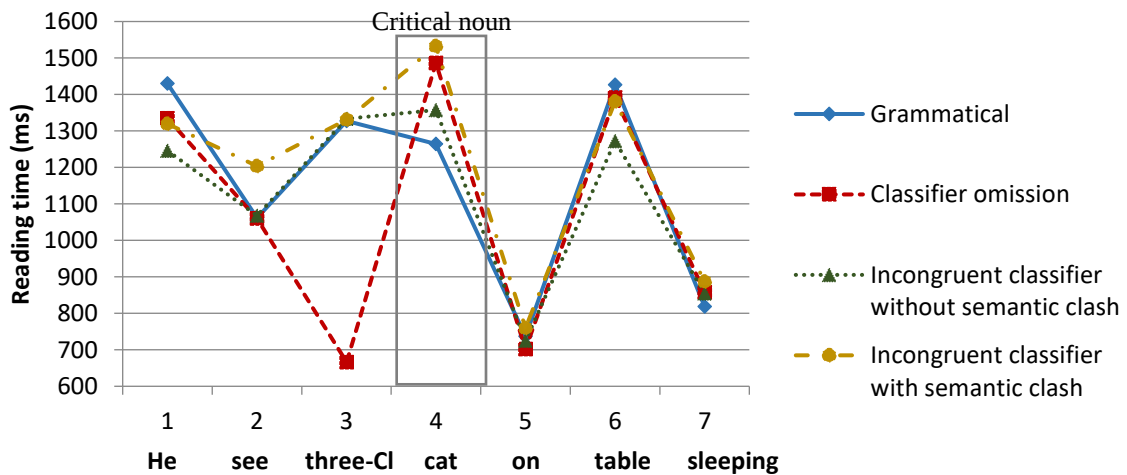


Figure 3.9.

Reading for each region: the lower-proficiency learner group



Region 4 was the critical region, the noun after the classifier; Region 5 and Region 6 were the two spill-over regions. Table 3.15 shows the mean reading times of these three regions and standard deviations of the higher-proficiency learner group; Table 3.16 shows the corresponding information of the lower-proficiency learner group.

Table 3.15

Mean reading times for the higher-proficiency learner group

	Grammatical		Classifier omission		Incongruent classifier without semantic clash		Incongruent classifier with semantic clash	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Critical region	1239.54	315.93	1145.84	378.27	1195.87	332.39	1351.43	572.55
Spill-over region 1	713.81	120.82	717.55	232.03	688.58	125.98	759.25	213.69
Spill-over region 2	1266.07	489.14	1218.29	431.06	1357.61	376.37	1334.53	512.21

Table 3.16

Mean reading times for the lower-proficiency learner group

	Grammatical		Classifier omission		Incongruent classifier without semantic clash		Incongruent classifier with semantic clash	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Critical region	1264.53	321.37	1486.27	469.48	1357.37	359.10	1532.78	489.69
Spill-over region 1	736.72	226.08	701.52	177.15	724.48	194.09	759.93	197.61
Spill-over region 2	1426.44	446.09	1391.56	484.12	1272.16	340.09	1380.58	477.02

It seems from the descriptive statistics that both the higher-proficiency learner group and the lower-proficiency group took longer to read the noun when it was preceded by an incongruent classifier from the same semantic domain as the correct classifier. This pattern was similar to that of the whole learner group. Statistical analyses were carried out to test whether the reading time differences had reached a significant level.

Statistical analyses. Generalized liner mixed effect models were used to analyze the data of the learner group. I added Proficiency level as a new independent variable in addition to Condition; therefore the model used for the analyses is:

$$\text{Reading Time (RT)} \sim \text{Condition} + \text{Proficiency level} + (1 \mid \text{Subject}) + (1 \mid \text{Item})$$

The dependent variable was the reading times of the learner group. For the fixed factor Condition, the grammatical condition was set as the reference category. As for Proficiency level, higher proficiency was set as the reference category. Random intercept of subject and item were still included in the model. We were also interested in the interaction between Condition and Proficiency level, which reflects whether learners treat various conditions differently. The model with interactions is:

$$\text{Reading Time (RT)} \sim \text{Condition} * \text{Proficiency level} + (1 \mid \text{Subject}) + (1 \mid \text{Item})$$

Whenever the interaction was of interest, the full models were reported after the simplified models to show more information regarding the interactions.

On the critical region, the noun after the classifier, the simplified model without interactions was used for the analysis on the critical region first. The model results of this region are shown in Table 3.17.

Table 3.17

Model results on the critical region with proficiency levels

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.96	0.09	11.12	<.001
Condition b	-0.05	0.03	-1.47	.14
Condition c	-0.04	0.03	-1.16	.25
Condition d	-0.10	0.03	-3.24	.001
Proficiency level: Low	-0.11	0.10	-1.16	.25

It can be seen from the model results that, similar to previous results of models without Proficiency level as the second independent variable, the learner group read the critical noun significantly more slowly when it was preceded by an incongruent classifier with semantic clash with the noun; other ungrammatical conditions including both classifier omission and incongruent classifier without semantic clash with the noun did not slow down the reading significantly. As for the influence of proficiency level, it was found that the lower-proficiency learner group read the critical noun more slowly than the higher-proficiency group. For instance, for Condition a, the higher-proficiency group had an average reading time of 1042ms (1/0.96s), while the lower-proficiency group had an average reading time of 1176ms (1/(0.96-0.11) s). However, this reading time difference was not statistically significant ($t=-1.16$, $p=.25$). In other words, learners' proficiency was not a significant predictor of the reading times on the critical region; higher-proficiency learners did not read the critical region significantly more quickly than lower-proficiency learners across all conditions.

I was particularly interested in whether proficiency level affected learners' sensitivity to different types of violation in classifier use. Therefore, the model results of

the full model with the interaction between Condition and Proficiency level are reported in Table 3.18.

Table 3.18

Full model results on the critical region with proficiency level

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.94	0.09	10.65	<.001
Condition b	-0.004	0.05	0.09	.92
Condition c	-0.02	0.05	-0.50	.62
Condition d	-0.10	0.04	-2.26	.02
Proficiency: Lower	-0.08	0.11	-0.74	.46
Condition b*Proficiency: Lower	-0.10	0.07	-1.60	.11
Condition c*Proficiency: Lower	-0.03	0.07	-0.46	.65
Condition d*Proficiency: Lower	-0.006	0.06	-0.09	.93

As shown above, there was no significant interaction between Proficiency level and any level of Condition. The lack of significant interaction between Condition and Proficiency level may indicate that higher-proficiency learner group and the lower-proficiency group did not perform in different ways with regard to different types of ungrammatical use of Mandarin classifiers; both groups only showed significant reading time slowdowns when an incongruent classifier from the same semantic domain was used and there was semantic clash between the classifier and the noun.

It can be seen from the descriptive statistics that the lower-proficiency group showed longer reading times for Condition b, the classifier omission condition (as shown in Table 3.16). The reading time difference between the grammatical condition and Condition b was in the reverse direction for the higher proficiency learner group (as shown in Table

3.1). Although no significant interaction has been observed, what is worth noticing in the full model results is that the slope for the interaction between Condition b and proficiency level is quite large, and is even comparable to that of the fixed effect of Condition d. The lack of significance might also be due to the large variance within the data, which was confirmed by the large standard error in the model results.

For the first spill-over region, the one-character prepositional word after the critical noun, the results of the simplified model are shown in Table 3.19 below. The full model was also carried out; however, no significant or marginal significant interaction was observed.

Table 3.19

Model results on first spill-over region with proficiency levels

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	1.49	0.09	15.98	<.001
Condition b	-0.01	0.05	-0.27	.79
Condition c	-0.03	0.05	-0.57	.57
Condition d	-0.05	0.05	-1.17	.24
Proficiency level: Low	-0.02	0.12	-0.13	.90

The model results show that on the first spill over region, the learner group did not read any ungrammatical conditions significantly more slowly than the grammatical condition, which again was similar to the results of the models without Proficiency level as another independent variable. What was also revealed in the analysis was that in this spill-over region, learners' proficiency level did not affect reading times significantly

either ($t=-0.13, p=.90$). There was no significant difference in reading speed between higher-proficiency learners and lower-proficiency learners.

For the second spill over region, the results of the simplified model are shown in Table 3.20 below. Again, the full model was carried out, but no significant interaction or noticeable slope of interaction was observed.

Table 3.20

Model results on the second spill-over region with proficiency levels

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.89	0.08	11.67	<.001
Condition b	-0.005	0.03	-0.15	.88
Condition c	0.02	0.03	0.72	.47
Condition d	0.006	0.03	0.20	.84
Proficiency level: Low	-0.08	0.08	-0.93	.35

For the second spill-over region, there was no significant difference between the reading times of the grammatical condition and any of the three ungrammatical conditions. Proficiency level was not a significant predictor of reading times of this region; there was no significant difference between the lower-proficiency learner group and the higher-proficiency learner group ($t=-0.93, p=.35$).

In summary, Proficiency level was not a significant predictor of reading times on the critical noun and the two words after the critical region; in another word, with higher levels of proficiency, although English-speaking learners of Chinese read these regions slightly more quickly (as it was shown by the negative estimate for the main effect of proficiency level in Table 3.17, Table 3.19, & Table 3.20), none of the differences in

reading times had reached the significant level. As suggested by the lack of significant interactions, Condition did not interact significantly with Proficiency level. These results indicate that learners of different proficiency levels in Chinese may not treat grammatical use of classifiers and different types of ungrammatical use of classifiers fundamentally differently. However, the large slope of the interaction between condition b and Proficiency level in the full model results on the critical region gives a suggestion that with higher L2 proficiency, the sensitivity to ungrammatical omission of classifiers was lower to some extent.

Knowledge of classifiers and online comprehension of classifiers

Descriptive statistics. It is possible that the performance of the English-speaking learners of Chinese on the online comprehension task was affected by their knowledge of Mandarin classifiers, as is the case with knowledge of grammatical gender for performance on grammatical gender tasks (Hopp, 2013). Along these lines, I divided the learner group into two sub-groups based on their score on the offline cloze task targeting their classifier knowledge. These were the higher-performance group (N=17), and the lower-performance group (N=15). Figure 3.10 shows the mean reading time of the higher-performance learner group on each region; Figure 3.11 shows the mean reading times of the lower-performance group.

Figure 3.10.

Reading for each region: the higher-performance learner group

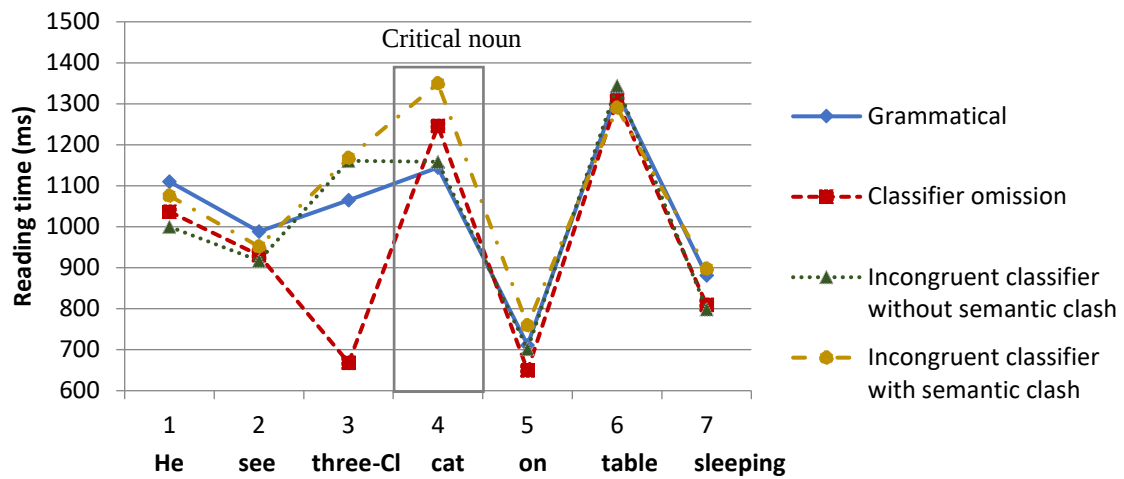
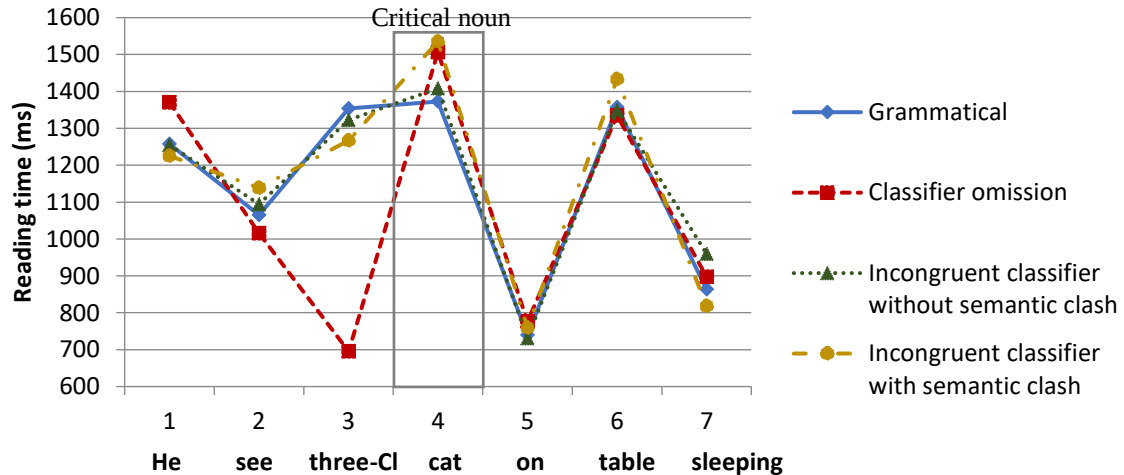


Figure 3.11.

Reading for each region: the lower-performance learner group



Region 4 was the critical region, Region 5 and 6 were the two spill-over regions. It can be seen from the figures that on the critical region, both learner groups showed elevated reading times for Condition d, in which an incongruent classifier from the same

semantic domain was used. In addition, both groups had longer reading times for Condition b, in which the classifier was omitted. Statistical analyses were carried out to investigate whether such reading time slowdown reached significant level, and the results are presented below.

Before moving on to the results of statistical analyses, more descriptive results are shown in the following tables. Table 3.21 shows the mean reading times and standard deviations of the higher-performance group on the critical region and the two spill-over regions. Table 3.22 shows this information for the lower-performance group.

Table 3.21

Mean reading times for the higher-performance learner group

	Grammatical		Classifier omission		Incongruent classifier without semantic clash		Incongruent classifier with semantic clash	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Critical region	1144.16	258.96	1245.95	345.33	1158.18	315.42	1349.66	473.11
Spill-over region 1	711.48	165.68	649.39	179.18	701.47	171.95	759.53	204.50
Spill-over region 2	1327.42	457.39	1307.67	427.30	1344.50	416.78	1290.45	470.44

Table 3.22

Mean reading times for the lower-performance learner group

	Grammatical		Classifier omission		Incongruent classifier without semantic clash		Incongruent classifier with semantic clash	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Critical region	1372.63	333.69	1506.53	591.25	1408.18	340.73	1534.90	604.34
Spill-over region 1	739.36	190.48	776.22	218.58	731.04	156.17	759.60	209.09
Spill-over region 2	1356.91	497.48	1334.53	516.76	1349.62	348.57	1433.48	516.40

Statistical analyses. To investigate whether learners' knowledge of Mandarin classifiers is related to their online comprehension, generalized linear mixed models were used to examine whether learners with different level of offline knowledge of classifiers showed different degrees of sensitivity to ungrammaticality in classifier use. Two independent variables were included in the model: Condition and Cloze level. Condition had four levels: Condition a, the grammatical condition; Condition b, the classifier omission condition; Condition c, the incongruent classifier condition in which an incongruent classifier from another semantic domain was used and there was no semantic clash between the classifier and the noun; Condition d, the incongruent classifier condition in which an incongruent classifier from another semantic domain was used and there was semantic clash between the classifier and the noun. Condition a, the grammatical category was set as the reference category. The other independent variable, Cloze level, had two levels, higher performance and lower performance. Higher

performance was set as the reference category. The dependent variable was the reading time on critical region and two spill-over regions. Random intercepts by subject and item were also included. An Inverse Gaussian distribution with inverse link was specified in the model. The models used are shown below:

Simplified model: Reading Time (RT) ~ Condition + Cloze level + (1 | Subject) + (1 | Item)

Full model: Reading Time (RT) ~ Condition * Cloze level + (1 | Subject) + (1 | Item)

Both a simplified model and full model were used for analyses. In cases where an interaction was of interest, both the simplified model and the full model were also reported.

For the critical region, the results of the simplified model are shown in Table 3.23. The results of the full model are also reported in table 3.24 to show information regarding the interactions.

Table 3.23

Model results on the critical region with cloze test levels

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.97	0.08	11.71	<.001
Condition b	-0.05	0.03	-1.47	.14
Condition c	-0.03	0.03	-1.17	.24
Condition d	-0.10	0.03	-3.25	.001
Cloze test level: Low	-0.15	0.09	-1.62	.11

There was a significant fixed effect of Condition d, suggesting that English-speaking learners of Chinese were sensitive to the incongruent classifier when there was semantic clash between the classifier and the noun ($t=-3.25$, $p=.001$). The average reading time for the critical noun was 1031ms in Condition a and 1149ms in Condition d. Although the lower performance group on the cloze test had longer mean reading time than the higher performance group (1220ms vs. 1031ms for Condition a), offline knowledge of Mandarin classifiers was not a significant predictor of the reading time of the critical noun ($t=-1.62$, $p=.11$).

The full model results showed a similar pattern to the simplified model. Learners of Chinese read Condition d significantly more slowly than condition a, the grammatical condition ($t=-2.73$, $p=.006$), which again confirmed that learners were sensitive to the incongruent classifier when there was semantic clash between the classifier and the noun. Overall knowledge of classifier was not a significant predictor of learners' reading speed on the critical noun. The lack of a significant interaction suggests that learners with different levels of knowledge of classifiers did not perform differently on any of the four conditions.

Table 3.24

Full model results on the critical region with cloze test levels

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.98	0.09	11.43	<.001
Condition b	-0.03	0.05	-0.68	.50
Condition c	-0.05	0.04	-1.04	.30
Condition d	-0.12	0.04	-2.73	.006

Table 3.24 (cont'd)

Cloze test level: Low	-0.16	0.10	-1.56	.12
Condition b*Cloze level: Low	-0.03	0.07	-0.46	.64
Condition c*Cloze level: Low	0.02	0.07	0.31	.76
Condition d*Cloze level: Low	0.04	0.06	0.57	.57

The overall score on the cloze test showed participants' knowledge of the target classifiers. It was found that this knowledge did not affect learners' sensitivity to classifier omission or incongruent classifiers before the nouns. However, using the overall score on the cloze test might not be the best way to examine this question. That is, learners with the same score may be able to provide correct classifiers to totally different nouns.

To investigate whether knowledge of individual classifier and noun pairs affects online comprehension, I divided all the observations from the 32 English-speaking learners of Chinese in the self-paced reading task into two sets based on whether a correct classifier was provided in the offline classifier knowledge task. In all of the 847 responses analyzed, 468 of them were categorized as classifier-consistent responses, meaning that in the cloze test a consistent classifier was provided for the noun; 367 of them were categorized as classifier-inconsistent responses, meaning that an incongruent or no classifier was provided in the cloze test. In 12 cases, an acceptable classifier was provided in the cloze test, but the classifier was not the one used in the reading comprehension, thus they were also categorized as the classifier-inconsistent group. After the grouping, 55.25% of the observations fell into the classifier-consistent group, for which participants showed knowledge of the classifier in the offline task; 44.75% of

the observations fell into the classifier-consistent group, for which participants failed to provide correct classifiers in the offline task, or the classifier provided was not the one used in the self-paced reading task.

The two sets of data were analyzed separately using generalized linear regression models, with an Inverse Gaussian distribution and an inverse link specified. Random intercepts by participant and item were also included. The model used is shown below:

$$\text{Reading Time (RT)} \sim \text{Condition} + (1 \mid \text{Subject}) + (1 \mid \text{Item})$$

The dependent variable was the reading time on the critical region and the two spill-over regions. The independent variable was the four conditions, and Condition a, the grammatical condition, served as the reference category. The results for the classifier-consistent group on the critical region, the first spill-over region, and the second spill-over region are shown in Table 3.25, Table 3.26, and Table 3.27 respectively.

Table 3.25

Model results of the classifier-consistent group on the critical region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.94	0.08	11.69	<.001
Condition b	-0.06	0.04	-1.29	.20
Condition c	-0.08	0.04	-1.74	.08
Condition d	-0.15	0.04	-3.48	<.001

Table 3.26

Model results of the classifier-consistent group on the first spill-over region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	1.52	0.08	18.08	<.001
Condition b	0.10	0.06	1.57	.11
Condition c	-0.03	0.06	-0.47	.64
Condition d	-0.12	0.06	-1.95	.05

Table 3.27

Model results of the classifier-consistent group on the second spill-over region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.90	0.07	12.81	<.001
Condition b	-0.03	0.04	-0.81	.42
Condition c	-0.03	0.04	-0.70	.49
Condition d	-0.03	0.04	-0.70	.49

On the critical region, the learners of Chinese had an average reading time of 0.94 in reverse time units, which was 1064ms, for the items that they provided the correct classifier for. When an incongruent classifier from the same semantic domain was used and there was semantic clash between the classifier and the noun, which was the case for Condition d, the learners read the critical noun significantly more slowly compared to the grammatical condition ($t=-3.48$, $p<.001$), with an average reading time of 1266ms. In addition, learners also showed a trend to being sensitive to incongruent classifiers without semantic clash with the noun, which was suggested by the marginally significant longer reading time for Condition c ($t=-1.47$, $p=.08$). There was no significant reading time slowdown between the grammatical condition and the classifier omission condition.

On the first spill-over region, the learner group had an average reading time of 658ms for the grammatical condition. They slowed down significantly on this region when an incongruent classifier that had semantic clash with the noun was used ($t=-1.95$, $p=.05$); the average reading time for Condition d was 714ms. There was no significant difference between the reading time of Condition a, and Condition b and c on this region.

On the second spill-over region, the average reading time for the grammatical condition was 0.90 in reverse time units, which was 1111ms. The participants did not slow down significantly on this region when the classifier was omitted ($t=-0.81$, $p=.42$), an incongruent classifier without semantic clash with the noun was used ($t=-0.70$, $p=.49$), or an incongruent classifier with semantic clash with the noun was used ($t=-0.70$, $p=.49$).

The results on the critical region and the first spill-over region showed that, for items for which the participants knew the correct classifier in the offline cloze task, English-speaking learners of Chinese were sensitive to the violations in which an incongruent classifier was used and there was semantic clash between the classifier and the noun in online comprehension, which was suggested by the significant reading time slowdown for Condition d on the critical noun and the prepositional word after the noun. Learners also showed a trend towards detecting the ungrammaticality when an incongruent classifier without semantic clash with the noun was used. On the other hand, they were not sensitive to omitted classifiers.

For the items for which participants were not able to provide a consistent classifier for the noun in the offline cloze task, the same generalized linear regression model was used for analysis. The model results on the critical region, the first spill-over region and

the second spill-over region are shown in Table 3.28, Table 3.29, and Table 3.30 respectively.

Table 3.28

Model results of the classifier-inconsistent group on the critical region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.92	0.09	10.62	<.001
Condition b	-0.02	0.05	-0.33	.74
Condition c	0.001	0.05	-0.04	.97
Condition d	-0.03	0.05	-0.71	.48

Table 3.29

Model results of the classifier-inconsistent group on the first spill-over region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	1.49	0.08	18.31	<.001
Condition b	-0.05	0.07	-0.75	.45
Condition c	-0.04	0.07	-0.54	.60
Condition d	0.01	0.07	0.20	.84

Table 3.30

Model results of the classifier-inconsistent group on the second spill-over region

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.86	0.08	10.80	<.001
Condition b	0.07	0.05	1.48	.14
Condition c	0.09	0.05	1.68	.09
Condition d	0.05	0.05	1.18	.24

The results showed that, for items that participants could not provide the congruent classifier for the nouns in the offline cloze task, they were not sensitive to any type of ungrammaticality in terms of classifier use, which was suggested by the lack of significant reading time slowdown in the three ungrammatical conditions, Condition b, c and d, both on the critical region, and the two spill-over regions. Although there was a marginally significant difference between Condition a and Condition c on the second spill-over region, the difference was in an reverse direction, meaning that learners read the region more quickly than for Condition c compared to the grammatical condition.

To summarize, in the investigation of the relationship between the knowledge of classifiers in the offline cloze task and the online comprehension of classifiers, it was found that knowledge of classifiers was not a significant predictor of sensitivity to classifier omission or incongruent classifiers. However, when I broke down all the data points into two subgroups, the classifier-consistent group for which participants were able to provide the congruent classifier for the noun, and the classifier-inconsistent group for which participants failed to provide an congruent classifier for the noun, or the classifier they provided was not the one used in the online comprehension task, it was found that when participants provided a congruent classifier for nouns in the offline task, they were sensitive to incongruent classifiers when there was semantic clash between the classifier and the noun, and they showed a trend towards being sensitive to incongruent classifiers when there was no semantic clash between the classifier and the noun, but they were not sensitive to classifier omission. For the nouns that participants did not know the congruent classifier or the classifier they knew was not the one used in the online

comprehension task, L2 participants did not show sensitivity to any type of ungrammaticality regarding classifiers.

Lexical retrieval and online comprehension of classifiers

In the lexical decision task modeled on LexTALE, reaction times to make decisions about each word, which indicates lexical retrieval speed, were recorded. To investigate the role of lexical retrieval speed in online comprehension of classifier-noun pairs, generalized linear mixed models were used for data analysis, with the retrieval speed of the nouns included as an independent variable in addition to Condition. The dependent variable was the reading time on the critical region and the two spill-over regions. Random intercepts by subject and item were also included. An Inverse Gaussian distribution with inverse link was specified in the model. The models used are shown below:

Simplified model: Reading Time $\sim$ Condition + Retrieval time + (1 | Subject) + (1 | Item)

Full model: Reading Time $\sim$ Condition * Retrieval time + (1 | Subject) + (1 | Item)

The new independent variable, Retrieval speed, was a continuous variable. The data of the learner group as well as the native group were analyzed separately using the same models. Extreme reaction times beyond 2.5 SD of the group mean in the lexical decision task were removed, which affected 1.52% of the natives' data and 1.07% of the learners' data. The reading time on the critical region and the two spill-over regions were used as the dependent variable in the analyses.

For the native group, the results of the simplified model are shown in Table 3.31. The interaction between Condition and Lexical retrieval speed reflects how lexical

retrieval affects sensitivity on classifier violations. The results of the full model are also shown in Table 3.32.

Table 3.31

Simplified model results of natives on the critical region with lexical retrieval time

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	2.39	0.09	11.43	<.001
Condition b	-0.20	0.06	-3.41	<.001
Condition c	-0.08	0.06	-1.24	.21
Condition d	-0.07	0.06	-1.08	.28
Lexical retrieval time	0.02	0.04	0.49	.62

As shown in the results, when the classifier was dropped, native speakers of Chinese read the noun significantly more slowly compared to the grammatical condition ($t=-3.41$, $p<.001$), suggesting that they were sensitive to the omitted classifier. Other types of ungrammatical use of classifiers did not lead to significant reading time slowdown on the critical noun. In addition, the retrieval speed of the noun in the lexical decision task was not a significant indicator of the reading speed of the noun in the self-paced reading task ($t=0.49$ $p=.62$).

Table 3.32

Full model results of natives on the critical region with lexical retrieval time

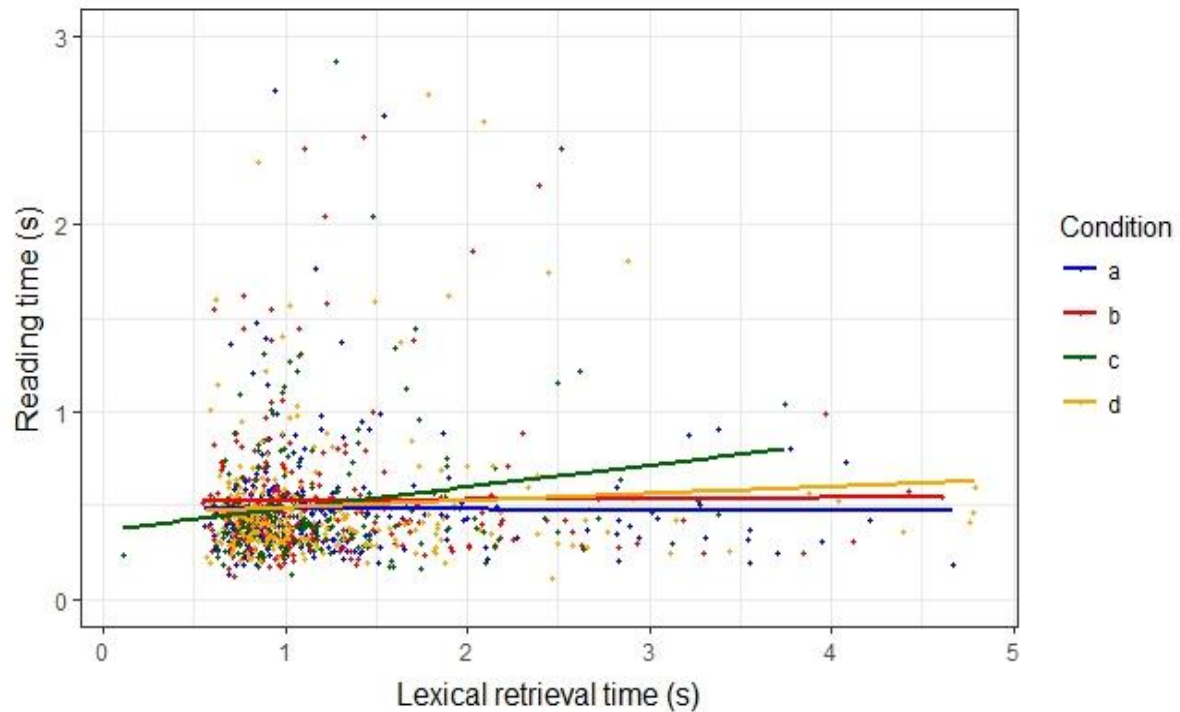
	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	2.29	0.16	14.11	<.001
Condition b	-0.19	0.13	-1.47	.14

Table 3.32 (cont'd)

Condition c	0.12	0.13	0.91	.36
Condition d	0.13	0.12	1.03	.31
Lexical retrieval time	0.10	0.07	1.52	.13
Condition b* Lexical retrieval time	-0.02	0.09	-0.19	.85
Condition c* Lexical retrieval time	-0.16	0.10	-1.65	.10
Condition d* Lexical retrieval time	-0.15	0.08	-1.80	.07

Figure 3.12.

Native reading time over lexical retrieval time on the critical region



The results of the full model showed that there was a marginally significant interaction between Condition c and Lexical retrieval speed ($t=-1.65$, $p=.10$), as well as Condition d and Lexical retrieval speed ($t=-1.80$, $p=.07$). Figure 3.12 shows the relationship between the native group's reading time on the critical region in the self-

paced reading task and lexical retrieval time in the lexical decision task with regression lines. It can be seen that for the grammatical condition, Condition a, the reading time on the critical noun did not increase with the growth of the lexical retrieval time, while for the two incongruent classifier conditions, Condition c and Condition d, participants read the critical noun more slowly if it took them longer to respond to the noun in the lexical decision task. In another word, the native participants tended to be more sensitive to incongruent classifiers when it took them longer to react to the noun in the lexical decision task, an unexpected result. But note that the lack of significant interaction between the reading time of the critical noun in the classifier omission and lexical retrieval time did not indicate that natives were not sensitive to classifier omission; their sensitivity to this ungrammaticality was independent of lexical retrieval speed.

For the first spill-over region, the results of the simplified model are shown in Table 3.33. Again, the full model was also used for analysis, and no significant interaction was detected in the results.

On the first spill-over region, the native speakers demonstrated significant reading time slowdowns for the three ungrammatical conditions, indicating that they were sensitive to all types of ungrammaticality with regard to classifiers. However, the lexical retrieval speed was not a significant predictor of reading time on this region ($t=0.33$, $p=.74$).

Table 3.33

Simplified model results of natives on the first spill-over region with lexical retrieval time

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	2.49	0.12	20.25	<.001
Condition b	-0.14	0.06	-2.43	.02
Condition c	-0.18	0.06	-3.16	.002
Condition d	-0.25	0.06	-4.50	<.001
Lexical retrieval time	0.01	0.03	0.33	.74

For the second spill-over region, the results of the simplified model are shown in Table 3.34. The full model was also carried out because of the importance of the interactions, but no significant interaction was spotted.

Table 3.34

Simplified model results of natives on the second spill-over region with lexical retrieval time

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	2.47	0.13	18.63	<.001
Condition b	-0.12	0.05	-2.25	.02
Condition c	-0.05	0.05	-1.00	.32
Condition d	-0.20	0.05	-3.80	<.001
Lexical retrieval time	0.01	0.03	0.27	.79

As shown in the model results, the native group read the second spill-over region significantly more slowly when the classifier was dropped ($t=-2.25$, $p=.02$), or when an incongruent classifier was used and there was semantic clash between the classifier and

the noun ($t=-3.80, p<.001$), compared to the grammatical condition. Again, the time participants spent to retrieve the corresponding noun in the lexical decision task was not a significant predictor of the reading time on this region ($t=0.27, p=.79$).

As for the learner group, on the critical region, the results of the simplified model are shown in Table 3.35. As shown in the model results, the learners read the noun significantly more slowly when an incongruent classifier with semantic clash with the noun was used ($t=-3.14, p=.002$); they were not sensitive to other types of ungrammaticality in classifier use. Meanwhile, the lexical retrieval time was a significant predictor of the reading time of the noun ($t=-2.01, p=.04$) in that the reading time increased with the lexical retrieval time.

Table 3.35

Simplified model results of learners on the critical region with lexical retrieval time

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.95	0.08	12.59	<.001
Condition b	-0.05	0.03	-1.59	.11
Condition c	-0.04	0.03	-1.29	.20
Condition d	-0.10	0.03	-3.14	.002
Lexical retrieval time	-0.02	0.01	-2.01	.04

The full model was also carried out; however, no significant interaction between the two independent variables was detected, suggesting that on this region, learners' lexical retrieval speed did not affect their reading time of different conditions in different ways.

On the first spill-over region, the results of the simplified model are shown in Table 3.36. The full model was also carried out because of the rich information interaction

could provide. However, no significant or approaching significant interaction was detected, and none of the interactions had a noticeably large slope.

Table 3.36

Simplified model results of learners on the first spill-over region with lexical retrieval time

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	1.56	0.08	19.89	<.001
Condition b	0.02	0.05	0.36	.72
Condition c	-0.03	0.05	-0.58	.56
Condition d	-0.05	0.05	-1.18	.24
Lexical retrieval time	-0.03	0.01	-2.49	.01

On the first spill-over region, English-speaking learners of Chinese did not read any of the ungrammatical conditions significantly more slowly than the grammatical condition. On the other hand, the lexical retrieval speed was a significant predictor of the reading speed ($t=-2.49$, $p=.01$), and learners read this region more slowly with longer lexical retrieval times.

On the second spill-over region, the results of the simplified model are shown in Table 3.37. It was suggested that learners did not show any sensitivity to any type of ungrammaticality on this region, as there was no significant reading time slowdown for any of the ungrammatical conditions. The lexical retrieval speed did not affect the reading speed either ($t=-1.28$, $p=.20$).

Table 3.37

Simplified model results of learners on the second spill-over region with lexical retrieval time

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.83	0.07	11.98	<.001
Condition b	0.02	0.03	0.57	.57
Condition c	0.02	0.03	0.63	.53
Condition d	0.01	0.03	0.24	.81
Lexical retrieval time	0.01	0.01	1.28	.20

The full model was also carried out because the interaction was very informative in this case, and the results are shown in Table 3.38 below. A significant interaction between Condition b and Lexical retrieval time was detected ($t=1.98, p=.05$). In addition, a marginally significant interaction between Condition d and Lexical retrieval time was also present.

Table 3.38

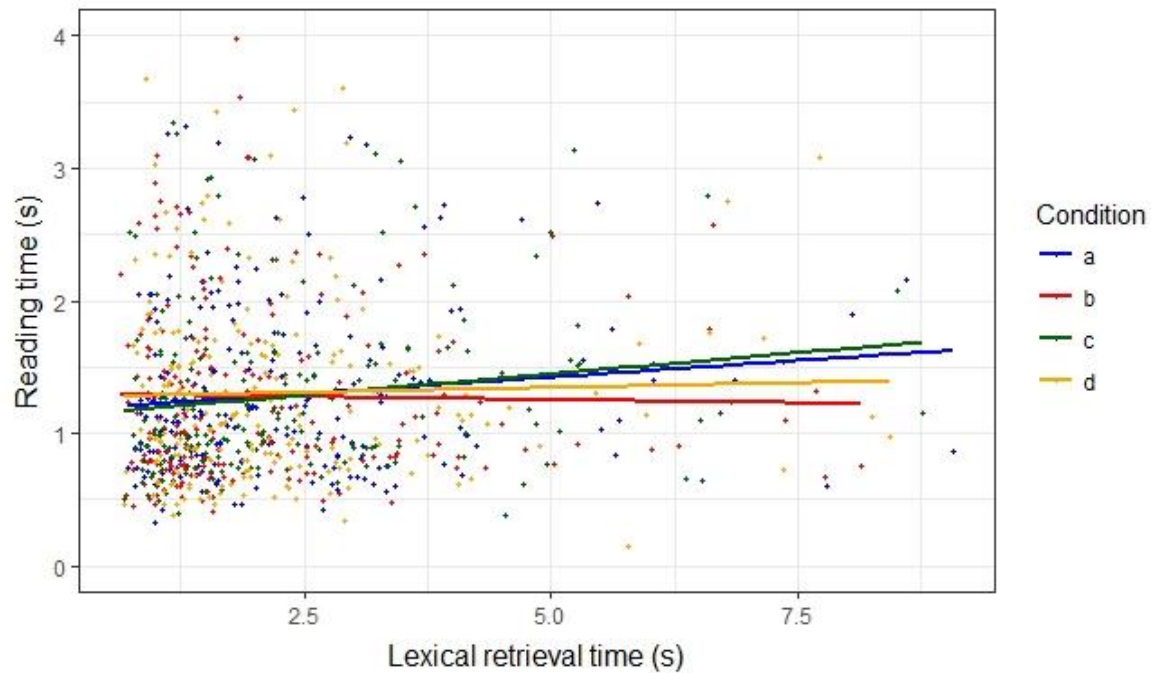
Full model results of learners on the second spillover region with lexical retrieval time

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	0.88	0.08	11.68	<.001
Condition b	-0.08	0.06	-1.36	.18
Condition c	0.001	0.06	-0.01	.99
Condition d	-0.08	0.06	-1.37	.17
Lexical retrieval time	-0.01	0.02	-0.63	.13
Condition b* Lexical retrieval time	0.04	0.02	1.98	.05
Condition c* Lexical retrieval time	0.01	0.02	0.43	.67
Condition d* Lexical retrieval time	0.04	0.02	1.77	.08

Figure 3.13 showed the relationship between the reading time on this spill-over region and the lexical retrieval time of the noun in the lexical decision task. The regression lines in the figure show the source of the interaction, which is that when participants retrieved the noun faster in the lexical decision task, they tended to be able to detect the ungrammaticality when the classifier was omitted (Condition b), or when an incongruent classifier with semantic clash with the noun was used (Condition d), as they tended to read these two conditions more slowly than the grammatical condition; however, when it took them longer to retrieve the noun, their sensitivity to those two types of ungrammaticality tended to decrease and even disappear, as it was shown by the shorter reading time of Condition b and Condition d compared to Condition a, the grammatical condition. The pattern suggested that the lexical representation of nouns may affect learners' sensitivity to classifier omission and incongruent classifier from the same semantic domain as the correct classifier.

Figure 3.13.

Learner reading time over lexical retrieval time on the second spill-over region



To summarize, for native speakers of Chinese, the time to retrieve nouns, did not affect overall reading time of either the noun or the two spill-over regions. However, when it took speakers longer to react to the noun in the lexical decision task, they tended to be more sensitive to incongruent classifiers, no matter whether there was semantic clash between the classifier and the noun or not. As for learners, when they reacted to the noun faster in the lexical decision task, they were able to read the noun, as well as the prepositional word right after the noun faster. What was more noticeable was that on the second spill-over region, it seemed that the lexical retrieval speed affected the reading speed of different conditions in different ways. When learners reacted to the noun faster in the lexical decision task, they tended to be sensitive to classifier omission and incongruent classifiers when there was semantic clash between the classifier and the

noun; with the increase of lexical retrieval time, this sensitivity decreased and even disappeared. The results suggested that higher quality noun representations helped with detection of specific types of ungrammaticality with regard to classifiers.

Summary of results

In the self-paced reading task targeting participants' online comprehension of Mandarin classifier-noun combinations, both a grammatical condition and three ungrammatical conditions were included. The type of ungrammaticality was also manipulated to include three types of incorrect use of classifiers: classifier omission, an incongruent classifier from a different semantic domain as the congruent classifier and without semantic clash with the noun (for instance, the clothes classifier *jian* was used for pants, though a shape classifier is required), and an incongruent classifier from the same semantic domain and with semantic clash with the noun (for instance, the shape classifier for flat objects *zhang* was used for pants, though the long-shaped classifier *tiao* is required). The aim of this task was to investigate whether English-speaking learners of Chinese were sensitive to different types of ungrammatical use of classifiers in online comprehension, and how their online comprehension was affected by factors including L2 proficiency, their offline knowledge of the classifiers, and their retrieval speed of the nouns. Native speakers of Chinese were also tested to serve as controls. The main findings of this task include:

- 1) Native speakers of Chinese demonstrated sensitivity to all types of ungrammatical use of Chinese classifiers, which was indicated by the elevated reading time on the critical region and the two spill-over regions for the three ungrammatical conditions

compared to the grammatical condition. On the other hand, English-speaking learners of Chinese were only sensitive to incongruent classifiers when there was semantic clash between the noun and the classifier, as they read the critical noun significantly more slowly for this condition compared to the grammatical condition.

2) Proficiency did not appear to be a significant predictor for L2 learners' performance on online comprehension of classifiers, both higher proficiency learners of Chinese and lower proficiency learners were able to detect the incongruent classifiers from the same semantic domain as the correct classifiers; however, the results suggested a trend that lower-proficiency learners might outperform higher-proficiency learners in terms of sensitivity to classifier omission, an unexpected result.

3) The overall score on the cloze task targeting offline knowledge of the classifiers used in the self-paced reading task was not a significant predictor of learners' sensitivity to different types ungrammatically used classifiers. However, for the nouns for which the learners managed to provide correct classifiers, learners were able to detect ungrammaticality when an incongruent classifier from the same semantic domain as the correct classifier and with semantic clash with the noun was used, and they tended to be sensitive to incongruent classifiers when there was no semantic clash between the classifier and the noun; on the other hand, in cases where they failed to provide the correct classifier for the noun in the offline task, they were not sensitive to any type of ungrammatical use of classifiers.

4) For native speakers of Chinese, retrieval speed of the nouns was not a significant predictor of reading times on the critical region or the two spill-over regions, but the observed significant interaction suggested that when it took them longer to react to the

noun in the lexical decision task, they were more sensitive to incongruent classifiers, no matter whether there was semantic clash between the classifier and the noun or not, an unexpected result.

English-speaking learners of Chinese read the critical noun and the prepositional word right after the noun faster if it took them less time to react to the noun in the lexical decision task; retrieval speed affected their sensitivity to different types of ungrammatical use of classifiers differently in that higher word retrieval speed predicted more sensitivity to classifier omission and incongruent classifiers when there was semantic clash between the classifier and the noun, and this sensitivity decreased and even disappeared with a longer reaction time to the noun in the lexical decision task.

CHAPTER 4

DISCUSSION AND CONCLUSION

Summary of results

Classifier production and online comprehension

The current study aims to investigate the process by which English-speaking learners of Chinese acquire Mandarin classifiers, with a focus on the source of the challenges L2 learners face in the process. An elicited production task, a self-paced reading task, a proficiency test, a classifier knowledge cloze task, and a word decision task were employed.

In the production task, native speakers of Chinese never omitted classifiers. L2 learners of Chinese omitted classifiers in very rare cases (0.5% of all their responses). The major difference between the native group and the learner group was that learners relied more on the general classifier, and specific classifiers were less frequent in their production, while the native group used specific classifiers more frequently than the general classifier. Such findings were consistent with Polio (1994). Native speakers also showed more variation of classifier use in that multiple classifiers were used for some nouns. The learner group used more questionable specific classifiers than the native group, and most of the ungrammatical choices could be attributed to the use of a classifier from a different semantic domain as the correct classifier, and the incorrect classifier matched the semantic features of the noun.

In the online comprehension task, when read sentences with classifier-noun combinations, native speakers of Chinese were sensitive to classifier omission, as well as incongruent classifiers, no matter whether there was semantic clash between the classifier

and the noun or not. As for the L2 group, they were not sensitive to classifier omission, and they only showed sensitivity to incongruent classifiers when there was semantic clash between the classifier and the noun, not to incongruent classifier without semantic clash with the noun.

The role of L2 proficiency, knowledge of classifiers, and lexical retrieval

The results of the current study suggested that at relatively higher levels of Chinese proficiency, L2 learners of Chinese behaved only slightly differently in both production and online comprehension compared to the lower proficiency learners. In production, the higher proficiency group showed a trend toward relying less on the general classifier and using more specific classifiers. In online comprehension, both the higher proficiency learners and the lower proficiency learners showed sensitivity to inconsistent classifiers from when semantic clash between the classifier and the noun was present. However, the lower proficiency group tended to be sensitive to classifier omission, while the higher proficiency learners did not.

L2 learners' knowledge of classifiers seemed to be an important contributing factor to their behavior pattern in production and online comprehension. The learner participants who performed better on the offline cloze task targeting classifier knowledge used significantly more specific classifiers and relied less on the general classifier in production. As for online comprehension, although the overall score on the classifier knowledge task was not a significant predictor of learners' performance, the learner group behaved differently on items that they were able to provide the correct classifier for in the offline task than on those for which they failed to do so. When L2 learners managed to provide the correct classifier for the noun in the offline classifier knowledge

task, they were sensitive to incongruent classifiers when there was semantic clash between the classifier and the noun, and they showed a trend toward being able to be sensitive to incongruent classifiers without semantic clash with the noun. When L2 learners were not able to provide the correct classifier for the noun in the offline task, they were not sensitive to any type of ungrammaticality with regard to classifiers.

Lexical retrieval speed was also found to be an influential factor in L2 learners' online comprehension of classifiers. Learners' faster retrieval of the noun predicted more sensitivity to classifier omission and incongruent classifiers when there was semantic clash between the classifier and the noun, and this sensitivity decreased and even disappeared with longer reaction time to the noun in the word decision task measuring lexical retrieval speed. As for native speakers of Chinese, a longer reading time of the noun after inconsistent classifiers was found with longer lexical retrieval time; natives speakers' sensitivity for classifier omission was not affected by the retrieval speed of the noun.

Answers to research questions

Research Question 1: Syntactic acquisition vs. lexical acquisition

Trying to provide insight on the broad issue of whether and how new features/functional categories can be acquired by L2 learners whose L1 lacks them, the current study focused on the source of difficulty observed in the L2 acquisition of Chinese classifiers. I investigated how English-speaking learners of Chinese use classifiers in production, as well as their sensitivity to different types of classifier errors in online comprehension. The characteristics of classifiers that inform the investigation include:

1) Classifier are separate morphemes that appear between determiners/numerals and nouns. They are not fused with nouns, or numerals/demonstratives; therefore, omission of classifiers is possible;

2) Noun categorization is involved in the classifier system, and L2 learners have to be able to use different types of information in order to show native-like performance on classifiers. First, semantic features of the classifier categories. Secondly, co-occurrence relationship between classifiers and nouns. It provides information on which semantic feature was picked by the classifier system for categorization, and consequently whether a taxonomy classifier or a shape classifier is required. For instance, *pants* are clothes, but the shape rather than the function is used for categorization, therefore the congruent classifier is the long shape classifier *tiao*. In addition, similar to grammatical gender assignment in which noun categorization is also involved, learners' familiarity with nouns also plays a role in the L2 acquisition of classifiers.

The first research question is: what is the sources of L2 learners' difficulty in classifier acquisition? I expected that if it is a representational deficit that underlies learners' difficulty in classifier acquisition, it is possible that L2 learners will omit classifiers in production and show limited sensitivity to classifier omission in online comprehension. On the other hand, if the source of difficulty in classifier acquisition lies at the lexical level, classifier omission should be rare in L2 learners' production, and learners should also be sensitive to classifier omission in online comprehension; meanwhile, they should have difficulty producing specific classifiers, and show limited sensitivity to incongruent classifiers. The lexical account could also be supported if there was a relationship between learners' behavior and lexical knowledge.

However, it seems that neither of the predictions was fully supported by the results, mainly because of the discrepancy between the learner group's performance in production and online comprehension with regard to classifier omission. The learner group rarely dropped classifiers in their production, but they did not show significant sensitivity to classifier omission in the online comprehension task. In addition, it was also suggested that lower proficiency learners may outperformed higher proficiency learners in terms of sensitivity to classifier omission tended to be sensitive to classifier omission. If English-speaking learners of Chinese have difficulty establishing the functional category of Cl in their L2 Chinese syntax, why did the learner group seldom omit classifiers in production? Why did lower proficiency learners tended to be sensitive to classifier omission in online comprehension? If the learners have the ability to establish the new functional category, then why L2 learners did not appear to be sensitive to classifier omission in online comprehension, especially the higher proficiency learners? There is no reason to speculate that the new functional category was established in L2 syntax in the early stage of acquisition, and then disappeared with higher L2 proficiency.

It might be possible that L2 learners of Chinese firstly regard the classifiers as bounded to the numeral/determiner (Polio, 1994), and the combination might have been used as a chunk before the noun. The closer connection between the classifier and numeral/determiner than between the classifier and the noun might be enhanced by the input, as in Chinese, the classifier and the noun can be separated by elements such as adjectives. For instance, 'a large cat' in Chinese is *one-Cl big cat*. A frequent behavior pattern in the production task was that, when L2 learners were not familiar with the

nouns, they would produce the numeral-general classifier chunk first, and paused before the noun. This behavior pattern in production also supports the speculation that classifiers were associated with numeral/determiners by some of the learners. If this is the case, lower-proficiency learners' sensitivity to classifier omission in online comprehension might be due to their sensitivity to missing elements in chunks, rather than the syntactic violation. Spinner (2013a) also argued that English-speaking learners of Swahili had difficulty segmenting the gender marker (which is realized as prefix) from the noun root at least at early stages of the acquisition. Therefore, their use of gender prefix in production and perception did not necessarily reflect the representation in their L2 syntax. Similarly, although the L2 learners in the current study seldom omit classifiers in production, it is possible that they do not really have classifiers as a functional category.

The learner group's performance in the production task can also be regarded as evidence for chunk storage. The L2 learners seldom omitted classifiers in production, which is an expected pattern if chunk storage is the case. In addition, they overused the general classifier, and the use of specific classifiers are limited. Specific classifiers are often used with familiar nouns such as *textbook*, *paper*, which presumably they have used together a lot. When the classifier is associated to numerals, the link between classifiers and nouns is weak, and the general classifier would be over used.

Being treated as a chunk, the numeral-classifier association could also be unpacked in later stages of acquisition (Myles, Mitchell, & Hooper, 1999). The transition from connecting the classifier with the numeral to connecting it with the noun is not beyond the realms of possibility, because it was found that faster retrieval speed of the noun in

the lexical decision task entailed sensitivity of classifier omission in learners' online comprehension, and the sensitivity decreased or even disappeared with slower retrieval speed. The results might be able to provide evidence that with sufficient lexical knowledge of the noun, the classifier could be associated with the noun under the syntactic projection of Cl in the L2 grammar.

The learner group's lack of sensitivity to classifier omission in online comprehension might be because even the higher proficiency group in the current study were not proficient enough to show native-like performance. The most of the L2 participants had no or limited experience living in the L2 environment, and most of their exposure to Chinese was in formal classroom setting. There is a possibility that the higher proficiency L2 group in the current study had not passed the transition stage from associating the classifier with the numeral to associating it with the noun. When the chunk was decomposed, and new connection had not been built, the sensitivity to missing element in the chunk would not be present anymore, and the sensitivity to syntactic violation had not fully developed. Further investigation with higher proficiency learners of Chinese is needed.

Grüter, Lew-Williams, and Fernald (2012) pointed out that children also rely on unanalyzed chunks of determiner-noun sequences in their acquisition of gender, which is a reflection of their learning through co-occurrence computation. The learning process makes the link between nouns and gender nodes remain strong at later stages of L1 development. Therefore, it seems that chunk storage contributes to L1 acquisition because of the co-occurrence computation of elements within chunks.

It seems contradictory here to argue that early L2 learners rely on chunks in classifier acquisition, as learners should have limited ability to do so as suggested in previous studies. However, there is reason to doubt whether the chunks children rely on in L1 acquisition and the chunks L2 learners use are the same. According to Gao (2010), L2 learners had difficulty noticing the semantic connection between classifiers and nouns. If L2 learners associate the classifier with numerals rather than nouns, as suggested by Polio (1994), it is reasonable to believe that the chunk L2ers rely on would not facilitate the acquisition of classifiers, L2ers' use of double classifiers in Polio (1994) was a suggestion of how the chunk affected classifier acquisition in a negative way.

Although previous studies suggested that children learn L1 features through association from the input, there are also studies arguing that native speakers break the complex words down in processing. For example, Cunnings (2017) discussed why inflection is difficult for L2 learners. He argued that L2 learners tended to store complex words as a whole, while natives decomposed them, making the morphosyntactic information less accessible to L2 learners. Limited access to the morphosyntactic information can constrain the building of strong links between lexical items and syntactic nodes. Therefore, in addition to the possibility that L2 learners of Chinese associate classifiers with numerals, it is also possible that L2 learners store the Cl phrase as a whole, a process that might be exacerbated by the fact that word boundaries are not overtly marked in Chinese. The lower-proficiency learner participants' tendency of being sensitive to classifier omission might thus be simply due to their sensitivity to missing part of the whole chunk stored in their memory as well. Less reliance on the whole chunk storage could result in less sensitivity to the missing element. This is also

consistent with the findings of Spinner (2013a), who suggested that English-speaking learners' performance on Swahili gender became worse once they started parsing words. In this sense, paradoxically, decreased accuracy in performance is a sign of acquisition taking place.

L2ers are likely to store morphologically complex words as a whole, and the less robust morphosyntactic features encoded in memory can result in more dependence on the strength of lexical representation (Cunnings, 2017). It is consistent with this proposal that that in the current study, L2 learners' lexical retrieval speed of the noun, which can be a reflection of the quality of lexical representation, or robustness of lexical knowledge, was associated with sensitivity to classifier omission and incorrect classifiers. With better quality of lexical representation, the morphosyntactic information carried on CI phrases becomes available to L2 learners.

The lexical learning account (Grüter, et al., 2012; Hopp, 2013) argued that the reason for L2ers' weak link between lexical and abstract syntactic nodes results from their lesser reliance on co-occurrence computations compared to natives acquiring the L1. Therefore, it is also possible that L2ers' non-native like performance on classifiers is due to less reliance on co-occurrence computations, rather than chunking. The current study also revealed L2 limitations in terms of access to lexical co-occurrence information. It was found that in online comprehension, the L2 participants were not able to detect the classifier error when semantic information was not useful. For instance, learners did not detect errors when a taxonomy classifier for clothes, *jian*, was used for *pants*, where a long shape classifier *tiao* was required. The incorrect classifier *jian* does not contradict with the semantic feature of the noun, as *pants* falls in the category of clothes. In this

case, familiarity with the semantic features of the classifier and the noun could not help to detect the error as no semantic conflict was involved. Co-occurrence relationships between the classifier and the noun was the only hint of whether a taxonomy classifier or a shape classifier is required. Only with extensive knowledge of the co-occurrence relationship could learners know whether a taxonomy or a shape classifier was required. It was also found that even with faster retrieval speed of the noun, there is no trend that the co-occurrence information could become more accessible to L2 learners in online comprehension, which further suggests that it is difficult for L2 learners to access the co-occurrence information.

In production, L2 learners' overreliance on semantic features and limited use of co-occurrence information was also an important source for their ungrammatical use of classifiers. The most frequent error type in the production task was to use a taxonomy classifier when a shape classifier was required, or to use a shape classifier when a taxonomy classifier was required. In this type of violation, no semantic conflict was involved (e.g., *pants* is a type of clothes, so semantically it is consistent with the taxonomy classifier *jian*, although the shape classifier *tiao* is required because of its shape). The lack of knowledge on the co-occurrence relationship between classifiers and nouns makes it difficult for L2 learners to decide which property of the noun was picked by the classifier system, hence when a taxonomy classifier should be used and when a shape classifier should be used.

Further evidence for the lexical learning account was that knowledge of classifiers facilitate L2 learners' error detection in online comprehension and more native-like use of classifiers in production. For items that L2 learners successfully provided the correct

classifier for in the classifier knowledge test, learners were more sensitive to classifier violations in the online comprehension task. L2 learners' higher score on the classifier knowledge test is also associated with more native-like performance in the production task. The findings suggested that native-like performance might be possible with sufficient lexical knowledge.

Also in line with the lexical learning account, rich semantic information encoded in classifiers could contribute to classifier acquisition. Evidence for this claim is that semantic information appears to be quite accessible to L2 learners of Chinese. L2 learners were sensitive to classifier errors when there is conflict between the semantic features of the classifier and the noun in online comprehension (e.g., the people classifier *wei* is used for *cat*, which requires an animal classifier *zhi*), and they seldom used semantically inconsistent classifiers for target nouns in production.

Admittedly, despite the potential explanation for the lower proficiency and higher proficiency L2 participants' performance in the online comprehension, it is worth further investigation as to why the higher proficiency L2 participants performed differently on classifier omission in production and online comprehension. There has been extensive discussion on the relationship between comprehension and production of L2 grammar. Previous studies have both suggested that comprehension can exceed production (Grüter, 2005) and production can exceed comprehension (Unsworth, 2007). The asymmetries between comprehension and production can occur either because comprehension and production are separate, or because the simultaneous development of comprehension and production is masked by factors such as pragmatic issues and/or existence of redundant forms (e.g., a listener can understand a sentence without processing the number marker)

(Spinner, 2013b). Although it was not directly investigated whether there is one grammatical system that is shared by both the productive and receptive system, or whether there are two grammatical systems, the current study can serve as another example of a discrepancy between production and comprehension, such as has been found before.

To conclude, the overall results suggest that although it is not easy for English-speaking learners of Chinese to establish the new functional category, classifiers, in their L2 syntax, the real constraint may lie at the lexical level. Sufficient lexical knowledge, which includes access to different types of lexical information, and fast lexical retrieval can contribute to native-like performance regarding Chinese classifiers.

Further remarks on the lexical account: Gradual L1/L2 difference

Hopp (2016b) investigated the interaction between verb frequency and difficulty in construction object clefts, and found that L2 performance on the high-frequency verbs resembled native reading of the sentences with low-frequency verbs. Word frequency was used as an indicator of lexical retrieval, because it has been found that slower lexical retrieval is associated with lower word frequency in the L2 lexicon (Tokowicz, 2015). In the current study, lexical retrieval speed was directly tested using the lexical decision task, and similar results to Hopp (2016b) were observed. The L2 group in this study was sensitive to ungrammatical use of classifiers when it took them less time to retrieve the noun, and a similar pattern was found in the L1 group with slower retrieval speed, although natives' sensitivity to classifier omission did not interact with the retrieval speed. In addition, learners' sensitivity to ungrammatical classifiers was also found to be related to their knowledge of classifiers. Altogether the results are consistent with

previous studies that have suggested a so-called gradual difference between L1 and L2 (Clahsen & Felser, 2017; Hopp, 2016b; Grüter et al., 2018).

The gradual L1/L2 difference accounts (Cunnings, 2017; Clahsen & Felser, 2017; Hopp, 2016b; Grüter et al., 2018), although differing in details, generally agree that the difference observed between L1 and L2 performance is not qualitative or categorical. One possible reason for the non-native like L2 performance is that lexical processing is the foundation for syntactic structure building; therefore, poor lexical representation constrains L2 performance (Hopp, 2016b). Another possibility is that the non-native like performance can be attributed to the lack of ability for efficient use of non-grammatical knowledge, which hinders the use of grammatical knowledge in processing tasks (Clahsen & Felser, 2017). A common ground of these proposals is that syntactic structure building is not unattainable for L2 learners. The current study provides evidence for the lexical account with evidence from the L2 acquisition of classifiers, which is new functional category for English-speaking learners of Chinese.

Research Question 2: Noun categorization in mental lexicon

The second research question is: Do learners organize their mental lexicon similarly to native speakers with regard to classifiers? Specifically, do learners rely on semantic information of classifier categories only? Is co-occurrence relationship between classifiers and nouns also accessible to learners for noun categorization?

It was found that native speakers of Chinese used both semantic features and co-occurrence relationships between the classifier and the noun for noun categorization. Evidence for this claim is that they never used classifiers that were not consistent with the semantic features of nouns, and they seldom used taxonomy classifiers when shape

classifiers were required, or used shape classifier when taxonomy classifiers were required. Only the co-occurrence relationship can provide cues on which semantic feature is relevant for noun categorization. Different from native speakers, English-speaking learners of Chinese over-relied on semantic information, and the co-occurrence information was less accessible to them, resulting in their non-targetlike categorization of nouns. They were not sensitive to inconsistent classifiers when no semantic conflict was involved in online comprehension, and they showed difficulty deciding whether a taxonomy classifier or a shape classifier should be used for a specific noun in production. Even with higher Chinese proficiency, there was no cue that the co-occurrence information became more accessible to L2 learners, as higher proficiency learners did not show any tendency of being able to be sensitive to classifier violations when a taxonomy classifier was used to replace a shape classifier, or vice versa.

L2 learners' limited ability to access the classifier-noun co-occurrence information might be related to the weak connections between classifiers and nouns in their lexical representation. As discussed before, English speaking learners of Chinese may tend to associate classifiers with numerals rather than nouns. If this is the case, the co-occurrence relationship between classifiers and nouns is not prominent in the L2 processing of classifiers. Another possibility is that although L2 learners use Cl phrases including numerals, classifiers and nouns as chunks, they have difficulty decomposing the chunks and accessing the morphosyntactic information encoded on the chunks. In this case, the functional projection of Cl cannot be reinforced with the co-occurrence relationship between classifiers and nouns, which may affect the native-like categorization of nouns as denoted by classifiers. There is also possibility that even the

higher proficiency L2 participants in the current study were not proficient enough to be able to access the co-occurrence information, but at least the lack of tendency to improve suggests that it is difficult for L2 learners to learn to use the information.

Research Question 3: L2 proficiency and classifier acquisition

The third research question is: How does the classifier system develop overtime with L2 Chinese proficiency? Do L2 learners of different Chinese proficiency levels use classifier differently in production and comprehension? Does their performance on classifiers improve with L2 proficiency?

In the current study, Chinese proficiency was not found to be a significant contributor to native-like performance regarding classifiers in production and online comprehension. In the production task, English-speaking learners of Chinese with higher L2 proficiency used classifiers in a slightly more native-like way compared to lower proficiency learners. In the online comprehension task, higher proficiency learners did not outperform their lower proficiency counterparts in that they were not more sensitive to ungrammatical use of classifiers, although as it was discussed, the non-native like performance might be a sign of acquisition taking place. The results suggest that, rather than the overall L2 proficiency, L2 learners' lexical knowledge and their lexical retrieval ability play the crucial role in their performance with regard to classifiers. An implication of this finding is that lexical retrieval ability does not go hand in hand with L2 proficiency. While it is not the focus of the current study, further study targeting the relationship between L2 proficiency and lexical representation would be helpful.

Although classifiers carry semantic information, to use the general classifier *ge*, or even to omit classifiers, does not affect communication to a large extent. The redundancy

of specific classifiers in communication may have played a role in the limited improvement of the classifier system with L2 proficiency of Chinese. For instance, even very advanced learners of Chinese may over use the general classifier, because most of the times it is grammatical and sounds acceptable or natural to native speakers. Another possibility is that the L2 participants in the current study did not cover large enough L2 proficiency range. The effect of proficiency may emerge with more advanced L2 participants who have extensive exposure to Chinese in the target language environment.

Pedagogical implications

Aiming to explore the source of the difficulty L2 learners go through in the acquisition of Mandarin classifiers, the current study generally suggests that lexical retrieval and access to lexical information is an issue for L2 learners of Chinese. Although it was not directly investigated, it is quite possible that rich and emphasized input on classifier-noun combinations can be helpful to L2 learners, in the sense that lexical familiarity can be enforced in this way.

As it was suggested that L2 learners are not able to use lexical co-occurrence information to the same extent as native speakers (Grüter et al., 2012), the rich information encoded on classifiers can serve as a facilitator for the L2 acquisition of classifiers. At the end of the study, some of the participants reported that they were aware of the existence of semantic rules for classifier membership, although their knowledge of the rules varied. However, some participants, especially those who scored relatively low in the classifier knowledge test, did not know or use any of the rules, not consciously at least. For those participants who appeared to have some knowledge of the semantic rules for classifier membership, most of the times they appeared to be cautious to apply the

rules extensively to unfamiliar nouns. For instance, some participants knew that small animal nouns such as *cat* and *dog* take an animal classifier, they were not willing to extend the same classifier to *chicken*, *bird*, or *goat/sheep*. Their lack of confidence in the semantic rules may have constrained their application of the rule to a larger range of nouns, especially to non-prototype nouns in their perception.

Although the results of the current study suggested that semantic information is accessible to L2 learners, it was also found that not all L2 participants in the current study were aware of the semantic rules. Learners, especially lower proficiency learners might also associate classifiers to numerals rather than to nouns. The observed pattern suggests that more emphasis can be put on semantic information of classifiers in classroom teaching, encouraging students to learning specific classifiers through analogy (Myers, 2000). For instance, when a new classifier is introduced in class, a common practice is to introduce several prototypical nouns that are associated with the classifier. A further step can be taken to encourage students to discuss any other nouns that they think may also fall in the category. Such activity can help students to explore the boundary of classifier classes, and encourage their application of semantic rules in the use of classifiers.

Meanwhile, it should also be made clear to L2 learners that semantic information itself does not ensure native-like use of classifiers, as a long object may take a taxonomy classifier, and an animal or an apparel may take a shape classifier. Enhanced input emphasizing classifier-noun association can draw learners' attention to the co-occurrence information (Sharwood Smith, 1991), and help learners to notice what semantic feature of the noun is used by the classifier system for categorization. Language instructors can try to use as many specific classifiers as possible, and emphasize classifiers in their oral and

written input provided to students. In addition, teaching classifiers in Cl phrases rather than abstract associations (e.g., *cat* goes with the classifier *zhi*) would be more helpful to L2 learners (Grüter et al., 2012).

In addition to emphasizing classifiers in instructors' input, it would also be helpful to expose learners of Chinese to classifiers in authentic language. Use of authentic materials in language classes benefits learners of different L2 levels including beginners (Zyzik & Polio, 2017). Incorporating authentic texts in Chinese language classes would offer learners opportunity to learn how specific classifiers and the general classifier are used in authentic language, which can help them to use classifiers in a more native-like way.

Concluding remarks

The current study investigated the L2 acquisition of Chinese classifiers, focusing on L2 learners' use of classifiers in production and online comprehension. The L2 participant group consisted of low intermediate to mid or high intermediate learners of Chinese, most with little or no experience living in the target language environment. A larger sample size and data from highly proficient learners of Chinese would be complementary to the present study. In addition, more work is needed to further investigate the underlying reason for the discrepancy between L2 learners' performance in production and online comprehension regarding classifier omission, including whether there is one grammatical system shared by both the productive and receptive system or two separate grammatical systems, and whether potential factors such as pragmatics contribute to the observed discrepancy. Studies targeting the relationship between production and comprehension can be insightful to this end.

As for the methodology, the online method, self-paced reading was used in the current study. The target language in the current study is Chinese, which has very different writing scripts from English. In this case, participants' reading fluency is a potential interfering factor in sentence processing, especially given that the L2 participants in this study were not highly proficient learners of Chinese. The L2 learner group read the sentences more slowly than the native group, which is not surprising. When look into the reading times of each region, it seems that on the region before the critical noun, which was composed of a numeral and a classifier (two characters in the grammatical condition and the two incongruent classifier conditions) or a numeral only (one character in the classifier omission condition), learners' reading times doubled when the region had two characters compared to one character, while the native group did not show such noticeable reading time difference. The length effect could be a suggestion of low reading fluency of the learner group in the current study. However, it is worth mentioning that Chinese characters for numerals are relatively easy to process because of their simple structure, while the characters for the target classifiers were more complex. Therefore, it is possible that learners' longer reading time with the presence of classifiers was amplified by the nature of the characters.

In the field of classifier acquisition, limited attention has been given to how classifiers are used by native speakers, which is closely related to the input for L2 learners of Chinese, and consequently the L2 acquisition of classifiers. Further studies on how classifiers are used, and how the usage varies across speakers, how the usage varies in different contexts and conditions would be informative.

The results presented generally support the lexical account of variability in L2 acquisition of new features and functional categories. Sufficient lexical knowledge, which includes access to different types of lexical information, and fast lexical retrieval can contribute to native-like performance regarding Chinese classifiers. The findings also lend support to the view that L1/L2 difference is not qualitative or categorical, instead it is gradual.

APPENDICES

APPENDIX A

Background questionnaires

BACKGROUND INFORMATION

All personal information you will provide is confidential. Please circle your answers and fill in the requested information.

Age: Sex: male female

City/Country of birth:

Are you a student? yes no

If yes, please indicate your current level of education:

What is your field of study?

Is English your native language? yes no

What language(s) does your mother speak? your father?.....

How old were you when you started to learn Chinese?

Where did you start to learn Chinese?

junior high school

high school

college

other.....

Please rate your proficiency in the following area for your Chinese:

	5 superior	4	3 average	2	1 poor
Speaking	5	4	3	2	1
Reading	5	4	3	2	1
Listening	5	4	3	2	1
Writing	5	4	3	2	1

Have you taken ACTFL before? yes no

(skip the next two questions if you choose 'no')

What's your proficiency according to the ACTFL?

Reading: Novice (Low/Mid/High) Intermediate (Low/Mid/High)

Advanced ((Low/Mid/High)

Speaking: Novice (Low/Mid/High) Intermediate (Low/Mid/High)

Advanced ((Low/Mid/High)

Listening: Novice (Low/Mid/High) Intermediate (Low/Mid/High)

Advanced ((Low/Mid/High)

When did you take the ACTFL?

If you speak languages other than English,

what languages are they?

please rate your proficiency in it:

I speak it fluently

I speak it
somewhat well

I have studied it, but
I don't speak it well

I speak it a little

Have you ever lived outside of the United States?

No

Yes. Describe briefly where, when, and for how long:

Have you spent any time longer than two months living in an environment where English is not the majority language? No.

Yes. Describe briefly where, when, and for how long:

Education background (circle all that apply, and please list the language, if applicable, on the right):

elementary school	in English	in another language
high-school	in English	in another language
college	in English	in another language
graduate school	in English	in another language

Location (circle all that apply, and please list the place, if applicable, on the right):

Where did you attend elementary school? in the continental US elsewhere.....

Where did you attend high-school? in the continental US elsewhere.....

Where did/do you go to college? in the continental US elsewhere.....

背景信息

All personal information you will provide is confidential. Please circle your answers and fill in the requested information.

年龄: 性别: 男 女

出生城市:

你是学生吗? 是 否

如果是学生, 目前教育水平为:

ELC Level.....

本科 专业.....

研究生 专业.....

其他 请注明.....

如果不是学生, 目前职业:

母语

你会说方言吗?

你的妈妈说什么语言(方言)?

你的爸爸说什么语言(方言)?

你什么时候开始学习英文?

你在美国生活了多长时间? ____ 年 ____ 个月

你第一次来美国时多大? ____ 岁

这是你第一次在英语国家生活吗?

是

如果不是, 你什么时候在哪个国家生活过?

几岁在那儿生活?

在那儿生活多久?

在那儿做什么?

你一共在英语国家生活了多长时间?

Please rate your proficiency in the following area for your English:

5 superior ---- 3 average ---- 1 poor

Speaking 5 4 3 2 1

Reading 5 4 3 2 1

Listening 5 4 3 2 1

Writing 5 4 3 2 1

除了母语和英语以外，你还会说其他语言吗? 是 否

如果是，你还会说哪种语言

什么水平（注明何种语言）:

我说得很流利 我说得还不错 我学过，但是说得不太好 我会说一点

什么水平（注明何种语言）:

我说得很流利 我说得还不错 我学过，但是说得不太好 我会说一点

教育背景（城市 / 国家）

你在哪儿读小学?

你在哪儿读中学?

你在哪儿读大学?

你在哪儿读研究生?

APPENDIX B

Proficiency test

Part 1: Choose a word to complete the sentence.

A. 其实 B. 感冒 C. 附近 D. 舒服 E. 声音 F. 把

E.g. 她说话的（ E ）多好听啊！

1. 电影马上就要开始了，（ ）手机关了吧。
2. 他很高，这张桌子太低，坐着很不（ ）。
3. 您可以选择火车站（ ）的宾馆，住那儿会更方便。
4. 天气冷，你多穿点儿衣服，小心（ ）。
5. 对一个女人来说，漂亮、聪明都很重要，但（ ）更重要的是快乐。

Part 2: Choose a word to complete the dialogue.

A. 工具 B. 收 C. 温度 D. 到底 E. 辛苦 F. 抱歉

E.g. A: 今天真冷啊，好像白天最高（ C ）才 2℃。

B: 刚才电视里说明天跟冷。

1. A: 丽丽说再等她几分钟，她马上就来。
B: 她（ ）在干什么呢，怎么这么慢？
2. A: 那个房间又脏又乱，星期六我去打扫了一下。
B: 原来是你啊，（ ）了，谢谢你！
3. A: 我刚从会议室过来，怎么一个人也没有？
B: 对不起，今天的会议改到明天上午了，您没（ ）到通知吗？
4. A: 语言是交流的（ ），只记字典里的字、词是不够的，要多听多说。
B: 对，这才是学习中文的好方法。
5. A: 真（ ），我迟到了。

B: 没关系，表演还有 5 分钟才开始。

Part 3. Choose a word to complete the paragraph.

从前有一位老人叫愚公，他家门前有两座山，又高又大， 1 ，全家人出门都很不方便。

一天，愚公把家里人叫到一起，说：“有山挡着，出门太困难了，我们把它搬走，好不好？”全家人都很 2 ，只有他的妻子没有信心。村里人看到愚公这么大年纪还在搬山，都很 3 ，也来帮助他们。有个叫智叟的老头儿看到了， 4 愚公太傻。愚公却说：“我死了还有儿子，儿子还有孙子，我们的人越来越多，山上的石头却越来越少，我们一定能 5 ！”

- | | | | |
|------------|------------|-------|-------|
| 1. A. 挡住了路 | B. 十分矛盾 | | |
| C. 因为无法推辞 | D. 犹豫了很长时间 | | |
| 2. A. 允许 | B. 注意 | C. 反对 | D. 赞成 |
| 3. A. 感动 | B. 生气 | C. 紧张 | D. 反对 |
| 4. A. 相信 | B. 考虑 | C. 笑话 | D. 确认 |
| 5. A. 发展 | B. 努力 | C. 到达 | D. 成功 |

APPENDIX C

Stimuli for lexical decision task

The columns contain the following information:

- First column: Item number. (Note that the first three items are dummies.)
- Second column: Item.
- Third column: word status: 0=non-word; 1=word.

Note:

- Real words include:
 - The 32 nouns used in the elicited production & self-paced reading task
 - Eight other words: unfamiliar words (e.g. tango) & familiar characters but unfamiliar combination (e.g. 酒馆 wine+room=bar)
- Non-words include:
 - Non-existing combination of characters (e.g. snow+umbrella)
 - Real words with one character replaced by another character that has similar meaning, pronunciation or shape (e.g. 味精 MSG-*味情)
 - Real words with the character order reversed (e.g. 眼睛 eye-*睛眼)

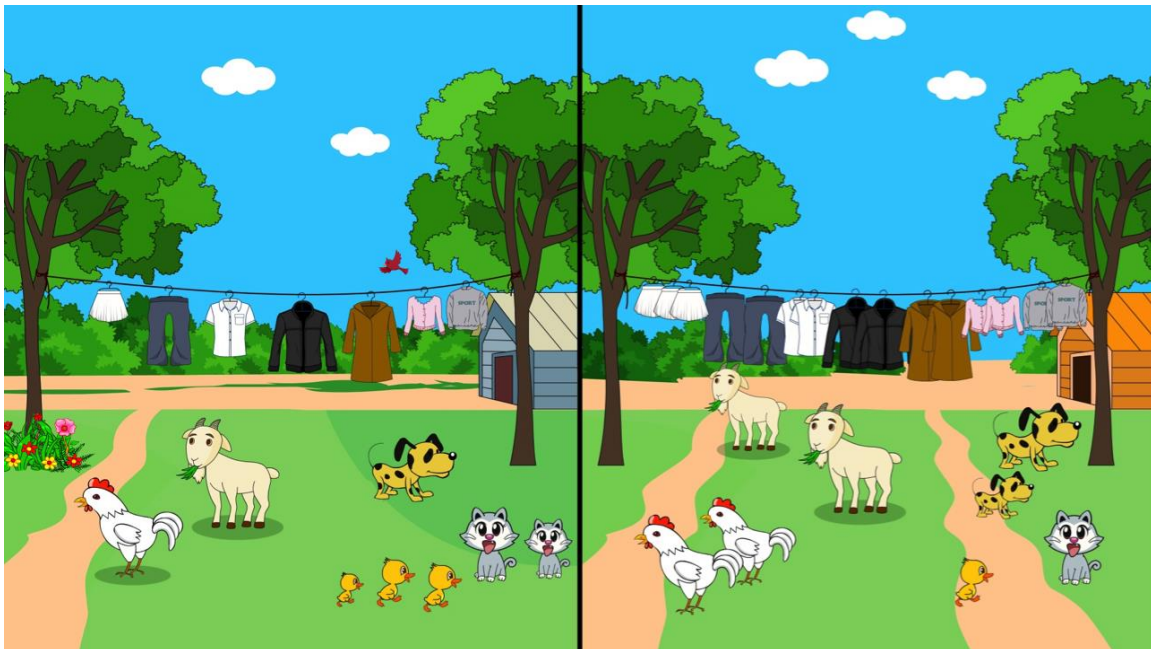
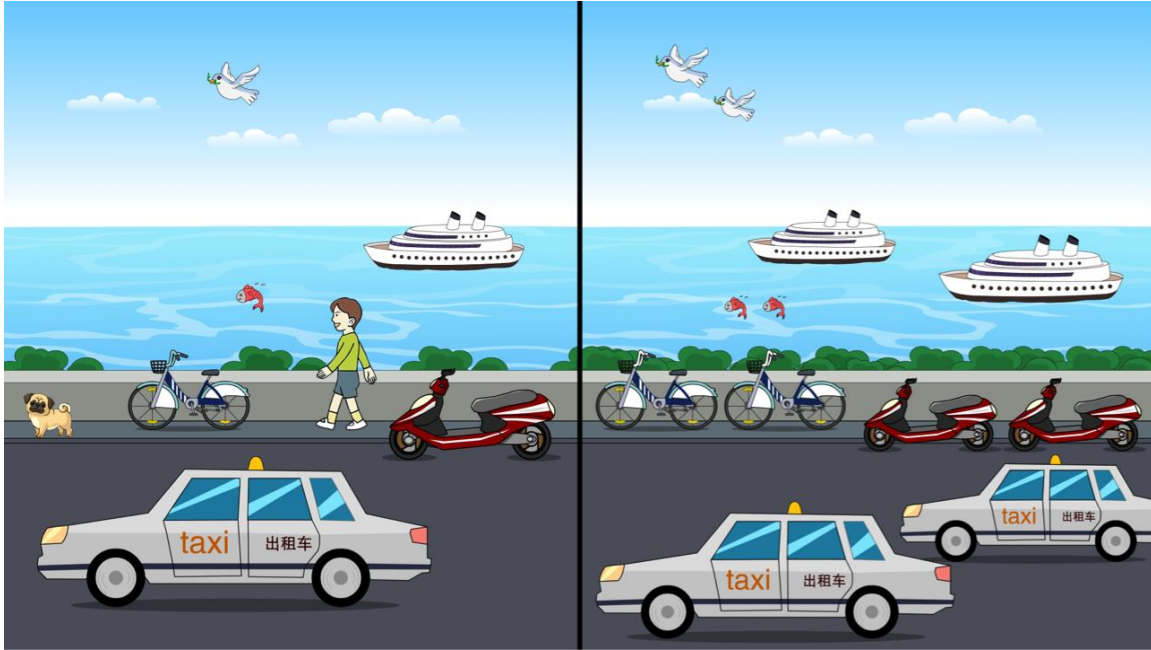
0	发沙	0
0	太阳 sun	1
0	柜子 cabinet	1
1	杂至	0
2	字典 dictionary	1
3	猫 cat	1
4	裙子 skirt	1
5	水树	0
6	电脑 computer	1
7	政客 politician	1
8	信用卡 credit card	1
9	羊 sheep	1
10	雪伞	0
11	外面 outside	1
12	心脏 heart	1
13	毛衣 sweater	1
14	乌云	1
15	毯子 blanket	1
16	昏	0
17	床 bed	1
18	摩托车 motorcycle	1
19	笼灯	0

20	船 boat	1
21	司机 driver	1
22	运动服 sweatshirt	1
23	味情	0
24	小鸟 bird	1
25	纸 paper	1
26	课本 text book	1
27	健体房	0
28	外套 overcoat	1
29	鱼 fish	1
30	几童	0
31	照片 photo	1
32	刚笔	0
33	饭杯	0
34	出租车 taxi	1
35	裤子 pants	1
36	鸭子 duck	1
37	音会	0
38	冰箱 refrigerator	1
39	T 恤衫 T-shirt	1
40	明友	0
41	小路 road	1

42	工作师	0
43	睛眼	0
44	探戈 tango	1
45	地图 map	1
46	电视 television	1
47	饺子 dumpling	1
48	电话	0
49	果苹	0
50	鸡 chicken	1
51	夹克 jacket	1
52	自行车 bike	1
53	书 book	1
54	腐豆	0
55	北瓜	0
56	桌子 table	1
57	务服	0
58	萄	0
59	衬衫 shirt	1
60	酒馆 bar	1

APPENDIX D

Pictures used in the elicited production task





APPENDIX E

Classifier-noun pairs in self-paced reading task

Table E1

Classifier-noun pairs used in self-paced reading task

Semantic domain	Classifier	Noun	Incorrect classifier from a different semantic domain	Incorrect classifier from the same semantic domain
Animacy	<i>zhi</i> (只 small animals)	<i>mao</i> ('cat')	<i>tiao</i> (long objects, e.g. fish, sneak)	<i>wei</i> (humans, e.g. teacher)
		<i>ji</i> ('chicken')	<i>tiao</i>	<i>wei</i>
		<i>xiaoniao</i> ('bird')	<i>tiao</i>	<i>wei</i>
		<i>yazi</i> ('duck')	<i>tiao</i>	<i>wei</i>
		<i>yang</i> ('sheep')	<i>tiao</i>	<i>wei</i>
Shape	<i>tiao</i> (long and slender objects)	<i>yu</i> ('fish')	<i>zhi</i> (small animals, e.g. cat)	<i>zhang</i> (flat surfaced objects)
		<i>kuzi</i> ('pants')	<i>jian</i> (clothes, e.g. shirt)	<i>zhang</i>
		<i>qunzi</i> ('skirt')	<i>jian</i>	<i>zhang</i>

Table E1 (cont'd)

		<i>tanzi</i> ('blanket')	<i>jian</i>	<i>zhi</i> (支 cylindric objects, e.g. pen)
		<i>chuan</i> ('boat')	<i>liang</i> (vehicles, e.g. car, bike)	<i>zhang</i>
		<i>xiaolu</i> ('road')	<i>liang</i>	<i>zhang</i>
zhang (flat surfaced objects)		<i>zhi</i> (纸 'paper')	<i>pian</i> 篇 (written items, e.g. essay)	<i>tiao</i>
		<i>zhaopian</i> ('photo')	<i>pian</i>	<i>tiao</i>
		<i>ditu</i> ('map')	<i>pian</i>	<i>tiao</i>
		<i>xinyongka</i> ('credit card')	<i>pian</i>	<i>tiao</i>
		<i>zhuozi</i> ('table')	<i>tai</i>	<i>zhi</i> (支)
		<i>chuang</i> ('bed')	<i>tai</i>	<i>zhi</i> (支)
Function	<i>jian</i> (clothes)	<i>maoyi</i> ('sweater')	<i>tiao</i> (e.g. pants)	<i>tai</i> (machines)
		<i>chenshan</i> ('shirt')	<i>tiao</i>	<i>tai</i>
		<i>waitao</i> ('overcoat')	<i>tiao</i>	<i>tai</i>

Table E1 (cont'd)

	<i>jiake</i> ('jacket')	<i>tiao</i>	<i>tai</i>
	<i>T-xushan</i> ('T-shirt')	<i>tiao</i>	<i>liang</i>
	<i>yundongfu</i> ('sweatshirt')	<i>tiao</i>	<i>liang</i>
<i>liang</i> (vehicles)	<i>chuzuche</i> ('taxi')	<i>tiao</i> (long and slender objects, e.g. boat)	<i>ben</i> (bounded item)
	<i>zixingche</i> ('bicycle')	<i>tiao</i>	<i>ben</i>
	<i>motuoche</i> ('motorcycle')	<i>tiao</i>	<i>ben</i>
<i>ben</i> (bound items)	<i>zidian</i> ('dictionary')	<i>zhang</i> (flat surfaced objects, e.g. paper)	<i>tai</i> (machines)
	<i>shu</i> ('book')	<i>zhang</i>	<i>tai</i>
	<i>keben</i> ('textbook')	<i>zhang</i>	<i>tai</i>
<i>tai</i> (machines)	<i>diannao</i> ('computer')	<i>zhang</i>	<i>liang</i> (vehicles)
	<i>dianshi</i> ('television')	<i>zhang</i>	<i>liang</i>
	<i>bingxiang</i> ('refrigerator')	<i>zhang</i>	<i>liang</i>

APPENDIX F

Stimuli for self-paced reading task

Displaying regions are divided by ‘/’; nouns (bolded) immediately after classifiers are the critical region.

1. 小王/看到/三只/猫/在/桌子/上/睡觉。

*Xiaowang kandao sanzhi **mao** zai zhuozi shang shuijiao*

Xiaowang see three-(Cl) cat on table sleep

‘Xiaowang saw three cats sleeping on the table.’

2. 我/听到/几只/小鸟/在/公园/里/唱歌。

*Wo tingdao jizhi **xiaoniao** zai gongyuan li changge*

I hear some-(Cl) bird in park inside sing

‘I hear some birds singing in the park.’

3. 小白/看到/四只/鸡/在/房子/后面/吃米。

*Xiaobai kandao jizhi **ji** zai fangzi houmian chimi*

Xiaobai see several-(Cl) chicken in house back eat rice

‘Xiaobai saw some chicken eating rice behind the house.’

4. 他们/看到/几只/羊/从/汽车/旁边/走过。

*Tamen kandao jizhi **yang** cong qiche pangbian zouguo*

Today see some-(Cl) sheep from car beside walk by

‘They saw some sheep passing by the car.’

5. 他/看到/几只/鸭子/在/小河/里/游泳。

Ta kandao jizhi yazi zai xiaohe li jingguo

He see some-(Cl) duck in river inside swim

‘He saw some ducks swimming in the river.’

6. 小美/以为/那条/裤子/在/衣柜/里/放着。

Xiaomei yiwei natiao kuzi zai yigui li fangzhe

Xiaomei think that-(Cl) pants in closet inside put

‘Xiaomei thought the pants are in the closet.’

7. 她/看到/那条/裙子/在/商店/打折/了。

Ta kandao natiao qunzi zai shangdian dazhe le

She see that-(Cl) skirt at store on sale

‘She see that skirt is on sale at store.’

8. 小美/看到/三条/船/在/小河/旁边。

xiaomei see santiao chuan zai xiaohe pangbian

Xiaomei see three-(Cl) ship in river side

‘Xiaomei saw three boats by the riverside.’

9. 请你/把/那条/鱼/从/冰箱/里/拿/出来。

Qingni ba natiao yu cong bingxiang li na chulai

Please that-(Cl) fish from refrigerator inside take out

‘Please take out the fish from the refrigerator.’

10. 房间里/有/三条/毯子/在/床上/放着。

Fangjianli you santiao tanzi zai chaungshang fangzhe

Room have three-(Cl) blanket on bed put

‘There are three blankets on the bed in the room.’

11. 我们/看见/一条/小路/从/门口/经过。

*Women kanjian yitiao **xiaolu** cong mekou jingguo*

We see one-(Cl) road from door pass by

‘We see a road in front of the door.’

12. 小爱/把/那张/桌子/从/客厅/搬了/出来。

*Xiaoai ba nazhang **zhuozi** cong keting banle chulai*

Xiaoai that-(Cl) table from living room Move-le out

‘Xiaoai moved that table out of the living room.’

13. 小李/看到/五张/照片/在/墙上/挂着。

*Xiaoli kandao wuzhang **zhaopian** zai qiangshang guazhe*

Xiaoli see five-(Cl) picture on wall hang

‘Xiaoli saw five pictures hanging on the wall.’

14. 请你/把/那张/纸/从/地上/拿/起来。

*Qingni ba nazhang **zhi** cong dishang na qilai*

Please that-(Cl) paper from floor pick up

‘Please pick up that piece of paper from the floor.’

15. 天明/把/那张/地图/从/书包/里/拿/出来。

*Tianming ba nazhang **ditu** cong shubao li na chulai*

Tianming that-(Cl) map from backpack inside take out

‘Tianming took the map out from the backpack.’

16. 他/把/那张/床/从/楼上/搬/下去了。

*Ta ba nazhang **chuang** cong loushang ban xiaqule*

He that-(Cl) bed from upstairs move down
'He moved the bed from upstairs to downstairs.'

17. 我/以为/那张/信用卡/在/钱包/里。

*Wo yiwei nazhang **xinyongka** zai qianbao li*
I think that-(Cl) credit card in wallet inside
'I thought my credit card is in the wallet.'

18. 我/把/那件/毛衣/从/衣柜/里/拿/出来。

*Wo ba najian **maoyi** cong yigui li na chulai*
I that-(Cl) sweater in closet inside take out
'I took that sweater out of the closet.'

19. 小英/问/这件/衬衫/在/哪儿/买的。

*Xiaoying ask zhejian **chenshan** zai nar maide*
Xiaoying forget this-(Cl) shirt in where buy
'Xiaoying asks where this shirt was bought.'

20. 她/爸爸的/那件/夹克/在/沙发/上/放着。

*Ta babade najian **jiake** zai shafa shang fangzhe*
She dad's that-(Cl) jacket on sofa top put
'Her dad's jacket is on the sofa.'

21. 小白/看到/那件/外套/在/椅子/上/放着。

*Xiaobai kandao najian **waitao** zai yizi shang fangzhe*
Xiaobai see that-(Cl) coat on chair top put
'Xiaobai saw that the jacket was on the chair.'

22. 小丽/买了/两件/T 恤衫/给/哥哥/过/生日。

Xiaoli maile yijian T-xushan gei gege guo shengri
Xiaoli buy two-(Cl) T-shirt to brother celebrate birthday
‘Xiaoli bought two T-shirts for her older brother to celebrate his birthday.’

23. 我/把/那件/运动服/从/学校/拿/回家/了。

Wo ba najian yundongfu cong xuexiao na huijia le
I that-(Cl) sweatshirt from school take back home
‘I brought the sweatshirt back home from school.’

24. 英爱/坐了/一辆/出租车/从/机场/回家。

Yingai zuole yiliang chuzuche cong jichang huijia
Yingai take one-(Cl) taxi from airport go home
‘Yiongai took a taxi from the airport to go home.’

25. 小天/骑了/一辆/摩托车/从/家里/去/学校。

Xiaotian qile yiliang motuoche cong tajia qu xuexiao
Xiaotian ride one-(Cl) motorcycle from home go school
‘Xiaotian rode a motorcycle from home to school.’

26. 天明/弟弟的/那辆/自行车/在/学校/坏了。

Tianming didide naliang zixingche zai xuexiao huaile
brother’ broken
Tianming that-(Cl) bike in school
s

‘Tianming’s brother’s bike is broken in school.’

27. 小天/看见/那台/冰箱/在/厨房/里面/放着。

Xiaotian kanjian naitai bingxiang zai chufang limian fangzhe

Xiaotian see that-(CI) refrigerator at kitchen inside put
'Xiaotian saw a refrigerator in the kitchen's.'

28. 小明/房间的/那台/电视/在/桌子/上/放着。

Xiaoming fangjiande naitai dianshi zai zhuozi shang fangzhe
Xiaoming room that-(CI) TV in table top put
'The TV in Xiaoming's room is on the table.'

29. 爸爸/送了/一台/电脑/给/小白/作/生日/礼物。

Baba songle yitai diannao gei Xiaobai zuo shengri liwu
Dad give one-(CI) computer to Xiaobai as birthday gift
'Last year Dad gave Xiaobai a computer as a birthday gift.'

30. 我/看到/几本/字典/在/书柜/上/放着。

Wo kandao jiben zidain zai shugui shang fangzhe
I see some-(CI) dictionary in book case top put
'I saw some dictionaries in the book case.'

31. 小白/把 /三本/课本/从/书包/里/拿/出来。

Xiaobai ba sanben keben cong shubao li na chulai
Xiaoybai three-(CI) textbook from backpack take out
'Xiaobai took out from the backpack the three textbook.'

32. 小明/拿着/几本/书/从/教室/里/出来。

Xiaoming nazhe jiben shu cong jiaoshi li chulai
Xiaoming take some-(CI) book from classroom inside out
'Xiaoming got out of the classroom with several books in hand.'

APPENDIX G

Stimuli for offline cloze task

Please complete the following phrases with a measure word other than “个”.

- | | |
|---------------|----------------|
| 1. 三_____裤子 | 17. 一_____小路 |
| 2. 一_____小鸟 | 18. 七_____运动服 |
| 3. 两_____衬衫 | 19. 两_____地图 |
| 4. 五_____床 | 20. 两_____摩托车 |
| 5. 两_____冰箱 | 21. 四_____T 恤衫 |
| 6. 四_____羊 | 22. 三_____信用卡 |
| 7. 一_____自行车 | 23. 十_____鸭子 |
| 8. 六_____书 | 24. 五_____裙子 |
| 9. 一_____夹克 | 25. 三_____照片 |
| 10. 两_____电脑 | 26. 四_____桌子 |
| 11. 四_____船 | 27. 三_____课本 |
| 12. 三_____字典 | 28. 两_____外套 |
| 13. 三_____电脑 | 29. 六_____鱼 |
| 14. 五_____猫 | 30. 一_____毯子 |
| 15. 一_____出租车 | 31. 八_____纸 |
| 16. 三_____毛衣 | 32. 六_____鸡 |

REFERENCES

REFERENCES

- Alarcón, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14 (3), 332–350
- Allan, K. (1977). Classifiers. *Language*, 53, 285–311.
- Aldwayan, S., Fiorentino, R., & Gabriele A. (2010). Evidence of syntactic constraints in the processing of wh-movement. A study of Najdi Arabic. learners of English. In B. VanPatten & J. Jegerski (Eds.) *Research in second language processing and parsing* (pp. 65-86). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67 (1), 1-48.
- Bi, Y., Yu, X., Geng, J., & Alario, F.-X. (2010). The role of visual form in lexical access: evidence from Chinese classifier production. *Cognition*, 116, 101-109.
- Carstens, V. (2000). Concord in minimalist theory. *Linguistic Inquiry*, 31, 319-355.
- Chafe, W. L. (ed.) (1980). The pear stories: Cognitive, cultural, and linguistic aspects of narrative production. Norwood, NJ: Ablex.
- Chen, B., Ning, A., Bi, H., & Dunlap, S. (2008). Chinese subject-relative clauses are more difficult to process than the object-relative clauses. *Acta Psychologica*, 129, 61–65.
- Chen, H.-C., & Tang, C.-K. (1998). The effective visual field in reading Chinese. *Reading and Writing: An Interdisciplinary Journal*, 10, 245–254.
- Cheng, L. L.-S., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30, 509–542.
- Chou, T.-L., Lee, S.-H., Hung, S.-M., & Chen, H.-C. (2012). The role of inferior frontal gyrus in processing Chinese classifiers. *Neuropsychologia*, 50, 1408-1415.
- Clahsen, H., & Felser, C. (2017). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*. Advance online publication. doi: 10.1017/S0272263117000250
- Corbett, G. (1991). *Gender*. Cambridge, UK: Cambridge University Press.

- Craig, C. (1986). Introduction. In C, Craig (Ed.), *Noun classes and categorization* (pp. 1-10). Amsterdam: John Benjamins Publishing Company.
- Croft, W. (1994). Semantic universals in classifier systems. *Word*, 45, 145-171.
- Cunnings, I. (2017). Interference in native and non-native sentence processing. *Bilingualism: Language and Cognition*, 20, 712–721.
- Dixon, R. M. W. (1982). *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin: Mouton.
- Erbaugh, M. S. (1986). Taking stock: The development of Chinese noun classifiers historically and in young children. In C, Craig (Ed.), *Noun classes and categorization* (pp. 399-436). Amsterdam: John Benjamins Publishing Company.
- Erbaugh, M. S. (2004). Chinese classifiers: Their use and acquisition. In P. Li, L. H. Tan, E. Bates & O. J. L. Tzeng (Eds.), *Handbook of East Asian psycholinguistics: Vol. 1. Chinese* (pp. 39-51). Cambridge, UK: Cambridge University Press.
- Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics*, 32, 187-220.
- Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, 66, 226-248.
- Franceschina, F. (2001). Morphological or syntactic deficit in near-native speakers? An assessment of some current proposals. *Second Language Research*, 17, 213–147.
- Franceschina, F. (2005). *Fossilised second language grammars: the acquisition of grammatical gender*. Amsterdam: John Benjamins.
- Gao, H. H. (2010). A study of Swedish speakers' learning of Chinese noun classifiers. In U. Bohnacker & M. Westergaard (Eds.), *The Nordic languages and second language acquisition theory, special issue of Nordic Journal of Linguistics*, 33, 197–229.
- Gao, Y. (1998). *Mental representations of Chinese numeral classifiers*. Doctoral dissertation, Lehigh University.
- Gebhardt, L. (2009). *Numeral classifiers and the structure of DP*. Doctoral dissertation, Northwestern University.
- Gebhardt, L. (2011). Classifiers are functional. *Linguistic Inquiry*, 42, 125-130.

- Grüter, T., Lau, E., & Ling, W. (2018). L2 listeners rely on the semantics of classifiers to predict. In A. B. Bertolini & M. J. Kaplan (Eds.), *Proceedings of the 42nd Annual Boston University Conference on Language Development* (pp. 303-316). Somerville, MA: Cascadia Press.
- Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: production or a real-time processing problem? *Second Language Research*, 28, 191–215.
- Hale, J. T. (2011). What a rational parser would do. *Cognitive Science*, 35, 399–443.
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22, 369-397.
- Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29, 33–56.
- Hopp, H. (2016a). Learning (not) to predict: Grammatical gender processing in adult L2 acquisition. *Second Language Research*, 32, 277–307.
- Hopp, H. (2016b). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, 37, 1253–1280
- Hopp, H. (2017). Individual differences in L2 parsing and lexical representations. *Bilingualism: Language and Cognition*, 20, 689–690.
- Hsiao., Y., & MacDonald, M. C. (2016). Production predicts comprehension: Animacy effects in Mandarin relative clause processing. *Journal of Memory and Language* 89, 87–109.
- Hu, Q. (1993). *The acquisition of Chinese classifiers by young mandarin-speaking children*. Doctoral Dissertation, Boston University.
- Huetting, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye-gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture*, 10, 39-58.
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Language Learning*, 57(1), 1–33.
- Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, 61, 940–967.
- Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research*

- methods in second language psycholinguistics* (pp. 20-49). New York: Routledge.
- Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, 59, 503–535.
- Keating, D. K., & Jegerski, J. (2015). Experimental design in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, 37, 1 – 32 .
- Ken, L.K., & Harrison, G. (1986). Young children's use of Chiense (Cantonese and Mandarin) sortal classifiers. In H. S.R. Kao & R. Hoosain (Eds.), *Linguistics, psychology, and the Chinese language* (pp. 125-146). Hong Kong: Center of Asian Studies, University of Hong Kong.
- Kramer, R. (2015). *The morphosyntax of gender*. Oxford: Oxford University Press.
- Kupisch, T., Akpınar, D., & Stöhr, A. (2013). Gender assignment and gender agreement in adult bilinguals and second language learners of French. *Linguistic Approaches to Bilingualism*, 3 (2), 150-179.
- Lau, E., & Grüter, T. (2015). Real-time processing of classifier information by L2 speakers of Chinese. In E. Grillo & K. Jepson (Eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development* (pp. 311-323). Somerville, MA: Cascadilla Press.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behav Res Methods*, 44, 325–343.
- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Liang, S.-Y. (2009). *The acquisition of Chinese nominal L2 adult learners*. Doctoral Dissertation, University of Texas at Arlington.
- Liu, Y., Yao, T-C., Shi, Y., Bi, N-P., Ge, L. (2009). Integrated Chinese Textbook (simplified and traditional character): Level 1, Level 2. Boston, MA: Cheng & Tsui.
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6, 1-16.
- Loke, K. K. (1996). Norms and realities of Mandarin shape classifiers. *Journal of the Chinese Language Teachers Association*, 31, 1-22.
- Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.

- Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, 39, 861-904.
- McCarthy, C. (2008). Morphological variability in the comprehension of agreement: an argument for representation over computation. *Second Language Research*, 24, 459-486.
- Myers, J. (2000). Rules vs. analogy in Mandarin classifier selection. *Language and Linguistics*, 1, 187-209.
- Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2. *Studies in Second Language Acquisition*, 21, 49-80.
- Polio, C. (1994). Non-native speakers' use of nominal classifiers in Mandarin Chinese. *Journal of the Chinese Language Teachers Association*, 29, 51-66.
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103-133.
- Qian, Z., & Garnsey, S. M. (2016). A sheet of coffee: An event-related brain potential study of the processing of classifier-noun sequences in English and Mandarin. *Language, Cognition and Neuroscience*, 31, 761-784.
- Ritter, E. (1993). Where is gender? *Linguistic Inquiry*, 24, 795-803.
- Saalbach, H., & Imai, M. (2012). The relation between linguistic categories and cognition: The case of numeral classifiers. *Language and Cognitive Process*, 27 (3), 381-428.
- Sabourin, L. & Stowe, L. (2008) Second language processing: When are first and second languages processed similarly? *Second Language Research*, 24, 397-430.
- Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language learning*, 61, 80-116.
- Senft, G. (2000). Nominal classification system. In Gunter Senft (Eds.), *Systems of nominal classification*. Cambridge: Cambridge University Press.
- Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, 7, 118-132.

- Snellings, P., Van Geldern, A., & De Glopper, K. (2002). Lexical retrieval: An aspect of fluent second language production that can be enhanced. *Language Learning*, 52, 723-754.
- Spinner, P., & Juffs, A. (2008). L2 grammatical gender in a complex morphological system: The case of German. *International Review of Applied Linguistics*, 46 (4), 315-348.
- Spinner, P. (2013a). The L2 acquisition of number and gender in Swahili: A feature reassembly approach. *Second Language Research*, 29, 455-479.
- Spinner, P. (2013b). Language production and reception: A processability theory study. *Language Learning*, 63, 704-739.
- Spinner, P., & Thomas, J. (2014). L2 learners' sensitivity to semantic and morphophonological information on Swahili nouns. *International Review of Applied Linguistics in Language Teaching*, 52 (3), 283-311.
- Stowe, L. (1986). Evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227-245.
- Srinivasan, M. (2010). Do classifiers predict differences in cognitive processing? A study of nominal classification in Mandarin Chinese. *Language and Cognition*, 2, 177-190.
- Sumiya, H. (2008). *The effect of familiarity and semantics on early acquisition of Japanese numeral classifiers*. Doctoral dissertation, University of Colorado.
- Tai, J. H-Y., & Wang, L. (1990). A semantic study of the classifier tiao. *Journal of Chinese Language Teachers Association*, 25, 35-56.
- Tien, Y.-M., Tzeng, O. J. L., & Hung, D. L. (2002). Semantic and cognitive basis of Chinese classifiers: A functional approach. *Language and Linguistics*, 3 (1), 101-132.
- Traxler, M. J. & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies. *Journal of Memory and Language*, 35, 454-475.
- Tse, S. K., Li, H., & Leung, S. O. (2007). The acquisition of Cantonese classifiers by preschool children in Hong Kong. *Journal of Child Language*, 34, 495-517.
- VanPatten, B., Keating, G. D., & Leiser, M. J. (2012). Missing verbal inflections as a representational problem: Evidence from self-paced reading. *Linguistic Approaches to Bilingualism*, 2 (2), 109-140.

- White, L., Valenzuela, E., Kozłowska-Macgregor, M., & Leung, Y-K. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25, 105-133.
- Zhang, J., & Lu, X. (2013). Variability in Chinese as a foreign language learners' development of the Chinese numeral classifier system. *The Modern Language Journal*, 97, 46-60.
- Zhang, Y., Zhang, J., & Min, B. (2012). Neural dynamics of animacy processing in language comprehension: ERP evidence from the interpretation of classifier-noun combinations. *Brain & Language*, 120, 321–331.
- Zhou, X., Jiang, X., Ye, Z., Zhang, Y., Lou, K., & Zhan, W. (2010). Semantic integration processes at different levels of syntactic hierarchy during sentence comprehension: An ERP study. *Neuropsychologia*, 48, 1551–1562.
- Zyzik, E., & Polio, C. (2017). *Authentic materials myths: Applying second language pedagogy to classroom teaching*. Ann Arbor: University of Michigan Press