Chinese Verb Complement Constructions of Manner and States: A Corpus-Based Comparison Between L1 and L2 Speakers

Hong Gang Jin, Jie Zhang, and Hongyin Tao

Abstract This paper deals with the acquisition of Verb Complement Constructions of Manner and States (VCM/S, 方式/情态补语) from a comparative and corpus-based perspective. An examination of L1 and L2 Chinese VCM/S production and development yields three main findings: (a) there are marked quantitative and qualitative differences between L1 and L2 VCM/S production at both construction and component levels; (b) these persistent productive differences reflect the indispensable roles of psycholinguistic factors, such as frequency, complexity, form-meaning mapping, and co-occurrence patterns of VP and VC, especially on verb choices; and (c) L2 VCM/S construction learning is like any other construction learning that follows a U-shaped learning path that consists of unique and distinctive stages. The process also involves both implicit factors and explicit classroom input and instruction. The theoretical and pedagogical implications of these findings are discussed.

Keywords Complement of Manner and States · Construction learning · Form-meaning mapping · Usage-based models · Complexity scale · L1 and L2 comparison

H. G. Jin (✉) · H. Tao
East Asian Languages and Literatures, Hamilton College, 198 College Hill Road, Clinton, NY 13323, USA
e-mail: hjin@hamilton.edu

J. Zhang
Department of Modern Languages, Literatures, and Linguistics, University of Oklahoma, 780 Van Vleet Oval, Norman, OK 73019, USA

J. Zhang · H. Tao
Department of Asian Languages and Cultures, UCLA, Los Angeles, CA 90095-1540, USA

H. H.-J. Chen et al. (eds.), *Learner Corpora: Construction and Explorations in Chinese and Related Languages*, Chinese Language Learning Sciences,
https://doi.org/10.1007/978-981-19-5731-4_6

1 Introduction

Complement constructions are a major syntactic pattern in the Mandarin Chinese grammatical system, where a variety of complements can be formed to express a range of meanings, such as result, direction, degree, and so forth (Li & Thompson, 1981). As such, they constitute some of the most unique features of the Chinese syntactic system (Shen, 2003). Following our L1 corpus-based study (Tao et al., 2020) on a similar construction, this paper deals with the acquisition of one type of complement construction, which we call Verb Complement Constructions of Manner and States (VCM/S, 方式/情态补语). VCM/S constructions typically consist of three key components: the verb predicate (VP), the complementizer *de* (得), and the complement of different syntactic structures. They indicate either the manner in which the action named by the verbal predicate is executed or evaluated or a state toward which the action is carried out (ibid.). Two quick examples illustrating these patterns can be found in (1) and (2).

(1) 这个人写得不好.
Zhe ge ren xie de bu hao.
This person write DE not well.
'This person does not write well.'

(2) 他变得很精神.
Ta bian de hen jingshen.
He become DE very energetic
'He becomes very energetic.'

In (1), the complement *bu hao* 'not well' can be seen as an evaluation ('how well') of the verbal predicate *xie* 'write'. In (2), on the other hand, the complement *hen jingshen* 'very energetic' can be understood to be the state toward which the action of *bian* 'change, become' is carried out.

As a construction, VCM/S involves multiple components and has posed challenges to learners of Chinese as a second language (CSL). Previous studies have shown that learners' error rate is at 25%–50% due to the uniquely grammaticalized structure and subtle functions associated with the construction (Sun, 2002; Feng, 2013; Jiang, 2019, among others). Few CSL studies, however, have examined VCM/S from the perspective of usage-based construction learning and dealt with both implicit and explicit learning factors, nor have they investigated VCM/S as an independent construction from other complement types and compared L2 learner development in connection with L1 production data. When learner data do get analyzed, however, existing studies tend to be descriptive in nature, focusing, for example, mostly on single VCM/S component and raw frequency counts of VCM/S sentences (e.g. D. Sun, 2002; Q. Sun, 2018; Zhou & Deng, 2009, among others).

The current study intends to address those shortfalls by carrying out a comparative corpus study, and it will be informed by usage-based approaches (Bybee & Hopper, 2001; Ellis, 2006, 2008, 2012) and Construction Grammar (Goldberg, 1995; Trousdale & Hoffmann, 2013), where form and meaning pairing and co-occurrence properties are argued to play a critical role in understanding grammatical patterns in both the first language (L1) and the second language (L2). Our data comprise corpora of compositions written by CSL learners in the US college setting and by L1 Chinese speakers in preparation for national university entrance examinations. By using both L1 and L2 data, we hope to (a) determine if there are similarities or differences between L1 and L2 speakers' production of the VCM/S construction in terms of frequency, form, function, form-function mapping, as well as distributional properties in the form of co-occurrence patterns (Ellis, 2002, 2012); and (b) delineate L2 construction learning paths at different stages. Our comparative empirical study will form the basis for further exploration of L2 acquisition theory and pedagogical practice.

2 Literature Review

2.1 Usage-Based Approaches to Construction Learning

Usage-based approaches to language acquisition explore how humans learn language from experience and view language acquisition as a process of learning constructions (Bybee & Hopper, 2001; Ellis, 2006, 2008, 2012; Hoey, 2005). This acquisitional model emphasizes associative and cognitive principles of learning and focuses on investigating psycholinguistic factors of construction acquisition such as frequency, contingency, and form-meaning mapping that drive the acquisition and use of linguistic constructions (Ellis, 2012). Two concepts play the central role in this model: constructions and distributional properties of constructions.

Constructions, according to Goldberg (1995) and Trousdale and Hoffmann (2013), are defined as form-meaning mappings, conventionalized in the speech community, and entrenched as language knowledge in the learner's mind. Constructions are the fundamental units of language and language acquisition. Factors affecting construction acquisition are believed to come from several dimensions: (1) form-related factors such as frequency and salience; (2) function-related factors such as prototypicality, generality, and redundancy; (3) contingency of form and function; and (4) learner related factors such as learner attention, automaticity, and transfer, among others (Ellis, 2002, 2012).

Distributional properties are found to affect language processing and learning. Research has shown that language users are sensitive to detailed distributional information at many levels of linguistic analysis and at different grain sizes: from phonemes, morphemes, words, multi-word phrases, and syntactic constructions. Both language comprehension and production are affected by distributional factors,

such as the overall frequencies of the syntactic construction (Gahl & Garnsey, 2004; Tily et al., 2009, among others), the frequency of different components in specific syntactic constructions (Clifton et al., 1984; Garnsey, et al., 1997; Arnon & Snider, 2010, among others), and co-occurrence relations between verbs and specific arguments/complements (Trueswell and Tanenhaus 1994; Tao et al., 2020).

2.2 Chinese VCM/S Acquisition Studies

CSL research on the Chinese VCM/S acquisition began mostly in the beginning of this century. Many studies focused on the umbrella Chinese *de* complements containing 5–8 different complement constructions, of which VCM/S was one of them. Among the acquisition studies of VCM/S constructions, a majority of them were descriptive in nature, focusing on the identification, categorization, and description of the VCM/S development in terms of raw frequency counts of VCM/S sentences (D. Sun, 2002; Q. Sun, 2018; Zhou & Deng, 2009), interlanguage error patterns (D. Sun, 2002; Feng, 2013; Jiang, 2019, among others), and some general developmental patterns (Q. Sun, 2018; Zhou & Deng, 2009; Feng, 2013, among others).

While most existing CSL VCM/S studies used elicited or survey data, a few utilized corpus data. In these cases, the majority of researchers used corpus data from two sources. One is the overseas students' interlanguage composition corpus collected from a Chinese proficiency test known as *Hanyu Shuiping Kaoshi* (HSK), and the other is a self-built corpus of written samples collected from one or several institutions. For example, D. Sun (2002) analyzed 184 sample sentences containing VCM/S constructions from the Beijing Language and Culture University (BCC) corpus. Based on the error rate, the author concluded that VCM/S was the easiest for L2 learners to acquire among all verb complement constructions as fewer errors were found in their VCM/S (N = 184) collection. This claim was supported by Feng (2013) and Jiang (2019), especially in comparison with other complement types such as resultatives and potentials. Zhou and Deng (2009), by contrast, investigated two types of VCM/S constructions, where the object is in different positions (VO and OV), with both corpus data and experimental tests, and revealed that the OV structure in VCM/S constructions was actually hard to acquire and it was absent in L2 learners' production until they reached advanced proficiency levels.

Overall, we find that previous studies are descriptive in nature and that the conclusions are generally mixed. Few have examined the acquisition of the VCM/S construction learning using large-scale corpus data for both L1 and L2 with a unified L1 background, and rarely have researchers investigated both forms and functions of VCM/S constructions and the psycholinguistic factors of construction learning, such as construction frequency and complexity that may impact learner development. More importantly, previous studies have mainly looked at either the verb predicate (VP) or the verb complement (VC) before and after *de*, without examining VCM/S

as a construction with the two co-occurring open slots of VP and VC. This study will attempt to address those issues by adopting usage-based approaches to language acquisition, focusing on construction learning with corpus evidence from both L1 and L2 Chinese.

Applying usage-based approaches to construction learning, specifically that of the Chinese VCM/S construction, this study is set out to investigate three research foci: (a) VCM/S production in frequency and distribution; (b) verb choices in the predicate; and (c) VCM/S complexity scales. For each of these research foci, we seek to explore similarities and differences between L1 and L2 as well as the developmental pattern of L2 speakers' production. In the end, we will further explore the implications of the results of the comparative data for both acquisition theory and pedagogical practices in CSL.

3 Data and Methodology

3.1 The Corpora

The corpus used in this study is composed of 1,284 compositions written by CSL learners and Chinese L1 speakers with a total word count of 376,387. The learner corpus consists of 1,136 compositions written by English-speaking CSL learners at roughly four levels based on the American Council on the Teaching of Foreign Languages (ACTFL) proficiency scale[1]: (1) Novice-Mid to Intermediate-Low, (2) Intermediate-Mid to Intermediate-High, (3) Intermediate-High to Advanced-Low, and (4) Advanced-Mid and higher. For ease of discussion, we will use the short forms L2-A, L2-B, L2-C, and L2-D to represent, respectively, these four learner groups. The L2 collection came from three sources. The first two levels are a collection of student compositions at a comprehensive public university in North America. The third level comes from an intensive US study abroad program whose students represented over 20 universities and colleges in North America. Finally, the fourth level is a selection of compositions retrieved from the *Hanyu Shuiping Kaoshi (*HSK*) Dongtai Zuowen Yuliaoku* (Chinese Proficiency Test Dynamic Composition Corpus) Version 1.1.[2] The compositions culled from this collection were mostly narrative, descriptive, or argumentative by genre, and only those who registered their nationality as either the United States or Canada were included.

[1] American Council on the Teaching of Foreign Languages (ACTFL) categorizes foreign language proficiency into five major scales: Novice, Intermediate, Advanced, Superior, and Distinguished. The Novice, Intermediate, and Advanced levels each have three sublevels: Low, Mid, and High (ACTFL, 2012).

[2] The HSK Advanced was designed for CSL learners who have completed at least 4 years of Chinese instruction or who have been immersed in Chinese speaking environments for more than 3,000 h (Zhang, 2011).

Table 1 Composition of the corpus

Proficiency levels	Number of samples	Mean sample length by words	Word type	Word token	TTR
L2-A	409	110	3,175	44,994	7
L2-B	248	181	4,269	44,968	10
L2-C	206	596	7,336	122,849	6
L2-D	273	320	8,850	87,405	10
L2 Mean	**284**	**302**	**5,907**	**75,054**	**8.25**
L1	**148**	**515**	**11,203**	**76,171**	**15**
Total	**1,284**		**34,833**	**376,387**	**Ave. 9.6**

To maximize compatibility, we also include 148 compositions written by Chinese L1 speakers to serve as the L1 benchmark. This subset is a collection of compositions written by Chinese high school students taking or preparing for the National Matriculation Test (*gaokao* 高考), a high-stakes standardized test taking place annually in China. Most of the students who wrote these essays were approximately 17–20 years old in their third year of high school. The essays were downloaded from the official educational websites *Zhongguo Jiaoyu Zaixian* (China Education Online) and *Renmin Wang* (People's Daily Online). Several genres were represented, including narrative, argumentative, expository, and prose.

Table 1 presents the composition of the corpus with counts of the number of writing samples, average length of the samples (word per sample (wps)), word type, word token, and type/token ratio (TTR). In terms of the length of the writing samples, the L1 group is over 500 wps, much longer than the average L2 samples, which is at 302wps (the exception is the L2-C group, which has an average length of near 600wps). In terms of the variety of words employed, as one would expect, L1 speakers' writing averaged more word types (11,203 vs. 5,907), tokens (76,171 vs. 75,054), and a higher TTR (15 vs. 8.25). Across all proficiency levels, the L2 learner data demonstrate a growth in the number of types, tokens, and TTR as their overall language proficiency improves, with the exception of the L2-B group, which has a relatively higher TTR of 10. This may be due to the fact that this group is composed of learners enrolled in two semesters (fifth and sixth semesters of the program sequence) with the likelihood of being assigned compositions on more topics than learners over only one semester.

3.2 VCM/S Tagging and Coding Decisions

Both L1 and L2 data were first tagged for parts of speech information with the POS tool developed by the Chinese Applied Linguistics Institute (http://corpus.zhonghuayuwen.org/). Then a regular expression under AntConc (Anthony, 2019) was used to extract all the constructions fitting the VCM/S pattern, with the results being further filtered manually before coding was performed.

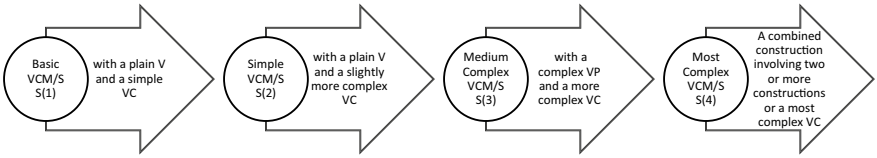

Fig. 1 Complexity scale of the Chinese VCM/S construction

Using an Excel spreadsheet, the first two authors of the paper went through all instances containing *de* and deleted cases that were deemed not the focus of the current study, such as potential verb complements, degree verb complements, and *de* used as a modal verb. Out of a total of 424 entries, there were 14 entries in discrepancy (3.3%). The two coders discussed these entries and reached an agreement for all 14 cases.

To capture how form and meaning are mapped and to explore the range of verbs and the prototypicality effect, we divided verbs in VCM/S constructions into four bands based on their raw frequencies in our corpus. A-list verbs are those with 18 or more token frequencies in the corpus; B-list includes verbs with 10–17 token frequencies; C-list verbs are common verbs with 2–9 token frequencies; in contrast, D-list includes infrequent words that have only one instance in our corpus. (For a full list of the verbs in frequency bands, please refer to Appendix.)

To capture the complexity of VCM/S constructions, we developed a coding scheme that treats each VCM/S instance as a holistic unit consisting of a verb phrase (VP) + DE + VERB COMPLEMENT (VC). Compared with previous approaches to VCM/S constructions, which focused on either the VP or the VC, this approach allows us to better capture the co-occurrence patterns of VPs and VCs, and their degrees of complexity as a construction. In Fig. 1, the complexity scales of the VCM/S construction are presented in four main categories: (1) Basic VCM/S, (2) Simple VCM/S, (3) Medium Complex VCM/S, and (4) Most Complex VCM/S. Again for ease of discussion, the short forms S(1), S(2), S(3), and S(4) will be used to refer to these complexity scales, respectively.

S(1) Basic VCM/S, represented as V + DE + (HEN) ADJ, is the most prototypical VCM/S structure. It is composed of a plain V (single syllabic or disyllabic) and an adjective as VC, denoting and evaluating the manner of the action. Because normally the unstressed adverb 很 *hen* 'very' is obligatorily required to co-occur with the adjective, we consider the use of *hen* an instance of S(1) instead of S(2). Here is an example from our corpus: 他站得很高 *Ta zhan de hen gao* 'He stands tall'.

S(2) Simple VCM/S is structurally more complex than S(1) in that different types of adverbial modifiers on the complement are used to convey different degrees. Depending on the complexity of the modifier, S(2) has two formulas: 2.1, V + DE + PREVERBAL ADV + ADJ or V + DE ADJ + POSTVERBAL ADV, such as preverbal 非常 *feichang* 'extremely', 太 *tai* 'too', and 特别 *tebie* 'especially' and postverbal 极了 *jile* 'extremely' or 一点 *yidian* 'a little'; 2.2, V DE + ADV PHRASE

+ ADJ, such as comparative 越来越 *yuelaiyue* 'more and more' /比 *bi* 'more than'/ 更 *geng* 'even more'. Examples from the corpus are 2.1 他吃得非常快 *Ta chi de feichang kuai* 'He eats extremely fast', and 2.2 他写得越来越好 *Ta xie de yuelaiyue hao* 'He writes better and better'.

S(3) Medium Complex VCM/S conveys its construction complexity through both VP and VC. The VP complexity is seen in its two forms when an object is required: VOV (taking an Object after the V and another reduplicated V) or OV[3] (preposing the Object before the V). The VC complexity is achieved through a variety of means. In addition to using an adverb or adverbial phrase as modifiers in S(2), it can also use a wider range of grammatical and lexical forms to achieve VCM/S functions, such as a complex adjective (two adjectives combined), a VP, or an idiom. Taking into consideration the complexities of both VP and VC, S(3) has three forms: 3.1, VP (VOV) + DE + (ADV/ADV PHRASE) + ADJ; 3.2, VP (OV) DE + (ADV/ADV PHRASE) + ADJ; 3.3, VP (VOV/OV) + DE + COMPLEX ADJ/VP/IDIOM. An example of 3.1 is 他写字写得非常快 *Ta xie zi xie de feichang kuai* 'He writes very quickly'; an example of 3.2 is 他字写得不太好 *Ta zi xie de butai hao* 'His handwriting is not very good'; and examples of 3.3 include 妈妈变得不通人情 *Mama bian de butong renqing* 'Mom becomes unsympathetic', 他们要吃得上点档次 *Tamen yao chi de shang dian dangci* 'They want to eat fancier food', and 他同学长得又高又胖 *Ta tongxue zhang de you gao you pang* 'His classmate is tall and chubby'.

S(4) Most Complex VCM/S achieves its highest complexity in two ways: the use of combined constructions and of a clause as a complement. Combined constructions allow two constructions to co-occur in a VCM/S. The added construction is often a disposal/passive/causative construction, introduced by 把 *ba or* 将 *jiang*[4] (the disposal markers), 被 *bei* (the passive marker), or 让 *rang,* 使 *shi or* 将 *jiang* (the causative markers). The additional information is to further specify the manner of the action via disposal, passive, and causative means. The use of a clause as a complement further enhances the speaker's evaluative stance through an expressed agent or patient of the action. The three S(4) formula are 4.1, V DE + (ADV/ADV PHRASE) CLAUSE, 4.2, CONSTRUCTION (BA/JIANG/BEI) + VCM/S OF (O)V + DE + (ADV/ADV PHRASE) + ADJ/IDIOMS/VP/CLAUSE (ADV) and 4.3, CONSTRUCTION (JIANG/RANG/SHI) + (O)V + DE + (ADV /ADV PHRASE) + ADJ/IDIOMS/VP/CLAUSE. Each combined construction can take different forms depending on the specific configuration of VPs and Cs. Here are some examples from the corpus: 4.1, 风吹得人站不住脚跟 *Feng chui de ren zhan bu zhu jiaogen* 'The wind blows so hard that people cannot stand still'; 4.2, 把巧克力的甜腻细滑刻画得淋漓尽致 *Ba qiaokeli de tiannixihua kehua de linlijinzhi* 'give the fullest and most vivid depiction of the sweetness and smoothness of chocolates'; and 4.3, 你使自己

[3] See more detailed discussion on the comparison of the two structures in Sect. 5.1 on VCM/S complexity.

[4] 将 *jiang* has two functions in Chinese: one is a disposal marker as 把 *ba* and the other is a causative marker as 让 *rang* or 使 *shi.* It thus can appear in two different constructions, denoting different functions.

Table 2 Complexity scales and formula of VCM/S constructions

Complexity	Formula
S(1) Basic VCM/S	1. V DE + (HEN) ADJ
S(2) Simple VCM/S	2.1. V DE + ADV + A (ADV) 2.2. V DE + ADV PHRASE + ADJ
S(3) Medium complex VCM/S	3.1. VP (VOV) DE + (ADV/ADV PHRASE) + ADJ 3.2. VP (OV) DE + (ADV/ADV PHRASE) + ADJ 3.3. VP (VOV/OV) DE + (ADV/ADV PHRASE) COMPLEX ADJ/VP/IDIOM
S(4) Most complex VCM/S	4.1. V DE + (ADV/ADV PHRASE) CLAUSE 4.2. C (BA/JIANG/BEI) + (O)V + DE + (ADV/ADV PHRASE) + ADJ/IDIOMS/VP/CLAUSE (ADV) 4.3. C (JIANG/RANG/SHI) + (O)V + DE + (ADV/ADV PHRASE) + ADJ/IDIOMS/VP/CLAUSE

的生活变得丰富有趣 *Ni shi ziji de shenghuo bian de fengfu youqu* 'You make your life richer and more fun'.

As summarized in Table 2, we proposed a four-scale VCM/S complexity scheme with 9 structural formulas to reflect the VCM/S complexity on form-meaning mapping and distributional properties between VP and VC as observed in the corpora. Later, we will testify to the validity of the scheme as it is taken as a measure to discriminate L1 and L2 speakers' VCM/S usage patterns both contrastively and developmentally.

4 Results

With the data and coding systems in place, this study has yielded a number of interesting results about L2 learners' VCM/S construction acquisition in areas of construction frequency distribution, verb choices in VPs, complexity scales, as well as differences between L1 and L2 production. In the following sections, we will report these findings along the lines of our proposed research foci.

4.1 VCM/S Frequency Distribution by L1 and L2 Speakers

As shown in Table 3, a total of 424 instances of VCM/S constructions were identified in the corpus out of 1,284 composition samples.[5] The L1 data had the most VCM/S instances (117). L2 learners were able to use VCM/S constructions as early as at

[5] To compensate for the varying sizes of the sub-corpus, we calculated mean frequencies of VCM/S constructions by a number of samples.

Table 3 Frequency distribution of VCM/S constructions in the corpus

Proficiency levels	Number of samples	Number of VCM/S	Mean VCM/S per sample
L2-A	409	74	0.18
L2-B	248	59	0.24
L2-C	206	101	0.49
L2-D	273	73	0.27
L2 Mean	**1136**	**307**	**0.27**
L1	**148**	**117**	**0.79**
Total	**1,284**	**424**	

the Novice-Mid to Intermediate-Low level, during which the VCM/S constructions were introduced to learners where the data were collected.

The comparative data demonstrated that L1 speakers' VCM/S production had a relatively higher frequency (at 0.79 per sample), while L2 learners in general exhibited a tendency of VCM/S underuse (at 0.27 per sample on average). A two-sample *Z*-test for the L1 and L2 data shows that the result is significant ($Z = 12.7$, $p = 0.01$).[6] The underuse pattern is conspicuous in the L2-A group (at 0.18 per sample). L2-B learners only increased their VCM/S production moderately (at 0.24 per sample). L2-C and L2-D learners had relatively higher VCM/S production per sample (at 0.49 and 0.27) among all learner groups.

The type and token frequencies of VCM/S constructions were tabulated in Table 4. Several trends emerged here. To begin with, there was a clear difference between L2 learners' usage of VCM/S constructions and that of L1 speakers. L1 speakers used 34 VCM/S construction types and 117 construction tokens, while L2-D, our highest proficiency L2 group, had 24 types and 73 tokens. Among L2 learners, the higher the proficiency level, the more construction varieties were exhibited. This is reflected by the raw frequencies of construction types (11 for L2-A, 14 for L2-B, 14 for L2-C, and 24 for L2-D) and the type-token ratio (0.149 for L2-A, 0.237 for L2-B, 0.139 for L2-C, and 0.329 for L2-D).

The data on VCM/S production frequency and its distribution indicated a significant difference between L1 and L2 speakers who exhibited underperformance in VCM/S types, tokens, and production quantity.

[6] We wish to thank Johnny Lin of UCLA Institute for Digital Research and Education for his statistical advice.

Table 4 VCM/S constructions by type and token frequency

Proficiency levels	VCM/S construction type	Number of samples	Average number of types by sample	VCM/S construction token	Average number of tokens by sample
L2-A	11	409	0.0269	74	0.1486
L2-B	14	248	0.0565	59	0.2373
L2-C	14	206	0.0680	101	0.1400
L2-D	24	273	0.0879	73	0.3288
L1	34	148	0.2297	117	0.2906

4.2 Verb Choices in the VCM/S Construction by L1 and L2 Speakers

Next, we report the results of verb choices in terms of type and token frequencies in VCM/S constructions. In our corpus, both L1 and L2 speakers' data included, there are altogether 117 verbs by type and 423 verbs by token.[7]

Based on our verb coding system, Fig. 2 shows the distribution of verb types by frequency band. Compared with the L2 data, the L1 speakers produced a markedly larger proportion of D-list verbs (63.1%), indicating a large lexical repertoire of verbs. In contrast, the proportions of A-, B-, and C-list verbs in the L1 data are quite small, about 36.9% combined. The L2 learner data, on the other hand, have an overall large proportion of A-, B-, and C-list verbs. The proportion of C-list verbs is noticeably high, indicating L2 learners' limited lexical repertoire of verbs. Compared with L1 speakers, L2 learners mainly relied on common, high-frequency, and general verbs while lacking a wide range of specific and abstract verbs. L2-D learners' distribution pattern is approaching L1 speakers but differed in quantity (44.1% vs. 24.6% for C-list and 32.4% vs. 63.1% for D-list), meaning the advanced learners like L2-D are still expanding their verbal repertoire and are beginning to accumulate more specific verbs.

The distribution of verb tokens by frequency band is presented in Fig. 3. Two opposite trends are worth noting. First, we can see more clearly that L2 learners across proficiency levels used a noticeably large number of A- and B-list verbs (64.8% for L2-A, 45.8% for L2-B, 75% for L2-C, and 57.5% for L2-D). Considering that there are only 11 verbs in the A- and B-list combined, the repeated use of these verbs (变 *bian* 'become', 考 *kao* 'test', 吃 *chi* 'eat', 说 *shuo* 'speak', 做 *zuo* 'do', 看 *kan* 'look', 过 *guo* 'lead [a life]', 长 *zhang* 'grow', 玩 *wan* 'play, have fun', 发展 *fazhan* 'develop', and 学 *xue* 'study') by L2 learners indicates learners' attention toward prototypical verb choices in VCM/S constructions. Second, on the opposite end of the scale, we notice

[7] The total number of verbs are 423 instead of 424 because in the L2-C group, there is a case of missing verb.

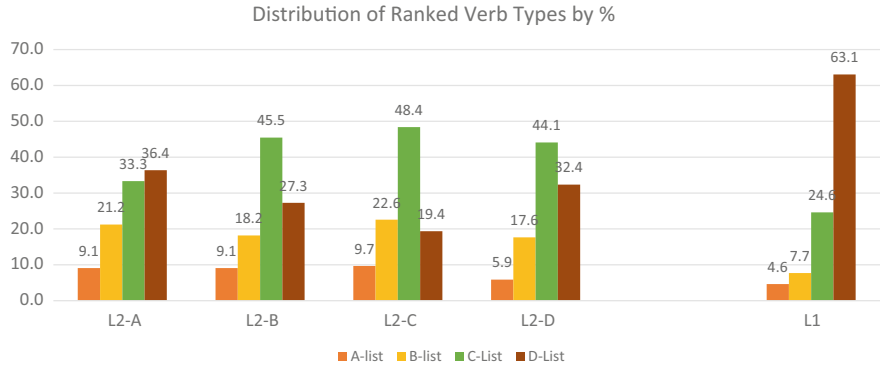

Fig. 2 Distribution of verb types by frequency band

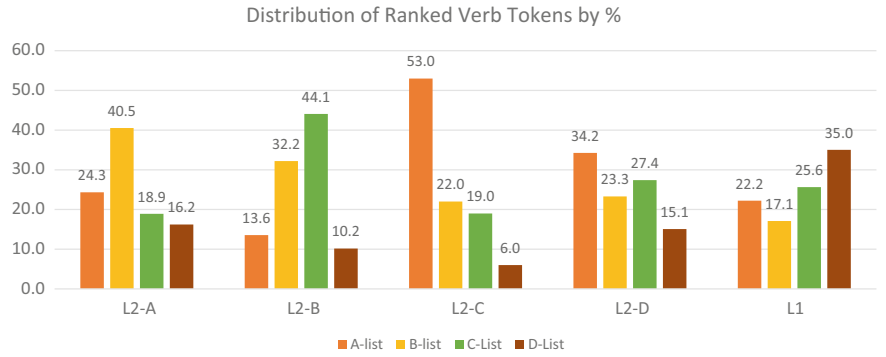

Fig. 3 Distribution of verb tokens by frequency band

that L2 learners' production of D-list verb tokens is almost negligible (16.2% for L2-A, 10.2% for L2-B, and 6% for L2-C). Even L2-D, the highest proficiency group in the corpus, only produced 15.1% worth of D-list verbs.

4.3 Complexity Scales of VCM/S Constructions by L1 and L2 Speakers

While the frequency data of VCM/S constructions and verb choices yielded interesting distributional and developmental patterns, we now turn to the issue of VCM/S complexity based on our proposed VCM/S complexity scales described in Sect. 3.2.

Here, we report three complexity-related results: (a) the VCM/S complexity distribution, (b) VP co-occurrence patterns, and (c) VC form-meaning mapping mechanisms. These three parameters are examined quantitatively and qualitatively as well as contrastively and developmentally.

In terms of the VCM/S complexity distribution, Fig. 4 reveals several useful trends about differences in production complexity. First, there is again a substantial difference between L1 and L2 production in terms of the VCM/S complexity. A chi-square test shows statistically significant between-group differences ($X^2 = 26.2$, $df = 12$, $p < 0.0001$). The L1 data exhibited an even and growing trend in the distribution of the four scales with S(4) VCM/S constructions, making the largest proportion. 59% of VCM/S constructions produced by Chinese L1 speakers belonged to S(3) and S(4), which are high-level complexity VCM/S constructions to convey more sophisticated and nuanced meanings. On the other hand, the L2 learner data showed quite different distribution patterns from L1 speakers as well as among themselves across proficiency levels. Unsurprisingly, L2 learners were generally restricted to S(1) and S(2) VCM/S constructions, especially at the first three levels, with the proportion of S(2) VCM/S constructions being particularly noticeable among intermediate and advanced learners (41.9% for L2-A, 47.5% for L2-B, and 62.4% for L2-C). S(3) VCM/S constructions presented an interesting U-shaped trajectory, which began with a higher percentage of 39.2% for L2-A, and later dipped down to 15.3%, 19.8%, and 16.4% for L2 B, C, and D, then went up to 23.9% for L1 speakers. Such a trend will be further discussed in Sect. 5.

Given the lower numbers in some of the categories (e.g. zero of the S(4) scale in L2-A), Fisher's Exact Test was conducted to identify significant intra- and inter-group differences. The results, as shown in Table 5, confirm two visual impressions obtained from inspecting the chart in Fig. 4: S(2) and S(4) VCM/S constructions seem to show important statistical information for intra- and inter-group differences.

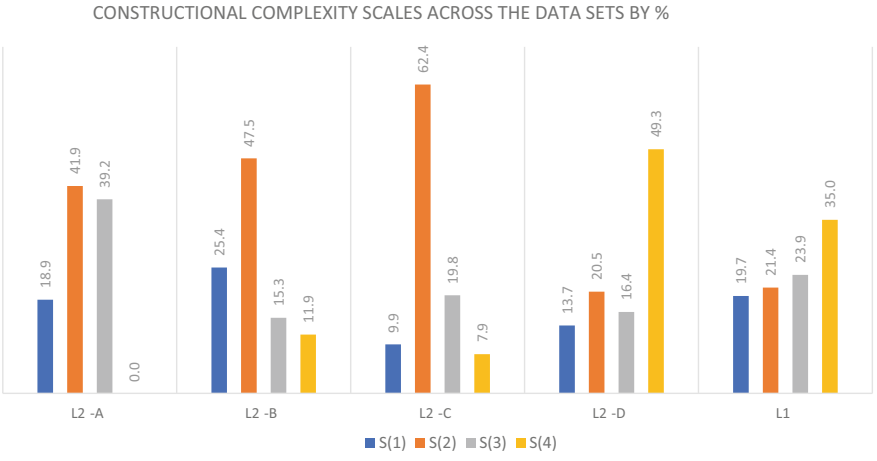

Fig. 4 Distribution of VCM/S constructional complexity scales by percentage

Table 5 Inter- and intra-group differences in VCM/S constructional complexity scales

	L2-A	L2-B	L2-C	L2-D	L1
S(1)	0.612	0.090	0.033	0.495	0.386
S(2)	0.511	0.148	**<0.0001**	**0.001**	**<0.0001**
S(3)	**0.001**	0.137	0.418	0.169	0.798
S(4)	**<0.0001**	0.060	**<0.0001**	**<0.0001**	**<0.0001**

Fisher's Exact Test, *p*-values: values displayed in bold are significant at the level alpha = 0.01.

We take the data to show that while S(2) VCM/S constructions clearly differentiate learners of L2-A, B, and C from L2-D and L1 speakers, S(4) VCM/S constructions further align L2-D learners with L1 speakers in terms of construction complexity.

In terms of VP co-occurrence patterns, our data showed an interesting difference between L1 and L2 speakers in their use of VO and OV in a VCM/S. While L1 speakers in our data used 100% OV form and 0% VO when an object is required, L2 learners' data, however, presented a mixed picture across proficiency levels. L2-A used VO and OV about 50% each with many repeats and errors, L2-B used VO 100%, L2-C VO 80% and OV 20%, and L2-D's pattern was VO 40% and OV 60%. This clearly indicates developmental changes as proficiency increases; L2 learners at higher levels seemed to have gone through a process of structure switching from VO to OV at the VP level. A detailed discussion on the peculiar L2-A data and switching process is offered in Sect. 5.

In terms of VC form-meaning mapping mechanisms, the data revealed a differential preference between L1 and L2 speakers along complexity scales. As described in Sect. 3.2, VCM/S has a variety of ways to map VC forms to their functions, such as the manner of action and the speaker's evaluative stance. Our qualitative analysis indicated that L1 and L2-D speakers preferred to use 5 high complexity types of VC pairings, some are lexical and others are structural mechanisms, to achieve nuanced VCM/S functions, especially for evaluative and affective stances. Examples of VC pairings from our data include (a) complex adjectives as 变得开朗与乐观 *bian de kailang yu leguan* 'become outgoing and optimistic', (b) verb phrases as 长得很像新疆人 *zhang de hen xiang xinjiangren* 'look very much like a Uyghur', (c) idioms as 听得耳熟能详 *ting de er-shu-neng-xiang* 'hear something frequently to the extent that it becomes very familiar', (d) clauses as 风吹得人站不住脚跟 *feng chui de ren zhanbuzhu jiaogen* 'The wind blew so hard that a person could not stand on their feet', and (e) double constructions of VCM/S co-occurring with disposal/passive/causality structures as 你使自己的生活变得丰富有趣 *ni shi zijide shenghuo biande fengfu youqu* 'you made your life both rich and interesting'. Among them, idioms were the most frequently used form by L1 speakers and L2-D learners for subtle meanings of VCM/S constructions. However, lower level L2 speakers (L2-A, -B, and -C) were found to prefer VC pairings within S(1) and S(2) and lexical modifiers for VCM/S function mapping. Examples of VC modifiers from our data include (a) adverbs as 太 *tai* 'too', 非常 *feichang* 'very', 特别 *tebie* 'especially', or 最 *zui* 'most', (b) adverbial phrases as 越来越 *yuelaiyue* 'more and more', 一天比一天 *yitian bi yitian* 'day by

day', and (c) comparative structures as 比…A *bi…A* 'more than', and 跟… (不)一样 *gen…(bu)yiyang* '[not] the same as', among others. As proficiency increased from L2-A to L2-B, and finally to L2-C, these mechanisms gradually expanded to include types with S(3) and S(4) complexity that use structural means to achieve VCM/S functions (as seen in Fig. 4).

To sum up, while L2 learners developed some fundamental abilities to use VCM/S constructions with increased complexity, it takes a long time for them to reach the L1 native level.

5 Discussion

The goal of the current study is to investigate the usage of the Chinese VCM/S construction within usage-based approaches to language acquisition. To our knowledge, this is the first comparative corpus study on construction learning that addresses psycholinguistic factors of frequency, co-occurrence properties, and form-meaning mapping between L1 and L2 speakers. Our discussion will focus on two areas based on our results and our proposed research foci: (a) similarities and differences between L1 and L2 VCM/S production in terms of the VCM/S frequency distribution, verb choices, and construction complexity; and (b) L2 VCM/S developmental patterns.

5.1 Similarities and Differences in VCM/S Production Between L1 and L2 Speakers

Regarding the VCM/S frequency distribution, our results indicate a marked difference between L1 and L2 VCM/S production. L2 learners produced substantially fewer VCM/S constructions than L1 speakers and their average token and type ratio per sample are much lower as compared with L1. This pattern persists across all proficiency levels, and the gap is not closed even when advanced proficiency is reached. Such results can be interpreted in several ways. First, our results of L2 persistent underperformance do not support D. Sun (2002), Feng (2013), and Jiang (2019)'s claim that VCM/S constructions are among the easiest verb complement constructions to learn for L2 learners but provide partial support to Zhou and Deng (2009)'s finding that VCM/S is a complex construction and certain features may not be acquired even if learners reach advanced proficiency. We postulate that different conclusions may be due to different sample sizes used, different analytical foci on VCM/S constructions, such as dynamic statistical information or static grammatical structures, and the interpretation of 'easiness' through limited measures (such as raw frequency and errors) without further examining construction complexity, verb choices, forms, and functions. Second, VCM/S learning is indeed a process of construction learning, whose usage experience and input exposure can lead to

quantitatively and qualitatively different production between L1 and L2 speakers. It is thus important to examine the VCM/S learning from usage-based approaches and to recognize the roles of both implicit (e.g. frequency, form-meaning mapping) and explicit (classroom instruction) learning factors. Additionally, the persistent L2 underperformance signals unique L2 learning needs in areas of systematic language use, expansion of lexical repertoire on construction verbs and complements, targeted attention-directing on nuanced VCM/S form-function mapping, and effective instruction at different stages.

Regarding verb choices, both similarities and differences are observed between L1 speakers and L2 learners. The differences are shown in that L2 learners tend to choose verbs which are limited in number but high in frequency (A- and B-lists), while L1 speakers have a balanced and diverse range of high- and low-frequency verbs but a clear preference for highly specific verbs (C- and D-lists) for their VCM/S constructions. The similarity is seen in L2 learners' repeated use of 变 *bian* 'become' among other 11 high-frequency verbs. First of all, such results point to L2 learners' increasing sensitivity toward and purposeful selection of prototypical verbs in VCM/S constructions as L1 speakers. The fact that prototypical verbs are used by L2-C and L2-D learners corroborates two existing findings in the field of second language acquisition: (a) highly frequent, salient, and prototypical verbs help L2 learners extract shared typical features among learned constructions that eventually help anchor the abstract construction category (Casenhiser and Goldberg, 2005; Childers & Tomasello, 2001); and (b) the prototypicality effect often does not take effect until certain input exposure or proficiency level is reached, such as Intermediate-Mid or higher in our study (Kellerman, 1979; Year & Gordon, 2009). Second, the results provide important L2 evidence to Tao et al. (2020)'s L1 study that shows the verb *bian* is ranked first in frequency and is the most prototypical verb in L1 VCM/S production, accounting for 17% of VCM/S tokens. Third, our data indicate a clear difference in vocabulary range that exists between L1 speakers and L2 learners. L2 learners' overall weakness to use specific verbs for VCM/S constructions and their limited verb choices point to the need for L2 learners to expand their lexical repertoire to convey more nuanced evaluative meanings in more complex VCM/S constructions.

Regarding construction complexity, our results again demonstrate a significant difference between L1 and L2 data measured by a 4-scale complexity scheme proposed by this study. Such a new complexity measure system allows researchers to examine the VCM/S development at both construction and component levels and also to delineate factors in VCM/S complexity distribution patterns, VP co-occurrence patterns, and VC form-function mapping in relation to complexity scales.

With VCM/S complexity distribution, L1 speakers and L2 learners are divided along the complexity scales. While lower level L2 learners tend to adhere to basic S(1) and simple S(2) complexity VCM/S constructions, L1 and L2-D speakers prefer to explore a wide variety of VCM/S constructions with low to high complexity, especially on the high end of S(3) and S(4). Such results uncovered several important findings. First, L2 VCM/S learning is closely linked to VCM/S complexity. Extensive usage and experience as reflected in proficiency levels are the major driving force

in increasing construction complexity (Ellis, 2012). Second, our L2 data analyses on complexity demonstrate that the construction complexity scheme proposed by this study can be used reliably in measuring VCM/S learning. The two high-yield scales are simple S(2) and most complex S(4). S(2) is the best indicator for VCM/S developmental changes at different stages and S(4) signals the gap and necessary alignment between L2 and L1 production. Third, our complexity analyses unveil two unusual and seemingly counter-intuitive VCM/S learning patterns. The first has to do with the U-shaped S(3) data distribution (gray bars in Fig. 4). The second is the L2-C data (orange bars in Fig. 4), which seems to deviate from the rest. We argue that these patterns reflect unique second language developmental stages commonly found in SLA studies (Gass & Selinker, 2013; N. Ellis, 2012). A detailed discussion will be provided in Sect. 5.2, which focuses on developmental patterns.

With VP co-occurrence patterns, an interesting L2 learning pattern emerges, showing differences from L1 speakers. When an object is required in a VCM/S, L2 learners, especially at lower levels, tend to use the VO form to co-occur with its complement, while L1 speakers and L2-D learners prefer the OV form. First, such results match the findings by Zhou and Deng (2009) in that L2 learners whose L1 languages are either alphabetical (Thai, Vietnamese) or non-alphabetical (Japanese and Korean) also tend to use VO in VCM/S constructions at all levels and even at the more advanced levels. Second, both our study and Zhou and Deng (2009) agree that L2 learners' VCM/S learning seems to have started in an unnatural and incorrect order, that is, from a complex VOV form before switching to a simpler and native-like OV form. Grammatically, either VO or OV can co-occur with a complement in a VCM/S, except that VO requires an extra step of verb reduplication into VOV before it can co-occur with other VCM/S components. VOV is thus argued to be more complex than OV structurally. VOV requires a reduplication transformation, whereas OV involves simple movement from postverbal to pre-verbal positions (Zhou & Deng, 2009). Based on our data and observation, we argue that the unnatural VOV usage is a possible result of formal and explicit instruction that needs to be further investigated. As is claimed by Bley-Vroman (1991) and DeKeyser (2003), adult L2 learners of formal classrooms are exposed to two types of input: (a) explicit input from text-books and classroom instruction, and (b) implicit input from language use involving frequency, co-occurrence patterns, and form-meaning mapping processes. Adult L2 learners' VOV preference is likely induced by their explicit formal instruction and textbook exposure. As L2 learners experience more Chinese, they begin to notice the gap between theirs and the native version. This awareness will trigger a restructuring process to help switch the usage to a more native-like version. This restructuring process is in line with SLA studies in Lightbown (1985) and McLaughlin and Heredia (1996).

With VC form-function mapping mechanisms, a differential preference is observed between L1 speakers and L2 learners. While lower level L2 learners like to use low complexity lexical means (adverbs and adverbial phrases) to map VC pairings to VCM/S functions, such as the manner of the action, L1 speakers and advanced L2 learners prefer to use highly complex structural means (idioms, clauses, and combined constructions) for VC pairings to map a full range of nuanced VCM/S

functions. Idioms and combined constructions (the *ba/bei/shi/rang* constructions) are also found to be abundant in L1 production. Such results directly point to several acquisitional implications. First, L1 speakers and L2 learners seem to use different mapping mechanisms (lexical vs. structural) for VCM/S functions. This result is in line with findings by Casenhiser and Goldberg (2005), who discovered that their L2 learners' construction forms first mapped to concrete functions before mapping onto more abstract and subtle functions. Second, our study finds that idiom use is a good indicator for judging whether a VCM/S is intermediate or advanced VCM/S usage. This result also corroborates the findings by Tao et al. (2020) that Chinese native speakers prefer to use complex VC pairings, such as idioms, VPs, and clauses, to convey more subtle and sophisticated evaluative and affective stances.

In summary, our study found similarities but marked differences between L1 and L2 VCM/S production in areas of the VCM/S frequency distribution, verb choices, and construction complexity. These differences point to the fact that VCM/S construction learning is experience-based and learners are sensitive to psycholinguistic factors of frequency, form-meaning mapping, and co-occurrence patterns, but it is also influenced by L2 learners' explicit formal exposure.

5.2 L2 VCM/S Developmental Patterns

The second focus of our discussion is on L2 VCM/S developmental patterns within usage-based approaches to language learning. We address two development-related issues: (a) the L2 VCM/S U-shaped learning patterns, and (b) two unique but independent stages of construction learning: the formulaic stage and the input-induced conservative stage.

As for L2 VCM/S learning patterns, we alluded in Sects. 4.3 and 5.1 that the L2 VCM/S learning process follows a typical U-shaped learning pattern known in the field of second language acquisition. This learning model was proposed and researched mainly by two groups of scholars under different frameworks and at different times.[8] We combine the two approaches to help explain our data. According to Gass and Selinker (2013), U-shaped learning refers to a L2 learning curve across three distinctive stages. The learning normally begins with a high-performance level (Stage 1) and over time it descends to a lower level (Stage 2). After another period of time, the performance once again ascends to a higher level qualitatively (Stage 3). N. Ellis (2012) further expanded the model with new elements and details. He adapted the U-shaped learning model to a three-stage process of construction development, which has distinct and unique stage characteristics. That is from formula to low-scope slot-and-frame pattern, to creative construction (see Fig. 5). Stage 1 is characterized by seemingly high performance in formulaic sequences, Stage 2 is characterized by

[8] Gass and Selinker represent the initial studies of the model within the framework of universal grammar and N. Ellis and others represent the construction learning research under usage-based approaches in recent times.

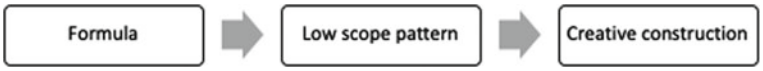

Fig. 5 The three-stage sequence model of construction learning

limited low-scope patterns and open slots to place elements with shared similarities, Stage 3 is characterized by extensive use of constructions creatively and in a wide variety. The transition from one stage to another is motivated by the need of forming and extending construction categories.

We now use the U-shaped construction learning model by Ellis (2012) to explain our L2 statistical data represented by gray bars (Scale 3 complexity) in Fig. 4 of Sect. 4.3. As is shown by the data, the learning curve starts with L2-A learners' high performance of 39.2% S(3) VCM/S constructions, then dips down to the bottom at 15.3% for L2-B, and gradually rises to 19.8% for L2-C, and finally ends at 20.5% for L2-D. We argue that our data reflect exactly a U-shaped VCM/S learning pattern. During this process, L2-A learners start with a rich repertoire of VCM/S formulaic sequences, possibly accumulated from their classroom input and textbooks (Stage 1). Because these formulaic sequences are used with rather high frequency, such as 变得很快 *bian de hen kuai* 'change quickly', and with prototypical verbs, such as *bian* 'change', these sequences quickly gain special statistical status and become concrete similarities (V *de* A) for a prototypical pattern that seeds a VCM/S category. This categorical formation then allows open slots in VP and VC to be substituted with similar elements in form and function, e.g. 变得很多 *bian de hen duo* 'change a lot' or 吃得很快 *chi de hen kuai* 'eat fast', generating different types of low-scope VCM/S patterns (Stage 2). The frequent usage of these low-scope patterns will soon increase in number and be extended to a full range of VCM/S constructions to be used productively and creatively (Stage 3). However, State 3 will last a long time before reaching the native level. We believe that this is the construction learning path that our L2 learners have gone through for VCM/S constructions and have resulted in the production data in our study.

As for the formulaic stage, we attempt to verify if L2-A learners' seemingly high performance on complex S(3) production is full-grown VCM/S constructions or memorized formulaic sequences. Given that L2-A consists of adult learners in formal classrooms and their VCM/S exposure is limited to 2–3 semesters, the complex S(3) VCM/S constructions are theoretically unlikely but in reality exist in the data. We argue that the S(3) production by L2-A is not full-fledged VCM/S constructions but are predominantly formulaic sequences taken from the input. Our evidence comes from three areas. First, the large repertoire of VCM/S constructions produced at this stage is mostly short, repetitive, and not productive. 70% of these VCM/S constructions occur only once, and many are obvious set phrases taken directly from textbooks or classroom instruction. Second, the VCM/S constructions produced by L2-A learners at this stage, though complex on the surface, are rather fixed sequences that are not breakable and cannot generate new forms. For example, 甜食吃得越来越少 *tanshi chide yuelaiyue shao* 'gradually, sweets are consumed less and less' is

a S(3) with a OV + ADV PHRASE structure that is repeated verbatim 7 times. There are no traces of open slots to substitute for new elements in these VCM/S constructions. According to Ellis (2012), the use of open slots is a crucial signal for moving beyond the formulaic stage and forming a construction category. We find S(3) VCM/S constructions are in fact memorized and unanalyzed formulaic strings taken verbatim from the textbooks after verifying from classroom teachers, e.g. the idiom use of 吃得津津有味 *chide jinjinyouwei* 'taste deliciously'. For purpose of further verifying our claim, we call for online empirical studies on VCM/S constructions.

As for the input-induced conservative stage, we attempt to use our L2-C data to explain this unique acquisition stage and its developmental trajectory. According to Goldberg and Boyd (2015), input-induced conservatism refers to a statistical preemption process where learners avoid using certain well-formed but slightly different constructions because similar constructions have been systematically witnessed. We argue that L2-C data on VCM/S frequency distribution, verb choices, and construction complexity mirror exactly this type of conservative stage. As Table 3 and Figs. 2, 3, and 4 indicate, L2-C has the highest VCM/S tokens (101), but these VCM/S constructions carry low types (14) as L2-B. Their verb choices fall on mostly high-frequency ones (75% of A-, B-, and 25% C-, D-lists). Their VCM/S complexity remained predominantly on S(1) basic to S(2) simple scales (72.3%). Clearly, L2 learners at this stage tend to clutch on familiar VCM/S uses and avoid using VCM/S constructions that are different. This seemingly low performance manifests a typical input-induced conservatism according to Goldberg and Boyd (2015) and is found in similar studies by Clark and Clark (1979) and Ellis (2012), where learners resort to limited but familiar VCM/S types and complexity. Ellis (2012) terms this conservatism as 'Teddy Bear phenomenon' and argues that the clutching to 'teddy bear constructions' indicates the learner's efforts to further form and consolidate the construction category. On the other hand, Markman and Gentner (1993) and Goldberg et al. (2007) argue that this is a typical instructed L2 developmental stage with restricted production diversity. Explicit instruction and classroom exemplars tend to lead L2 learners to narrowly focus on their familiar criteria governing the category membership. As L2 learners' proficiency level goes up, they will break this conservatism and expand the construction category widely as L2-D learners did.

To sum up, our findings above contribute to the field in three ways: (a) the U-shape learning exists in VCM/S development and the model by N. Ellis is valid in explaining the construction learning stages and their transitions from one to another; (b) adult L2 construction learning involves both implicit associative learning and explicit classroom instruction, which often dictate the course and characteristics of the formulaic stage and input-induced conservative stage, and (c) construction complexity analyses are the key to understanding construction development, and our proposed VCM/S complexity scale scheme is an important contribution to the field.

6 Conclusions

This comparative study on L1 and L2 Chinese VCM/S production and development contributes to the research on construction learning with three findings: (a) there are marked quantitative and qualitative differences between L1 and L2 VCM/S production at both construction and component levels; (b) these persistent productive differences reflect the indispensable roles of psycholinguistic factors, such as frequency, complexity, form-meaning mapping, and co-occurrence patterns of VP and VC, especially on verb choices; and (c) L2 VCM/S construction learning is like any other construction learning that followings a U-shape learning path that consists of unique and distinctive stages. The process also involves both implicit factors and explicit classroom input and instruction. Such a study has important theoretical and pedagogical implications for the field of second language learning and teaching in general and for Chinese as a second language field in particular.

Theoretically, to our knowledge, this study represents the first attempt that examines the Chinese VCM/S as a construction using usage-based approaches to language acquisition. While existing literature has mainly focused on a single component of VCM/S in isolation, in particular either the VP or VC, this study examined VCM/S acquisition as constructions both at the construction level and at the component level. At the construction level, we proposed a complexity scale of VCM/S constructions that encompasses four major categories and nine formulae. At the component level, we examined the two open slots of VP and VC jointly. This approach not only allowed us to capture the finer path of construction development, from simple to increasing complexity of VCM/S constructions structurally and semantically, but it also provided us a window to observe L2 learners' prototypical choices of VP and co-occurrence patterns on VC in the two open slots, as well as the range of form-meaning mapping mechanisms by L1 and L2 speakers. Second, this study utilized a large-size corpus with both L1 and L2 speakers' production data. The L2 learners' proficiency levels covered a wider range of the spectrum, from Novice-Mid to Advanced-Mid and higher. This allowed us to examine L2 speakers' usage patterns of VCM/S constructions both in comparison with L1 speakers and across developmental stages. The production data supported our proposed categorization and complexity scales of VCM/S as a construction. This study proves that corpus can be used in a fruitful way in examining the L2 acquisition of linguistic constructions in the usage-based paradigm.

Pedagogically, several implications can be drawn from the present study. First, CSL instructors should understand the complexity involved in the CSL learning process of VCM/S construction and should constantly monitor learners' language use experience and encourage learners to notice the possible open slots and their co-occurrence rules for form-meaning pairs with each VCM/S. Second, CSL instructors should use adequate input exemplars with strategic frequency (containing fewer number of construction types but an abundance of tokens) to help learners establish a correct mental category of basic and simple VCM/S constructions early on. Then at a later stage, use diverse input and purposeful attention-direction to expand

the VCM/S category to include many different types of nuanced VCM/S functions. Finally, explicit classroom instruction can lead learners to a possible narrow or excessive focus on a single or irrelevant dimension of the construction. CSL instructors should be more conscious of exposing learners to a wide range of native-like VCM/S construction types and at different complexity scales, especially at advanced levels.

Appendix: List of Verbs of Different Frequency Bands (Type = 117 and token = 423)

A	变	85
	考	23
	吃	22
B	说	17
	做	16
	看	16
	过	13
	长	13
	玩	12
	发展	11
	学	10
C	开	9
	走	9
	写	6
	打	6
	进行	6
	恢复	5
	站	5
	听	4
	唱	4
	表现	4
	跑	4
	弹	3
	搞	3
	活	3
	演	3
	体现	2
	去	2
	发	2

(continued)

(continued)

	安排	2
	弄	2
	惹	2
	扩大	2
	死	2
	涨	2
	熏	2
	爬	2
	睡	2
	管	2
	经过	2
	耐	2
	锻炼	2
	骂	2
D	买	1
	争	1
	住	1
	决定	1
	准备	1
	刮	1
	到	1
	刷	1
	刻画	1
	办	1
	包	1
	反应	1
	变化	1
	叫	1
	吓	1
	吹	1
	呕吐	1
	呛	1
	困	1
	失败	1
	定	1
	害	1
	差	1
	帮	1

(continued)

(continued)

	干	1
	心-信	1
	忙不迭	1
	想	1
	想象	1
	战	1
	打扫	1
	打扮	1
	扫	1
	折	1
	抛	1
	捏	1
	接近	1
	摔	1
	旅游	1
	显	1
	晒	1
	来	1
	梳	1
	模仿	1
	洗	1
	演戏	1
	烘	1
	生	1
	生活	1
	用	1
	画	1
	留	1
	皱	1
	相处	1
	穿	1
	笑	1
	管理	1
	绽放	1
	落	1
	行	1
	表演	1
	装点	1

(continued)

(continued)

	见	1
	订	1
	讲	1
	读	1
	谈谈	1
	起	1
	跳	1
	踮	1
	蹭	1
	运算	1
	进步	1
	适应	1
	逼	1
	飞	1
	骑	1

References

ACTFL. (2012). ACTFL Proficiency Guidelines 2012. Retrieved from https://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf.

Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*(1), 67–82.

Bybee, J., & Hopper, P. J. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. John Benjamins.

Bley-Vroman, R. (1991). The logical problem of foreign language learning. *Linguistic Analysis, 20*(1–2), 3–49.

Casenhiser, D., & Goldberg, A. E. (2005). Constructional fast mapping. *Berkeley Linguistic Society*.

Chao, Y. R. (1968). *A Grammar of spoken Chinese*. University of California Press.

Childers, J. B., & Tomasello, M. (2001). The Role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology, 37*(6), 739–748.

Clark E. V., & Clark, H. (1979). When nouns surface as verbs. *Language, 55*(4), 767–811.

Clifton, C., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 23*, 696–708.

DeKeyser, R. (2003). Implicit and explicit learning. In J. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 313–348). Oxford, MA: Blackwell.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics, 32*, 17–44.

Ellis, N. C., & Larsen-Freeman, D. (2009). Language as a complex adaptive system (Special Issue). *Language Learning, 59*(Supplement 1).

Ellis, N. C. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 end state. In P. Robinson & N. C. Ellis

(Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372–405). New York & London: Routledge.
Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics, 27*(1), 1–24.
Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188.
Feng, L. 冯丽莉. (2013). "De Zi Buyu Leixing de Xide Nandu yanjiu" 得字补语类型的习得难度研究 (Research on learning difficulties of *De* complement constructions). Shanghai Normal University MA Thesis.
Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language, 80*, 748–775.
Garnsey, S., Pearlmutter, N., Myers, E., & Lotocky, M. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*, 58–93.
Gass, S., & Selinker, L. (2013). *Second Language Acquisition*. New York & London: Routledge.
Goldberg, A., & Boyd, J. (2015). A-adjectives, statistical preemption, and evidence: Reply to Yang (2015). *Language, 91*(4), 184–187.
Goldberg, A., E., Casenhiser, D., & White, T. (2007). Constructions as categories of language. *New Ideas in Psychology, 25*, 70–86.
Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences, 7*(5), 219–224.
Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago & London: University of Chicago Press.
Hoey, M. P. (2005). *Lexical priming: A new theory of words and language*. London & New York: Routledge.
Jiang, T. 姜天琦. (2019). "Qingtai Buyu de Xide Yanjiu" 情态补语的习得研究 (Research on complement constructions of manner). Yangzhou Normal University MA thesis.
Kellerman, E. (1979). Transfer and non-transfer: Where we are now. *Studies in Second Language Acquisition, 2*, 37–57.
Li, C., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley, Los Angeles, & London: University of California Press.
Lightbown, P. (1985). Great expectations: Second language acquisition research and classroom teaching. *Applied Linguistics, 6*, 173–189.
Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*(4), 431–467.
McLaughlin, B., & Heredia, R. (1996). Information-processing approaches to research on second language acquisition and use. In W. Ritchie & T. Bhatia (Eds.), *Handbook of Second language acquisition* (pp. 213–228). Academic Press.
Shen, J. (2003). Xiandai Hanyu dongbu jiegou de leixingxue kaocha [A typological investigation of verb complement constructions in modern Chinese]. *Shijie Hanyu Jiaoxue, 2003*(3), 17–23.
Sun, D. 孙德金. (2002). "Waiguo Liuxuesheng Hanyu '*De*' Zi BuyuXide Kaocha" 外国留学生汉语"得"字补语句习得情况考察 (An investigation on CFL learners' acquisition of 'DE' complements). 语言教学与研究 *Yuyan Jiaoxue yu Yanjiu, 6*, 42–50.
Sun, Q. 孙群. (2018). Oumei Xuesheng Hanyu Qingtai Buyu Ju Xide Yanjiu 欧美学生汉语情态补语句习得研究 (L2 acquisition on Chinese complement constructions of manner by CSL learners of European languages). 海外华文教育 *Haiwai Huawen Jiaoyu,* 2018.101/6. 59–67.
Tao, H., Jin, H., & Zhang, J. (2020). A corpus-based investigation of manner/state complement constructions in Mandarin Chinese. To appear in Bianca Basciano, Franco Gatti, and Anna Morbiato (Eds.), *Corpus-based research on Chinese language and linguistics, Sinica venetiana* (pp. 1–40). Edizioni Ca' Foscari Digital Publishing, Venice, Italy: Università Ca' Foscari Venezia (Ca' Foscari University) Press.

Trueswell, J., & Tanenhaus, M. (1994). Toward a lexical framework of constraint-based syntactic ambiguity resolution. In C. Clifton & L. K. R. Frazier (Eds.), *Perspectives on sentence processing* (pp. 155–179). Erlbaum.

Trousdale, G., & Hoffmann, T. (Eds.). (2013). *Oxford handbook of construction grammar*. Oxford: Oxford University Press.

Tily, H., Gahl, S., Arnon, I., Kothari, A., Snider, N., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language & Cognition, 1*, 147–165.

Year, J., & Gordon, P. (2009). Korean speaker's acquisition of the English ditransitive construction: The role of verb prototype, input distribution, and frequency. *The Modern Language Journal, 93*, 399–417.

Zhang, J. (2011). *Acquisition of the Chinese resultative verb complements by learners of Chinese as a foreign language: A learner corpus approach.* Unpublished doctoral dissertation, The Pennsylvania State University.

Zhou, X. 周小兵, & Deng, X. 邓小宁 (2009). Liangzhong 'De' Zi Buyuju de Xide Kaocha (An investigation on two types of complement constructions of manner). 汉语学习 *Hanyu Xuexi, 2*, 65–71.