This is the Pre-Published Version.

This is the accepted version of the publication Ke, S., Jin, R., & Koda, K. (2023). Adult L2 learners' morphological sensitivity in a morphosyllabic language. The Mental Lexicon, 18(3), 446-471. The Version of Record is available online at: https://doi.org/https://doi.org/10.1075/ml.22004.ke.

Adult L2 Learners' Morphological Sensitivity in a Morphosyllabic Language

Sihui (Echo) Ke, The Hong Kong Polytechnic University

Rui Jin, Shenzhen University

Keiko Koda, Carnegie Mellon University

Abstract

This study examined adult L2 learners' morphological sensitivity in Chinese, a morphosyllabic language, and explored whether there is any modulating effects of L2 proficiency. Two word naming experiments (segment shifting and standard word naming) were administered to three participant groups, including native Chinese speakers, higher L2 Chinese proficiency learners, and lower L2 Chinese proficiency learners. In both experiments, reaction times (RTs) displayed only main effects of Chinese proficiency group and word type. This suggests that the morphological processing of L2 learners did not differ from that of native speakers, although the RTs of L2 learners were longer and exhibited more variability. Concerning error rates, both experiments showed that learners with higher and lower L2 proficiency had significantly higher error rates for words with unreliable morphological cues compared to those with reliable cues. Taken together, these findings indicate that L2 learners developed sensitivity to intraword morphological structure and employed decompositional strategies when reading Chinese words, irrespective of their L2 proficiency levels.

Keywords: morphological processing, Chinese, word naming, second language proficiency

INTRODUCTION

During the past decade, there has been a debate over late (adult) second language (L2) learners' processing of morphologically complex words (for a recent review, see Feldman & Kroll, 2019), which has mainly focused on whether learners who acquire an L2 post-puberty can develop morphological sensitivity to the distributional properties of morphologically complex words (e.g., affix productivity and morphological family size), and decompose complex words as native speakers do. This line of investigation is of both theoretical and pedagogical significance: theoretically speaking, research of the storage and processing of morphologically complex words can lead to insights of human mental lexicon (Embick et al., 2021); pedagogically speaking, the development morphological knowledge has been found to be dynamic and challenging for L2 learners (Larsen-Freeman, 2010), thus, empirical evidence of factors that influence L2 morphological sensitivity could provide implications for classroom intervention.

To date, some researchers have found evidence that supports decomposition in L2 morphological processing and a similar mechanism in both native language and L2 groups (e.g., Coughlin & Tremblay, 2015; Diependaele et al., 2011; Feldman et al., 2010); some have held a contrasting position and provided counter-evidence for L2 learners' insensitivity to intraword morphological structure (e.g., the Shallow Structure Hypothesis, Neubauer & Clahsen, 2009; Silva & Clahsen, 2008); still others argue that it depends on a range of factors including the properties of the first (L1) and second (L2) languages and the relationship between the two (e.g., Vainio et al., 2014), as well as L2 learners' characteristics such as length of residence, age of arrival and sex (e.g., Babcock et al., 2012), and L2 proficiency (e.g., Deng et al., 2016; Gor, 2017; Liang & Chen, 2014), just to name a few. Notably, the majority of previous research has examined English as the target L2, recruited advanced L2 learners whose L1 and L2 are both alphabetic, and designed behavioral lexical decision tasks to tap into adult L2 learners (in)sensitivity to morphological structure with *inflected* words as stimuli. There is little evidence with respect to how L2 learners process morphologically complex words in Chinese, a non-alphabetic language whose main word formation rule is compounding (Arcodia, 2012; Chinese virtually has no inflection according to Li et al., 1993), and to what extent it is affected by learner- and experimental-task-related factors.

Against this background, the present study investigated whether adult L2 learners develop intraword morphological sensitivity with English-speaking adult L2 learners of Chinese, a morphosyllabic language (DeFrancis, 1989), which shows syllable-to-character mapping. This study has two aims: one was to examine morphological decomposition in L2 visual word processing in Chinese, a nonalphabetic written language; the other was to explore the moderating effects of L2 proficiency and task. Two word-naming experiments (i.e., segment shifting and standard word naming) were administered in three groups (16 native speakers, 20 higher-proficiency L2 learners, and 19 lower-proficiency L2 learners). It is anticipated that the findings will contribute to a crosslinguistically valid account of the mechanism of L2 morphological processing.

Morphological decomposition in L2 word processing: Evidence from Chinese

Chinese (Mandarin) is distinctive due to its morphosyllabic writing system and the prevalence of morphologically complex words (for instance, approximately 72% of Chinese words are two-character compound words, according to the Lexicon of Common Words in Contemporary Chinese, 2008). It is widely accepted that native Chinese speakers decompose these complex words into constituents during reading (Crepaldi et al., 2012). A common research approach to gauge morphological sensitivity in native Mandarin speakers involves recording their reaction times and accuracy rates in visual lexical decisions in response to actual Chinese words, pseudowords, and nonwords (e.g., Gao et al., 2022). Gao et al. (2022) suggested that the performance differences between real words or pseudowords and nonwords reflect semantic effects, while the differences between pseudowords and real words alone indicate morphological sensitivity. Their study found that compared to real words, pseudowords elicited slower and more erroneous responses, suggesting that native Mandarin speakers are sensitive to morphological constraints and may experience difficulty in structural parsing. The question, to date, remains as to whether L2 Chinese learners adopt a similar processing mechanism. Although a substantial number of studies have been conducted to examine L2 Chinese visual word processing, the majority of interest lies in orthographic processing (e.g., Shen & Ke, 2007; Wang, Liu & Perfetti, 2004; Xu et al., 2014; Yang, 2000), and there are only a limited number of studies that have paid attention to morphological processes (e.g., Chen, 2018; Ke & Koda, 2017, 2019; Maeng & Kim, 2023; Zhang, 2017).

Ke and Koda (2017) examined morphological sensitivity in American university L1 English-L2 Chinese learners with a L2 Chinese segment shifting task designed after Feldman et al. (1995). Learners were asked to shift an orthographic segment (i.e., a character) from a three-character multisyllabic word, combine the segment with a two-character base word, and then name a new three-character multisyllabic word. It was found that L2 Chinese learners' efficiency was higher for the suffixoid condition than the prefixoid condition, followed by the nonaffixoid condition, which was similar to L1 English segment shifting performance (i.e., the participants were most efficient shifting suffixes, followed by prefixes, and least efficient in shifting nonaffixes). In addition, Ke and Koda (2017) found that, via regression analysis, L1

English morphological sensitivity significantly predicated L2 Chinese morphological sensitivity over and above L2 linguistic knowledge. But no comparison group (L1 Chinese group) was included in Ke and Koda's (2017) study.

Recently, Maeng and Kim (2023) examined whether Korean-speaking L2 Chinese speakers are able to construct hierarchically structured representations of morphologically complex words as efficiently as L1 Chinese speakers. The results from a mask-primed lexical decision task suggested that L2 speakers displayed similar priming patterns to L1 speakers for morphologically related prime-target pairs. However, the L2 participants also demonstrated semantic and orthographic priming effects, which were not observed in the L1 group. Maeng and Kim inferred that their findings support the Shallow Structure Hypothesis, as L2 Chinese participants heavily relied on semantic and orthographic cues when processing morphologically complex words in Chinese. Nonetheless, they did not account for the Korean-speaking participants' character/hancha cognate knowledge. Moreover, akin to previous studies of alphabetic languages, they utilized a lexical decision task instead of a word naming task.

To our knowledge, Zhang (2017) might be the only study that investigated the role of morphology in both L1 and L2 Chinese word naming. To be specific, Zhang compared the performance between L1 and L2 Chinese word naming accuracy among fourth graders in Singapore. The participants were asked to read a list of 20 single-character and 20 two-character words printed on cards. According to Zhang (2017), the L1 readers were those who reported using Mandarin as the home language for communication with both parents; the L2 readers were those who reported using only English as their home language. The two groups behaved differently, as L1 Chinese readers relied more on morphological processing in word reading while L2 Chinese readers, who spoke English as their L1 and received four years of formal

Chinese instruction, relied more on phonological processing in standard word naming. According to Feldman and Prostko (2002), existing morphological processing experiments bear different task demands (e.g., more explicit morphological structure manipulation in the segment shifting task versus less explicit morphological structure manipulation in the standard word naming task). It is possible that the segment shifting task in Ke and Koda (2017) tapped more explicit morphological processing in adult L2 learners whereas the standard word naming task used in Zhang's (2017) study elicited more implicit and automatized response in bilingual children. Nevertheless, morphological processing task demand was not the research focus of Ke and Koda (2017) or Zhang (2017).

There might be doubt that Zhang's (2017) study focused on child English-Chinese bilingual learners and the findings might not be directly applicable to the research of adult L2 learners. A pertinent study was conducted by Chen (2018), in which he measured morphological awareness in paper-and-pencil tasks in more skilled and less skilled adult L2 Chinese learners studying abroad in mainland China. More or less skilled learner grouping was based on the results of a standardized test—The Unified Comprehensive Chinese Examination of the Preparatory Education for International Students in China (Wang et al., 2016). Chen (2018) found that the more-skilled learners could take greater advantage of their sensitivity to morphological cues in lexical inferencing whereas the less killed learners could not use their morphological sensitivity to infer the meanings of unknown words. In view of the contrasting evidence from the three studies (Chen, 2018; Ke & Koda, 2017; Zhang, 2017), there is a niche for further research of L2 Chinese morphological processing based on different experimental tasks.

To sum up the review above, two major gaps remain in existing literature pertaining to L2 morphological processing and its (in)decomposition mechanism. First, there is an Anglocentric trend in previous research (see also Frost, 2012; Share, 2008, 2021), and very few studies on a non-alphabetic L2 such as morphosyllabic Chinese (an exception is Maeng & Kim, 2023). Second, it is still unclear as to how learner-related factors (e.g., L2 proficiency) and experimental-task-related factors might affect L2 morphological processing. As mentioned earlier, prior studies yielded contrasting findings about the effect of L2 proficiency. Also, they predominantly used masked or unmasked lexical decision tasks instead of word naming tasks.

THE CURRENT STUDY

The primary goal of this study was to examine whether adult L2 learners with limited exposure to the target language (about three years of formal foreign language education) can develop morphological sensitivity in a morphosyllabic L2 (Chinese), which is typologically distinct from their L1 (English). Native Chinese speakers were also recruited for the research. Three hypotheses were generated: (1) Native Chinese speakers decompose morphologically complex words into their constituents in word naming, regardless of task demands. (2) Adult L2 Chinese learners adopt similar decompositional processing strategies to native Chinese speakers, yet there is a graded morphological effect in L2 Chinese word processing subject to the influence of L2 proficiency. In other words, adult learners of higher proficiency develop morphological sensitivity in L2 Chinese whereas adult learners of lower L2 Chinese proficiency do not; (3) Higher L2 proficiency learners' morphological sensitivity is subject to the influence of task demands—they are more sensitive to intraword morphological structure in a task that involves more explicit structural manipulation (i.e., segment shifting) than a task that taps more implicit morphological processing (i.e., standard word naming).

GENERAL METHOD

Participants

Sixteen native Chinese speakers were recruited from a major university in Shanghai, China. Their age range was between 18 and 23 years old. They were all females.

In addition, 39 L2 Chinese learners who had studied Chinese formally for about three years were recruited from universities in the U.S. They had no knowledge of Chinese prior to university study. The age range was between 18 and 21 years old. There were 17 males and 22 females. The 39 learners were furthered categorized as higher L2 proficiency learners (*n* = 19) and lower L2 proficiency learners (*n* = 20) via a two-step cluster analysis of their scores gathered from a paper-and-pencil Chinese vocabulary knowledge test (adopted from Liu, 2013; cronbach's alpha = .92). The cluster quality was good based on Sihouette measure of cohesion and separation (see a similar approach in Chen, 2018). Description of the vocabulary knowledge test and participants' performance can be found in Appendix A.

Method

This study included two computerized experimental tasks, including segment shifting and standard word naming (as described below), a paper-and-pencil vocabulary knowledge test, as well as a language background questionnaire. The two experimental tasks were administered first and randomized across participants. The paper-and-pencil vocabulary knowledge test was distributed to the participants afterward, followed by the questionnaire. The tasks and the questionnaire were administered individually by the first author in a quiet room on the university campuses. The time to complete all tasks took approximately 45 minutes.

EXPERIMENT 1: SEGMENT SHIFTING

The segment shifting task was designed after Feldman et al. (1995). We treated data gathered from higher L2 proficiency learners as the baseline data of this experiment, and expected to find a pattern attuned to Chinese, a morphosyllabic language: Faster reaction times and higher accuracy rates for affixoid[1] shifting than for nonaffixoid shifting in Chinese. It was hypothesized that compared with higher L2 proficiency learners, it takes a shorter time for native speakers to react in this task, but that the reaction times are longer in lower L2 proficiency learners (for L2 proficiency modulating effect, see Chen, 2018; Miller & Koda, 2018), and that, when compared with higher L2 proficiency learners, the native speakers' accuracy rates are higher, and lower L2 proficiency learners are less accurate.

Materials

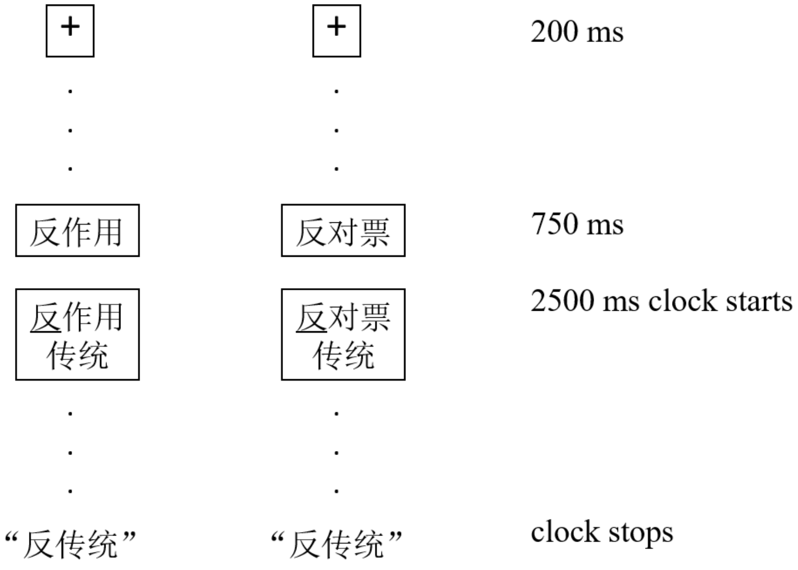
The experimental items included 16 source words formed with affixoids, 16 source words formed with nonaffixoids, and 16 target words (see Appendix B in online supplementary materials). Two types of three-character source words (i.e., formed with affixoids versus nonaffixoids) were constructed for this task (see item selection procedures in Appendix C in online supplementary materials). The affixoid condition consisted of one of the 16 affixoids and one of semantically opaque two-character words in the item pool (e.g., the character 反 corresponds to a prefixoid indicating anti/counter in the three-character word 反作用, fǎn-zuòyòng, counteraction) while the nonaffixoid items (e.g., 反对票, fǎnduì-piào, veto) were formed by combining a semantically opaque two-character words containing the same 16 characters/orthographic forms as their components (e.g., 反, fǎn, opposite in 反对, fǎnduì, to *object*) with another single-character word (e.g., 票, *piào*, *tickect*). The mean frequency (occurrences per million words in the SUBLEXCH corpus, see Cai & Brysbaert, 2010) was 5.88 for source words formed with affxioids, 2.03 for source words formed with nonaffixoids, and 74.80 for target words.

Procedure

This segment shifting experimental task was adapted from Feldman et al. (1995), controlled by *Presentation* (see Neurobehavioral Systems, 2020), and administered to individual participants in Experiment 1. In Feldman et al.'s (1995) study, the participants were first presented with an affixed word in upper case (e.g., LEADER), asked to strip the affix (e.g., ER) from the word, and attach the stripped affix to another word (e.g., KICK). The participants then were asked to name the resulting word (KICKER) aloud as quickly as possible. In this study, we have adapted the task with Chinese-specific items. Each individual participant was asked to detach a designated character/orthographic segment (e.g., 反 *fǎn*, *anti/counter*) from the source word (e.g., 反作用, *fǎn-zuòyòng*, *counteraction* or 反对票, *fǎnduì-piào*, *veto*) and attach the character to the target word (i.e., 传统 *chuántǒng*, *tradition*). The participant then had to name the resulting word (i.e., 反传统 *fǎn-chuántǒng*, *counter-tradition*) aloud as rapidly as possible. The target character (orthographic segment to be shifted) has one pronunciation only, regardless of the experimental conditions.

In the task (as shown in Figure 1), the stimuli were presented in 38-point fixed width font (*Simsun*) on a computer screen. For each trial, first, a fixation signal "+" was presented at the center of the screen, followed by the onset of a source word after 200 milliseconds (ms) at the same place. The source word was presented for 750 ms to ensure lexical access. Thereafter, the target word appeared below the source word. The participants were instructed to name the resulting word by combining the underlined segment and the target word. The source word and target word remained on the screen for 2500 ms; the fixation signal then appeared to indicate the beginning of the next trial. A lapse between the onset of the target word's presentation and the participant's voice onset was measured in ms together with oral response accuracy.

Figure 1.
The segment shifting procedure.



Data analysis

Observations with response times with inaccurate or missing recordings, or with response times greater than 2500 ms, were removed. Response times and error rates were analyzed with via R.

Gaussian Location-Scale Models and Linear Mixed Models were used for the analysis. The performance was then compared between two source word segment conditions (i.e., affixoid and nonaffixoid), both of which share an orthographic segment (e.g., 反 fǎn). If participants responded more slowly and made more errors when shifting source words formed under the nonaffixoid condition (e.g., 反对票, fǎnduì-piào, veto), they were considered to be sensitive to the intraword morphological structure (see also Ke & Koda, 2017). The models included reaction times (RTs) and error rates (ERs) as the respective dependent variables (see Models 1 and 2 specified in Table 1). Model 1 incorporated the following fixed variables: source word condition (affixoid versus nonaffixoid), Chinese proficiency group (native versus higher L2; lower L2 versus higher L2), base word frequency, and initial phoneme frequency. Model 2 included two additional fixed factors: whole word frequency and character family size. Furthermore, Model 1 incorporated participants and items as random intercepts. This allowed the model to associate random effects with items that differed across various participant groups, thereby capturing how group characteristics influenced item responses. It should be noted that only significant or marginal effects are reported in the next section. The complete data set as well as statistical analysis results of Experiment 1 are available at: osf.io/gqh85.

Table 1.

Gaussian Location-Scale Model and Generalized Linear Mixed Model of Participant Responses to Experiment 1.

Model	Dependent variable	Model specification
1	Response time (RT)	gam(list(Resp_time ~ morphology +group+LogBaseFreq +LogInitPhonFreq +s(subject, bs=”re”) +s(item, group, bs=”re”), ~ morphology+group+LogBaseFreq), family=”gaulss”, data=dat3)
2	Error rate (ER)	glmer(error~morphology*group+LogFreq+LogBaseFreq+ LogFamSize+LogInitPhonFreq+(1|subject) + (1|item), data=filter(dat), family=binomial)

Note. Model 1 utilized the *gam()* function from *mgcv* package, while Model 2 employed the *glmer* function from *lme4* package with the specification of “family=binomial.” morphology, source word condition; group, native/higher L2/ lower L2 group; LogFreq, log character frequency; subject, participants; item, 32 items.

Results

The descriptive statistics of Experiment 1 are shown in Table 2.

Table 2.

Descriptive Statistics of the Segment Shifting Task

Chinese proficiency	Source word segment condition	Response time (in ms)			Accuracy rate		
		Mean	SD	95% CI	Mean	SD	95% CI
Native	Affixoid	771.96	215.72	742.55, 801.38	0.82	0.38	0.77, 0.86
	Nonaffixoid	873.05	289.59	833.66, 912.45	0.82	0.39	0.77, 0.87
Higher L2	Affixoid	1013.31	372.62	969.15, 1057.46	0.77	0.42	0.73, 0.82
	Nonaffixoid	1034.10	375.95	986.98, 1081.21	0.87	0.34	0.83, 0.90
Lower L2	Affixoid	1008.51	359.30	961.73, 1055.30	0.76	0.43	0.72, 0.81
	Nonaffixoid	1047.34	412.30	987.03, 1107.64	0.59	0.49	0.54, 0.65

Table 3 presents the results of Gaussian Location-Scale Models with RTs as dependent variables: there was a significant main effect of source word segment condition (i.e., affixoid versus nonaffixoid) (estimate = −70.891, *p* = .0011) and a significant effect of base word frequency (estimate = −17.892, *p* = .0211). Although higher L2 proficiency group responded more slowly as compared to the native speaker group (estimate = −216.239, *p* = .003), and higher L2 proficiency groups’ RTs were not significantly different from lower L2 proficiency group’s (estimate = 31.029, *p* = .6008), two-way interactions between word- and learner-group related factors were not supported by the data and thus not reported in Table 3. Moreover, the variance was significantly lower for the affixoid source condition. In other words, when reliable morphological cues were available, the participants were able to shift the segments with reduced variability. Taken together, the findings based on the analyses of the RT data suggested that the native, higher L2 and lower L2 groups responded similarly and decomposed morphologically complex words in Experiment 1.

Table 3.

Results of the Effects between Source Word Segment Condition and Chinese Proficiency on Segment Shifting Response Times.

Model 1	Estimate	SE	*z* value	*p* value
Intercept	803.587	192.705	4.170	<.0001
Source word segment condition (affixoid vs. non affixoid[a])	−70.891	21.716	−3.264	.0011
Lower L2 vs. Higher L2	31.029	59.311	0.523	.6008
Native vs. Higher L2	−216.239	59.794	−3.616	.0003
LogBaseFrequency	−17.892	7.756	−2.307	.0211
LogInitialPhonemeFrequency	31.14	16.828	1.850	.0642
Intercept.variance	5.561	0.359	15.479	<.0001
Source word segment condition.variance[b]	−0.155	0.041	−3.756	.0002
Lower L2.variance[c]	0.014	0.047	0.286	.7750
Native.variance[c]	−0.500	0.048	−10.473	<.0001
LogBaseFrequency.variance	−0.030	0.014	−2.071	.0384
LogInitialPhonmeFrequency.variance	0.042	0.032	1.313	.1892

Note. Est., estimate; NSs, native speakers. [a], the nonaffioxid condition was treated as the refence; [b], the nonaffixoid condition was treated as the reference; [c], the higher L2 group was the reference group.

The results of the generalized linear mixed effects modeling for error rates (ERs) are presented in Table 4. Model 2 results indicated that there was a significant interaction effect between Chinese proficiency group and word condition: (1) within the higher L2 groups, participants made more errors in the nonaffixoid condition (estimate= −0.785, *p* = .0047). (2) When we compared the three groups' error rates for the affixoid condition, we found that the lower L2 learners made more errors than higher L2 proficiency speakers (estimate = 0.997, *p* = .0029), but the error rates of the higher L2 learners were not significantly different from those of the native speakers (estimate = −0.403, *p* = .2701). (3) There was no main effect of Chinese proficiency on the nonaffixoid condition, so these results were not reported.

Table 4.

Results of the Effects between Source Word Segment Condition and Chinese Proficiency on Segment Shifting Error Rates.

Overall model	Estimate	SE	z value	P value
Intercept	−2.007	1.928	−1.041	.2980
Source word segment condition (affixoid vs. nonaffixoid in higher L2[a])	−0.785	0.278	−2.284	.0047
nonAffixoid lower L2 vs. nonAffixoid higher L2	0.997	0.366	2.982	.0029
nonAffixoid native vs. nonAffixoid higher L2	−0.403	0.366	−1.101	.2701
LogCharacterFrequency	0.376	0.148	2.536	.0112
LogBaseFrequency	−0.244	0.074	−3.277	.0010
LogFamilySize	−0.234	0.197	−1.188	.2349
LogInitialPhonemeFrequency	0.218	0.171	1.276	.2020
Affixoid*lower L2 (vs.nonAffixoid*higher L2)	−0.232	0.295	−0.786	.4316
Affixoid*native (vs. nonAffixoid*higher L2)	0.762	0.330	2.313	.0207

Note. Est., estimate; NSs, native speakers. [a], The nonaffixoid condition was treated as the reference.

To compare the findings of the segment shifting task against our hypothesis formulated above (i.e., in segment shifting, native speakers and higher L2 learners show sensitivity to intraword morphological structure and respond less rapidly and make more errors in to words formed with nonaffixoids, whereas lower L2 learners are insensitive to different source word conditions), our hypothesis was partially confirmed, and there were two major findings: (1) Overall, although the native group responded significantly faster than the higher and lower L2 groups, we did not identify any interaction effect between source word condition and Chinese proficiency. Notably, all three groups (native, higher L2, and lower L2) demonstrated sensitivity to the intraword structure of morphologically complex words. Their reaction times (RTs) in the segment shifting task were significantly shorter for the affixoid condition than the nonaffixoid condition. Also, there was a significant effect of base word frequency on segment shifting RTs, which suggested that all three groups adopted decomposition strategies in morphologically complex word processing. (2) The lower L2 group produced significantly more errors than the native and higher L2 groups, which was not surprising. The higher L2 group did not differ significantly from the native group in terms of errors rates, and they produced significantly fewer errors in affixoid shifting compared to nonaffixoid shifting. Therefore, the analyses of both reaction time and error rate data suggested that L2 learners developed sensitivity to the internal morphological structure of words, regardless of their L2 proficiency levels.

EXPERIMENT 2: STANDARD WORD NAMING

As mentioned earlier, the participants of Experiment 1 also completed Experiment 2.

Materials

Four sets of word stimuli, each consisting of 16 items, were used in this task, with a total number of 64 items (see Appendix D in online supplementary materials). One set was comprised of three-character legally formed pseudowords generated by combining the affixoids with two-character base words. For example, 高兴度, *gāoxìngdù*, *happiness* or *delightfulness*) contains 高兴, *gāoxìng*, *happy* and the suffixoid 度 *dù*, roughly equivalent to the agent noun suffix *-ness* in English). This set of words was used to establish the base-line levels of word naming efficiency. A second set included 16 three-character illegally formed words, consisting of two-character base words and the 16 affixoids appearing at illegal positions (i.e., prefixoids are placed at the end of the multi-character string and suffixoids at the beginning of the string). A third set was comprised of three-morpheme three-character unrelated character strings (e.g., 爱常爸, *ài cháng bà*, meaning *love*, *often*, and *father* respectively). A fourth and final set of words included 16 two-character real words (e.g., 世界, *shìjiè*, *world*). The mean frequency counts of the two-character words (occurrences per million words in the SUBLEXCH corpus, Cai & Brysbaert, 2010) were 289.61, 283.48 and 284.10 in the first, second and last sets respectively. All characters were selected from Bands One and Two – the two lowest levels of GSCVCC (Chinese Proficiency Test Center, 2001).

Procedure

Experiment 2 was also controlled by *Presentation*. All stimuli were presented individually on a computer screen. The presentation order was randomized across participants. For each trial, first, a fixation signal "+" appeared at the center of the screen for 200 ms, followed by the stimulus which the participants read aloud. The stimulus disappeared after 2000 ms, followed by the fixation point to signal the beginning of a new trial. The participants were told to be as accurate as they could. There were four practice trials. Both reaction time (defined as the amount of time between the onset of a stimulus item and the onset of voice) for correct items and accuracy were recorded.

Data analysis

The analysis procedures were similar to Experiment 1. The time between the onset of the stimuli's presentation and the participant's voice onset was measured in ms. We also analyzed their oral response accuracy. Observations with response times with inaccurate or missing recordings were removed; items with response times greater than 2000 ms were removed as well. RTs were analyzed along with error rates. We anticipated that the differences in reaction times (RTs) and error rates (ERs) between illegally formed words and legally formed words, as well as between three-character strings and legally formed words, would suggest a morphological effect. Furthermore, it was hypothesized that, in a standard word naming task that involved more implicit morphological processing, morphological effects would be observed among native speakers the idea being that they would respond more slowly and would make more errors to illegally formed words than legally formed words, whereas no significant differences in higher or lower L2 learners' responses to the three types of words are expected. Gaussian Location-Scale Model and Linear Mixed-effects Model were implemented via R, with reaction times (RTs) and error rates (ERs) as respective dependent variables. Word condition and Chinese proficiency (native, higher L2, and lower L2) were included as fixed variables if they significantly improved the model fit (as shown in Table 5). Model 1 incorporated participants and items as random intercepts due to a better model fit. Again, only significant or marginal effects are reported below. The complete statistical analysis results of Experiment 2 are also available at: osf.io/gqh85.

Table 5.

Gaussian Location-Scale Model and Generalized Linear Mixed-effects Model of Participant Responses to Experiment 2.

Model	Dependent variable	Model specification
3	Response time (RT)	gam(list(Resp_time ~ group + wordtype + s(subject, bs="re") +s(item, bs="re"), ~ group + wordtype),family="gaulss", data=dat4)
4	Error rate (ER)	glmer(error ~ wordtype+Learner_Group +tcw_ar+ log(Ini_pho_fre) + (1 + wordtype | subject) + (1 + Learner_Group | item), data =dat,family=binomial)

Note: Model 3 utilized the *gam()* function from *mgcv* package, while Model 4 employed the *glmer* function from *lme4* package with the specification of "family=binomial." group, native/higher L2/ lower L2 group; tcw.ar, accuracy rate of two-character word type; initial_pho_fre: initial phoneme frequency; subject, participants; item, 64 items.

Results

As documented in Table 6, native speakers processed the three types of three-character items differently in the standard word naming task. The mean response times were shortest for three-character legally formed words, followed by three-character illegally formed words, and longest for unrelated three-character strings; a similar pattern was found for higher L2 proficiency learners; in contrast, it took a longer time for lower L2 proficiency learners to process three-character illegally formed words. As to accuracy rates, for all three groups, the accuracy rates of three-character legally formed words were higher than the other two sets of three-character items.

Table 6.

Descriptive Statistics of the Standard Word Naming Task.

Chinese proficiency	Word condition	Response times (in ms)			Accuracy rate		
		Mean	SD	95% CI	Mean	SD	95% CI
Native	TCW	654.02	153.25	633.66, 674.39	0.86	0.35	0.81, 0.90
	LFW	724.11	178.32	700.31, 747.92	0.86	0.34	0.82, 0.91
	IFW	742.07	147.29	722.46, 761.69	0.85	0.36	0.81, 0.90
	TCS	847.90	207.09	819.10, 876.70	0.78	0.41	0.73, 0.83
Higher L2	TCW	845.08	274.39	813.15, 877.02	0.93	0.25	0.91, 0.96
	LFW	902.87	252.96	871.30, 934.45	0.82	0.38	0.78, 0.86
	IFW	934.84	271.17	901.47, 968.22	0.84	0.37	0.80, 0.88
	TCS	952.79	290.65	910.40, 995.18	0.59	0.49	0.53, 0.64
Lower L2	TCW	877.02	277.23	839.57, 914.46	0.75	0.43	0.70, 0.80
	LFW	912.71	308.69	868.88, 956.54	0.66	0.48	0.60, 0.71
	IFW	956.23	282.89	916.27, 996.18	0.65	0.48	0.60, 0.71
	TCS	916.10	347.22	849.23, 982.97	0.36	0.48	0.31, 0.42

Note. TCW, two-character real word; LFW, three-character legally formed word; IFW, three-character illegally formed word; TCS, three-character unrelated character strings.

Table 7.

Results of the Effects between Word Condition and Chinese Proficiency on the Standard Word Naming Response Times.

Model 3	Estimate	SE	*z* value	*p* value
Intercept	910.923	32.972	27.627	< 0.001
Native vs. Higher	−172.534	44.628	−3.866	.0001
Lower vs. Higher	12.348	43.933	0.281	.7787
IFW vs. LFW	24.949	19.967	1.249	.2125
TCS vs. LFW	87.923	21.231	4.141	< 0.0001
Intercept.variance	5.349	0.037	144.518	< 0.0001
Native.variance[a]	−0.322	0.041	−7.932	< 0.0001
Lower.variance[a]	0.226	0.043	−7.932	< 0.0001
IFW.variance[b]	−0.059	0.399	−1.479	.1392
TCS.variance[b]	0.164	0.0440	3.734	.0002

Notes. LFW, three-character legally formed word; IFW, three-character illegally formed word; TCS, three-character unrelated character strings. Est., estimate. [a], the higher L2 group was treated as the reference group; [b], three-character legally formed word condition was treated as the reference.

First of all, there was no significant interaction effect between word type and Chinese proficiency, thus not reported in Table 7. Native speakers' reaction times (RTs) were significantly shorter than those of higher L2 learners, and there was no significant difference in RTs between higher and lower L2 learners. For all three groups, the RTs for three-character strings were significantly longer than for legally formed words, but there was no significant difference in RTs between illegally formed words and legally formed words. Similar patterns were observed for variance analyses.

Table 8 reports the results of error rates. No significant interaction effect between word type and Chinese proficiency was found. In Model 4, the error rates for three-character strings were significantly higher than for legally formed words (estimate = 1.283, *p* = .003), and there was no significant difference between illegally formed and legally formed words (estimate = 0.169, *p* = .6213). Overall, native speakers' error rates were significantly lower than those of higher L2 learners (estimate = −1.194, *p* = .0027), and the error rates were not statistically different between higher and lower L2 learners (estimate = 0.499, *p* = .0895).

Table 8.

Results of the Effects between Word Condition and Chinese Proficiency on the Standard Word Naming Error Rates.

Model 4	Estimate	SE	*z* value	*p* value
Intercept	2.647	2.364	1.119	.2630
IFW vs. LFW[a]	0.169	0.341	0.494	.6213
TCS vs. LFW	1.283	0.359	3.578	.0003
Native vs. Higher L2[b]	−1.194	0.398	-3.001	.0027
Lower L2 vs. Higher L2	0.499	0.294	1.698	.0895
TCW accuracy rate	−3.962	0.839	−4.721	< 0.0001
Log(initialphonemefrequency)	−0.086	0.199	−0.433	.6647

Notes. TCW, one-morpheme two-character real word; LFW, three-character legally formed word; IFW, three-character illegally formed word; TCS, three-character unrelated character strings. Est., estimate. [a], three-character legally formed word condition was treated as the reference; [b], the higher L2 group was treated as the reference group.

In sum, Experiment 2 refuted our hypothesis that L2 learners processed all three types of three-character items similarly. Despite responding more slowly and less accurately than native speakers, both higher and lower L2 proficiency learners demonstrated morphological sensitivity since their response times for three-character strings were significantly longer than for legally formed words, and their error rates were higher for three-character strings compared to legally formed words.

DISCUSSION

This study examined L2 morphological sensitivity to intraword morphological structure in a morphosyllabic language (i.e., Chinese). Participants of different Chinese proficiency backgrounds (i.e., native, higher L2 and lower L2) completed two experiments. Experiment 1 adopted a segment shifting task; Experiment 2 used a standard word naming task. In the segment shifting task, native speakers responded more quickly, and made fewer errors, than higher and lower L2 learners. Affixoids were responded to more quickly in the mean than nonaffixoids. Notably, base frequency was significantly facilitatory. In addition, the variance in response times and error rates were lower for native speakers for the affixoid condition, and the participants' response efficiency decreased with increasing base frequency. In the standard word naming task, the results for different Chinese proficiency groups mirrored those of the segment shifting task. Native speakers also responded more quickly than higher and lower L2 learners. However, for all three groups, the three-character-string/TCS condition elicited more errors and longer response times than the legally formed word/LFW condition. These results indicate that native and L2 Chinese readers alike are sensitive to the morphological structure of the Mandarin words, as witnessed by shared affixoid and TCS-vs.-LFW effects.

Interestingly, native speakers responded more quickly and showed less variance. These are considered the hallmarks of more learning experience. Importantly, the evidence from this research does not support the hypothesis that L2 learners are morphologically insensitive and do not adopt a decompositional strategy in processing morphologically complex words like native speakers, and that L2 processing is significantly slower than that of native speakers (Maeng & Kim, 2023; Neubauer & Clahsen 2009; Silva and Clahsen, 2008). If this were the case, we would have found similar response times and error rates across word conditions, regardless of Chinese proficiency and task demands.

Also, our findings regarding the response differences between native and L2 Chinese readers were consistent with those in Wade-Woolley and Geva (1999)’s study of L1 Russian-L2 English readers, which found that L1 English learners were more efficient word readers and were sensitive to morphologically complexity in response times, whereas less efficient L1 Russian readers of L2 English only demonstrated L2 morphological sensitivity in accuracy rates. A potentially unique finding of our study is that we did not observe the graded morphological effect identified in previous research that administered a lexical decision task to learners with increasing L2 proficiency (e.g., Coughlin & Tremblay, 2015). Both higher and lower L2 proficiency learners in our study demonstrated some morphological sensitivity in both experiments. This suggests that even learners with lower L2 Chinese proficiency were sensitive to the intraword morphological structure of Chinese.

There are several possible explanations for our findings. First, the L1 background and L2 proficiency of the adult L2 learners might interact and affect L2 processing strategies. For example, Zhang (2017) found that in a study of word decoding accuracy in Singapore, L1 English-speaking children of L2 Chinese did not reply on morphological decomposition as much as native Chinese-speaking children do. Although the adult L2 learners in the present study were categorized into higher and lower Chinese proficiency based on a print vocabulary test, perhaps a superior-level L2 learner group or more refined measurement of L2 proficiency (e.g., a combination of standardized, researcher-designed and learner-reported indexing) is needed to make further progress. Second, the word naming paradigm adopted in the present study might be cognitively and linguistically more demanding than the lexical decision approach or lexical inferencing used in the majority of previous L2 Chinese research (e.g., Chen, 2018; Koda & Miller, 2018). Making a lexical decision often involves YES/NO judgement, while word naming requires L2 learners to produce sounds in L2 within limited time lapses across stimuli. Finally, the finding that neither native speakers nor L2 learners' word naming was slowed down by affixoid positional constraint violation surprised us. Recent research has suggested that character transposition affects native Chinese speakers' word meaning retrieval (e.g., Zhang et al., 2021). Perhaps because the present study involved word decoding instead of meaning access, affixoid positional constraint violation did not affect participants' responses. However, we did identify a morphological effect in the segment shifting task. More future research is needed to determine the potential interactional effects of L2 proficiency and experimental tasks on L2 morphological processing.

The present study has several limitations that invite further research. First, only one researcher-designed L2 vocabulary measure was used as a proxy to L2 proficiency, while previous L2 research has used a standardized test (e.g., Chen, 2018), or measured both vocabulary and grammar knowledge (e.g., Miller & Koda, 2018), or measured both oral and print vocabulary knowledge (e.g., Zhang & Koda, 2018). Second, we did not control word status (i.e., real words vs. pseudowords) in this study. Last, word length was not controlled in the standard word naming task. The reason for this was that it is unrealistic or impossible to find a sufficiently large set of three-character real words in modern Chinese lexicon. Another direction for future research is the application of the Gaussian-Location-Scale models. We used this modeling approach to analyze the response time data in this research, mainly because the mean and variance can be defined independently from each other in the Gaussian distribution. According to Coupé (2018), Gaussian-Location-Scale models, which are still very rarely used in language science research, can be employed to investigate complex linguistic variables (see also Baayen et al., 2022).

Note.

1. According to Booji (2005), affixoid is a third type of morpheme in between lexemes and affixes. An increasing amount of modern Chinese lexicon are formed by affixoids (Zeng, 2008). For instance, the morpheme 学*xué* in Chinese can be used as an independent verb, meaning ‘to study;’ or as a suffixoid, a bound and productive form that has a limited meaning (i.e., ‘branch of study’, like ‘-logy’ in English) and a fixed position (i.e., the end of a multicharacter and multisyllabic word) (e.g., 文学 ‘literature’ and 人类学 ‘anthropology’) (Arcodia, 2011). By the same logic, a prefixoid has a fixed position at the beginning of a multicharacter and multisyllabic word (e.g., 反*fǎn* in 反作用, like ‘counter-’ in ‘counteraction’). More recently, Tseng et al. (2020) proposed three quantitative features in a computational model of affixoid behavior in Mandarin Chinese: (1) morphological productivity, (2) syntactical productivity, and (3) semantic diversity (see also Tian & Baayen’s 2023 exploration of Chinese word formation in terms of productivity and semantic transparency).

REFERENCES

Arcodia, G. F. (2011). A Construction Morphology account of derivation in Mandarin Chinese. *Morphology*, *21*(1), 89–130

Babcock, L. Stowe, J.C., Maloof, C. J., Brovetto, C., & Ullman, M. T. (2012). The storage and composition of inflected forms in adultlearned second language: A study of the influence of length of residence, age of arrival, sex, and other factors. *Bilingualism: Language and Cognition*, *15*, 820840. https://doi.org/10.1017/S1366728912000053

Baayen, R. H., Fasiolo, M., Wood, S., & Chuang, Y. Y. (2022). A note on the modeling of the effects of experimental time in psycholinguistic experiments. *The Mental Lexicon*, *17*(2), 178-212. https://doi.org/10.1075/ml.21012.baa

Booij, G. E. (2005). *The grammar of words: an introduction to linguistic morphology*. Oxford: Oxford University Press.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One*, *5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Chen, T. (2018). The contribution of morphological awareness to lexical inferencing in L2 Chinese: Comparing more-skilled and less-skilled learners. *Foreign Language Annals*, *51*(4), 816-830. https://doi.org/10.1111/flan.12365

Chinese Proficiency Test Center. (2011). *Grading syllabus for Chinese vocabulary and Chinese characters (in Chinese)*. Beijing: Economy and Science Press.

Crepaldi, D., Che, W. C., Su, I. F., & Luzzatti, C. (2012). Lexical-semantic variables affecting picture and word naming in Chinese: A mixed logit model study in aphasia. *Behavioural Neurology*, *25*(3), 165–184. https://doi.org/10.3233/BEN-2012-119002

Coughlin, C. E., & Tremblay, A. (2015). Morphological decomposition in native and non-native

French speakers. *Bilingualism: Language and Cognition*, *18*(3), 524–542. https://doi.org/10.1017/S1366728914000200

Coupé, C. (2018). Modeling linguistic variables with regression models: Addressing non-Gaussian distributions, non-independent observations, and non-linear predictors with random effects and generalized additive models for location, scale, and shape. *Frontiers in Psychology*, *9*, 307831. https://doi.org/10.3389/fpsyg.2018.00513

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. University of Hawaii Press.

Deng, T., Shi, J., Dunlap, S., Bi, H., & Chen, B. (2016). Morphological knowledge affects processing of L2 derivational morphology: An event-related potential study. *Journal of Neurolinguistics*, *37*, 47–57. https://doi.org/10.1016/j.jneuroling.2015.09.001

Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *64*(4), 344–358. https://doi.org/10.1016/j.jml.2011.01.003

Embick, D., Creemers, A., & Goodwin Davies, A. (2021). Morphology and the mental lexicon: Three questions about decomposition. In A. Papafragou, J. C. Trueswell, & L. R. Gleitman (Eds.), *The Oxford Handbook of the Mental Lexicon*. Oxford: Oxford University Press.

Feldman, L. B., Frost, R., & Pnini, T. (1995). Decomposing words into their constituent morphemes: Evidence from English and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 947–960. https://doi.org/10.1037/0278-7393.21.4.947

Feldman, L. B., Kostić, A., Basnight-Brown, D. M., Đurđević, D. F., & Pastizzo, M. J. (2010). Morphological facilitation for regular and irregular verb formations in native and non-native speakers: Little evidence for two distinct mechanisms. *Bilingualism: Language and Cognition*, *13*(02):119–135. https://doi.org/10.1017/S1366728909990459

Feldman, L. B., & Prostko, B. (2002). Graded aspects of morphological processing: Task and processing time. *Brain and Language*, *81*(1-3), 12–27.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, *35*(5), 263–279. https://doi.org/10.1017/S0140525X11001841

Gor, K. (2017). The mental lexicon of L2 learners of Russian: Phonology and morphology in lexical storage and access. *Journal of Slavic Linguistics, 25*(2), 277–302. https://doi.org/10.1353/jsl.2017.0011

IELTS. (2018).The IELTS scale. Retrieved from: https://www.ielts.org/en-us/about-the-test/how-ielts-is-scored

Ke, S. & Koda, K. (2017). Contributions of morphological awareness to adult L2 Chinese word meaning inferencing. *The Modern Language Journal*, *101*(4), 742–755. https://doi.org/10.1111/modl.12428

Ke, S., & Koda, K. (2019). Is vocabulary knowledge sufficient for word-meaning inference? An investigation of the role of morphological awareness in adult L2 learners of Chinese. *Applied Linguistics*, *40*(3), 456–477.

Ke, S. & Koda, K. (2021). Transfer facilitation effects of morphological awareness on multicharacter word reading in Chinese as a second language. *Applied Psycholinguistics*, *42*(5), 1263–1286.

Larsen-Freeman, D. (2010). Not so fast: A discussion of L2 morpheme processing and acquisition. *Language Learning*, *60*(1), 221–230.

Lexicon of Common Words in Contemporary Chinese (2008). *Lexicon of Common Words in Contemporary Chinese*. Beijing: Commercial Presss.

Liang, L., & Chen, B. (2014). Processing morphologically complex words in second-language learners: The effect of proficiency. *Acta Psychologica*, *150*, 69–79. https://doi.org/10.1016/j.actpsy.2014.04.009

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and Competence in Second Language Acquisition* (pp. 35–53). Cambridge: Cambridge University Press.

Maeng, J., & Kim, S-A. (2023). L2 processing of morphologically complex words in Chinese. *Linguistic Research*, *40*, 1-31. https://doi.org/10.17250/khisli.40..202309.001

Neubauer, K., & Clahsen, H. (2009). Decomposition of inflected words in a second language: An experimental study of German participles. *Studies in Second Language Acquisition*, *31*(3), 403–435.https://doi.org/10.1017/S0272263109090354

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110–114.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an" outlier" orthography. *Psychological Bulletin*, *13*4(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Share, D. L. (2021). Is the Science of Reading Just the Science of Reading English?. *Reading Research Quarterly*, *56*, S391-S402.

Shen, H., & Ke, C. (2007). Radical awareness and word acquisition among nonnative learners of Chinese. *The Modern Language Journal*, *91*, 97–111. https://doi.org/10.1111/j.1540-4781.2007.00511.x

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, *11*(2), 245–260. https://doi.org/10.1017/S1366728908003404

Su, Y-F., & Samuels, S. J. (2010). Developmental changes in character-complexity and word-length effects when reading Chinese script. *Reading and Writing*, *23*, 1085–1108. https://doi.org/10.1007/s11145-009-9197-3

Tseng, Y. H., Hsieh, S. K., & Chen, P. Y. (2020, December). Computational modeling of affixoid behavior in Chinese morphology. In D. Scott, N. Bel & C. Zong (eds.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 2879-2888). Barcelona: International Committee on Computational Linguistics.

Uchihara, T., & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *The Language Learning Journal*, *47*(1), 64–75. https://doi.org/10.1080/09571736.2016.1191527

Vainio, S., Pajunen, A., & Hyönä, J. (2014). L1 and L2 word recognition in Finnish: Examining L1 effects on L2 processing of morphological complexity and morphophonological transparency. *Studies in Second Language Acquisition*, *36*(1), 133–162. https://doi.org/10.1017/S0272263113000478

Wade-Woolley, L., Geva, E. (1999). Processing inflected morphology in second language word recognition: Russian-speakers and English-speakers read Hebrew. *Reading and Writing 11*, 321–343. https://doi.org/10.1023/A:1008097300762

Wang, J., Huang, L., & Guo, S. (2016). The overall design and quality analysis of the unified comprehensive Chinese examination of the preparatory education for international students in China [in Chinese]. *Language Teaching and Linguistic Studies*, *38*, 53–58.

Wang, M., Liu, Y., & Perfetti, C. A. (2004). The implicit and explicit learning of orthographic structure and function of a new writing system. *Scientific Studies of Reading*, *8*, 357–379. https://doi.org/10.1207/s1532799xssr0804_3

Xu, Y., Chang, L.–Y., & Perfetti, C. A. (2014). The effect of radical-based grouping in character learning in Chinese as a foreign language. *The Modern Language Journal*, *98*, 773–793. https://doi.org/10.1111/modl.12122

Yang, J. (2000). Orthographic effect on word recognition by learners of Chinese as a Foreign Language. *Journal of Chinese Language Teachers Association*, *35*, 1–18.

Zeng, L. (2008). Quantitative and qualitative study of Chinese semi-affixes (in Chinese). *Chinese Teaching in the World*, *4*, 75–87.

Zhang, D. (2017). Word reading in L1 and L2 learners of Chinese: Similarities and differences in the functioning of component processes. *The Modern Language Journal*, *101*(2), 391–411. https://doi.org/10.1111/modl.12392

Zhang, J., Lin, T., Wei, J., Anderson, R. (2014). Morphological awareness and learning to read Chinese and English. In X. Chen, Q. Wang, & Y. C. Luo (Eds.), *Reading development* *and difficulties of monolingual and bilingual Chinese children* (Vol. 8, pp.3-22). Springer Science: Dordrecht, Netherlands.

Zhang, H., & Koda, K. (2018). Word-knowledge development in Chinese as a heritage language learners: A comparative study. *Studies in Second Language Acquisition*, *40*(1), 201–223. https://doi.org/10.1017/S0272263116000450

Zhang, E. H., Lai, X. X., Li, D., Lei, V. L. C., Chen, Y., & Cao, H. W. (2021). Electrophysiological correlates of character transposition in two-character Chinese word identification. *Journal of Neurolinguistics*, *60*, 101024. https://doi.org/10.1016/j.jneuroling.2021.101024

Appendix A. Description and Results of L2 Vocabulary Knowledge Task

Following Meara (1996) and Uchihara and Saito (2019), we treated vocabulary knowledge as a proxy for general L2 proficiency in this research. In the paper-and-pencil vocabulary knowledge task, the participants were asked to provide English translations for 60 single- and two-character Chinese words. Each response with correct translation was awarded with one point (total score possible = 60). The respective mean accuracy rates for higher and lower proficiency learner groups were 44.65 (74.42%) (median = 44.00, SD = 4.97) and 30.42 (50.70%) (median =32.00, SD = 5.18). There was a significant difference in the vocabulary knowledge test performance between the higher L2 proficiency and lower L2 proficiency learners ($t = 8.80$, $p < .001$).

Appendix B. Segment Shifting Task Items

Segment	Affixiod condition	Nonaffixoid condition	Target word
大	大部分	大使馆	世界
小	小汽车	小吃店	学校
总	总收入	总结会	人口
老	老朋友	老虎机	同学
多	多功能	多少钱	语言
反	反作用	反对票	传统
家	音乐家	回老家	文学
业	银行业	非专业	旅游
方	合作方	打比方	挑战
面	对立面	一方面	知识
气	孩子气	冷空气	男子
生	服务生	好先生	研究
期	实习期	下星期	生长
式	美国式	非正式	自由
度	知名度	加速度	对比
力	生产力	强有力	影响

Appendix C. Item Selection Procedures

Three steps were taken to construct the item pool. In the first step, 16 productive bound morphemes were adapted from Zeng's (2008) database, which, to the best of our knowledge, is the only available resource that provides a list of productive word formation morphemes in Chinese, accompanied by morpheme frequency. This database is critical because it provides an explicit description of four inclusion criteria: (a) productivity, (b) position stability, (c) desemantization (with weakened lexical meaning), and (d) boundness (cannot be used as an independent lexical unit). The database includes 34 prefixoids (productive morphemes with fixed positions at the beginning of multicharacter words) and 54 suffixoids (productive morphemes with fixed positions at the end of multicharacter words). Of the 16 morphemes selected for the proposed study, six were prefixoids and ten were suffixoids (the same set of affixoids was also used to form critical items in the standard word naming task described in Experiments 2. Second, based on the intended participants' language learning backgrounds, Chinese base words (morphologically simple two-character words) were selected from Bands One and Two – the two lowest levels of the Grading Syllabus for Chinese Vocabulary and Chinese Characters (GSCVCC, Chinese Proficiency Test Center, 2001). Third, all characters' visual complexities were controlled within the moderate range (mean number of strokes around 13) (see Su & Samuels, 2010).

Appendix D. Standard Word Naming Task Items

TCW	TCS	LFW	IFW
火车	跟找已	大名字	文化大
机会	美西哪	小工作	辛苦小
世界	前才最	多颜色	水平多
经济	直次信	反关系	考试反
公司	爱常爸	总决定	问题总
电影	孩放而	老活动	事情老
水果	之呢候	高兴度	度成功
生活	错机正	手表家	家点心
安全	中今它	学习力	力改变
可以	克感明	未来式	式自己
非常	杀帮先	比赛生	生准备
完全	头进应	衣服业	业电话
告诉	法实谁	特别面	面紧张
希望	年比从	重要期	期知道
教育	晚被快	教授气	气农村
认识	死像等	开始方	方喜欢

Notes. The affixoids are underlined. TCW, two-character real word; LFW, three-character legally formed word; IFW, three-character illegally formed word; TCS, three-character unrelated character strings.

Address for correspondence

Sihui (Echo) Ke

Department of Chinese and Bilingual Studies

Faculty of Humanities

EF406c, The Hong Kong Polytechnic University

Hung Hom, Kowloon, Hong Kong SAR, China

sihui.ke@polyu.edu.hk or echoecho.ke@gmail.com

Co-author information

Rui Jin, Faculty of Education, Shenzhen University, Rji227@outlook.com

Keiko Koda, Department of Languages, Cultures & Applied Linguistics, Carnegie Mellon University, kkoda@andrew.cmu.edu

Initial submission date: 14 April, 2022

Acceptance date: 1 May, 2024