The effect of orthographic form-cuing on the phonological preparation unit in spoken word production

Chuchu Li · Min Wang · William Idsardi

Abstract Two experiments using the form-preparation paradigm were conducted to investigate the effect of orthographic form-cuing on the phonological preparation unit during spoken word production with native Mandarin speakers. In both experiments, participants were instructed to memorize nine prompt-response monosyllabic word pairs, after which an associative naming session was conducted in which the prompts were presented and participants were asked to say the corresponding response names as quickly and accurately as possible. In both experiments, the response words in the homogeneous lists shared the same onsets, or shared the same rimes; the response names had no common aspects of pronunciation in the heterogeneous lists. Chinese characters (Experiment 1) and Pinyin (phonetic transcription of the characters) (Experiment 2) were used to investigate the effect of the orthographic form. Significant onset facilitation and rime inhibition was shown for Pinyin syllables but not for characters. The contrasts of the onset and rime effect in the two orthographic forms suggest that a specific phonological unit is promoted in spoken word production in a certain orthographic form. Pinyin cued the participants to prepare the onset whereas Chinese characters did not. The rime interference effect may arise as a result of lexical competition in spoken word production.

Keywords Orthography · Phonological preparation unit · Spoken word production

C. Li · M. Wang (✉)
Department of Human Development and Quantitative Methodology, University of Maryland, 3304C Benjamin Building, College Park 20742, MD, USA
e-mail: minwang@umd.edu

W. Idsardi
Department of Linguistics, University of Maryland, College Park, MD, USA

Introduction

Spoken word production involves the operation of a series of cognitive mechanisms. A general top-down architecture of production starts from message or concept encoding, to lemma selection, lexeme retrieval, phonemic segment retrieval, syllable construction, and, finally, to articulation (Ferreira, 2010). Phonological retrieval and encoding is an indispensable process in language production. The WEAVER++ model (Levelt, Roelofs, & Meyer, 1999) suggests that at the beginning of phonological encoding in production, metrical and segmental units (e.g., stress and phonemes) are accessed in a parallel fashion. Later on, the phonemic segments are linearized in a syllabified organization that guides articulation. However, is the process of phonological retrieval and encoding in spoken word production the same across different languages? If not, what factors are responsible for the differences? In particular, does the orthographic form that represents the language matter? The present study investigated whether the use of different orthographic forms for the same language has an impact on phonological retrieval and encoding in spoken word production. We use the term "preparation unit" to refer to the phonological unit that is retrieved from the lexicon at the beginning of phonological encoding.

The form-preparation paradigm

The form-preparation task, also known as the implicit priming paradigm, has been frequently used to investigate the nature of the preparation unit in spoken word production (e.g., Chen, Chen, & Dell, 2002; Cholin, Schiller, & Levelt, 2004; Kureta, et al., 2006; Meyer, 1990, 1991; O'Seaghdha, Chen, & Chen, 2010). The task involves an

associative-learning session and a naming session. In the associative-learning session, participants memorize some prompt-*response* word pairs (e.g., night-*day*, wet-*dew*, and bread-*dough*). After participants have informed the experimenters that they have memorized all of the pairs, an associative naming session is immediately conducted in which prompt words appear unpredictably and the participants are required to say the response word as quickly and accurately as possible, while their response time is recorded (e.g., when the word *night* is presented, participants need to say the word *day*). The rationale of this paradigm is that, compared with the *heterogeneous* (or *control*) context in which response words do not share any elements (e.g., three response words are *day*, *sea*, *pie*), in a *homogeneous* context where the response words share the same initial element (e.g., the initial phoneme is always /d/ for *day*, *dew* and *dough*), the fore-knowledge of the initial element allows the participants to prepare their first phonological unit in production, thus facilitating their naming latency. The smallest ingredient that can lead to such a form-preparation effect is referred to as the preparation unit.

Meyer (1990, 1991) studied the preparation unit in Dutch using the form-preparation paradigm. She found that the preparation unit did not differ within a language when words with different lengths were produced. Native Dutch speakers benefited from the fore-knowledge of the onset of a set of words regardless of whether the words were short (e.g., monosyllabic words) or long (e.g., disyllabic words). Furthermore, the fore-knowledge of later shared components of a set of words did not elicit any significant effects, since manipulating the similarity of the rime, coda, or the second syllable of a word did not lead to a form-preparation effect. A possible explanation is that the assembly of the phonological units is sequential. Participants always need to prepare the utterance of the onset of response words first, and then proceed to the rime. Given that measuring the response time of spoken word production is about the time participants take to produce the first sound of a word, fore-knowledge of the later components does not lead to a faster response time. Roelofs (1999) further showed that the benefit from the fore-knowledge of shared onset in the form-preparation paradigm is driven by shared segmental information but not phonetic features. Neither the place nor the manner of articulation of the initial phoneme affected the form-preparation effect in Dutch (e.g., although /p/ and /b/ are both bilabial and stop phonemes, they did not yield a preparation effect when speakers produced names such as *bajes*, *bami*, *paling*); and only sharing the exact same initial phoneme provided a benefit. Finally, fore-knowledge of simply metrical properties such as the number of syllables, primary stress location, or tonal information did not benefit the preparation of spoken word production, although variability in these properties may reduce the benefits from the advance

knowledge of the initial segment (Chen, et al., 2002; Roelofs & Meyer, 1998).

The Preparation Unit in Different Languages and the Influence of Orthography

Research using the form-preparation task suggests that the preparation unit differs across languages (Chen, et al., 2002; Kureta, et al., 2006; Meyer, 1990, 1991; O'Seaghdha, et al., 2010). Native speakers of languages which are typically written using an alphabet such as Dutch (Meyer, 1990, 1991) and English (Jacobs & Dell, 2014; O'Seaghdha et al., 2010) have been shown to benefit from fore-knowledge of the initial phoneme, suggesting that the primary preparation unit in spoken word production in alphabetic languages is the phonemic segment. In O'Seaghdha et al. (2010, Experiments 5 and 6), native English speakers were asked to memorize a set of word pairs or picture-word pairs. Regardless of the format of the prompt (i.e., visual words or pictures), participants showed significant facilitation when a set of response words shared the same onset phoneme. Most recently, Jacob and Dell (2014) showed that the benefit from the fore-knowledge of the initial phoneme of a word may also be applied to compound words when two free morphemes are combined to form a morphologically complex word. The onset phoneme of the second morpheme of the compound (e.g., /d/ in *sawdust*) did not facilitate the word production. Chen et al. (2002) and O'Seaghdha et al. (2010), on the other hand, showed that native Mandarin speakers benefited from the same initial syllable but not from the same initial phoneme, suggesting that in Mandarin the preparation unit is the syllable rather than the phoneme. O'Seaghdha et al. asked native Mandarin speakers to memorize a set of monosyllabic word pairs when the prompts and responses could form a compound word. For example, in their Experiment 2, 雀 (que4) was the prompt and 麻 (ma2) was the response, and 麻雀 means *sparrow* in Chinese. In their Experiment 3, the prompts and responses were either semantically related words (e.g., 碰 (peng4, *touch*) served as the prompt and 摸 (mo1, *stroke*) as the response) or unrelated symbol-word pairs (e.g., & served as the prompt and 摸 (mo1, *stroke*) as response). In all situations, participants failed to show faster response times when the response words shared the same onset phonemes. Participants were also asked to memorize disyllabic word pairs (Experiments 1 and 7), and still failed to show onset facilitation; they only showed significantly faster response times when the response words shared the same initial syllable segment. In Kureta et al. (2006), native Japanese speakers showed facilitation from response words with the same initial CV (consonant-vowel) mora but not the same phonemic segment, suggesting that CV mora may be the primary preparation unit in Japanese.

Interestingly, the preparation units in those three different languages (i.e., English, Mandarin, and Japanese) are also consistent with how the writing systems most commonly used in these languages are designed. For languages using alphabetic orthographies such as Dutch and English, the phonemic segment is adopted in planning spoken words in these languages. For Mandarin whose primary orthography is morphosyllabic in which each character represents a syllable and a morpheme, speakers plan spoken words in syllables. Japanese speakers tend to plan their speech mora by mora, which is consistent with the language's moraic orthography—each Japanese Kana letter represents a mora. Here a natural question arises: in a task in which orthographic information is explicitly represented, is it possible that a certain orthographic form would promote the selection of a particular preparation unit? Orthographic knowledge is not required in production, but it is possible that orthographic information affects spoken word production in literate speakers. Damian and Bowers (2003) conducted a series of experiments to investigate this issue using the form-preparation paradigm. Using stimuli that were presented in either visual or spoken form, they found that native English speakers did not benefit from overlapping of the initial phoneme if the phoneme was spelled in different ways in a set of response words (e.g., *kennel*, *coffee*, and *cushion*). However, these results were not replicated in Roelofs (2006). In an associative learning-naming task, native Dutch speakers showed a significant facilitative effect when the response words shared the same initial phoneme but not initial letters (e.g., *sandaal*, *circuit*, and *CD*). The inconsistency in the findings of these two studies was interpreted as a consequence of the difference in orthographic depth between Dutch and English. Compared to Dutch, sound and spelling correspondence is more complicated in English, and so orthography and phonology may interact more in speech production among native English speakers.

Another factor that was thought to contribute to the orthographic effect in Damian and Bowers (2003) is that speakers may use orthographic codes to facilitate completion of a complicated spoken word production task, such as the word associative learning and production task in the form-preparation paradigm (Alario, Perre, Castel, & Ziegler, 2007). Alario et al. (2007) found that, in a simple picture-naming task where orthographic information was not presented, speakers did not show the orthographic effect that was found in Damian and Blowers (2003). A number of studies suggest that skilled readers retrieve and manipulate phonological units of visual words when they comprehend or memorize the printed materials (Crowder, 1982; Mann, 1986; Perfetti & McCutchen, 1982, Stanovich, 1982). In an associative naming task with the form-preparation paradigm, it is possible that exposure to the visual words in the associative-learning session allows participants to encode the word pairs based on the orthographic codes of the language, such that the orthographic information (i.e., spelling) of the visual stimuli has an effect on spoken word production in the subsequent associative naming session. However, this influence may be dependent upon other factors such as the orthographic depth of the target language. The present study investigated the influence of orthography from a different perspective, that is, we examined whether different orthographic forms (e.g., alphabetic vs. morphosyllabic) cue different preparation units in spoken word production.

The preparation unit in Chinese: Evidence from different paradigms Previous studies using the form-preparation paradigm suggested that the preparation unit in Mandarin Chinese is the syllable segment (or atonal syllable, i.e., the syllable without including tonal information) (Chen, et al., 2002; O'Seaghdha, et al., 2010). However, studies using other methodologies have not yielded the exact same conclusion. Using a simple picture-naming task without explicit orthographic information or a learning session, Qu, Damian, and Kazanina (2012) provided ERP evidence (i.e., event-related potentials) that the phonemic segment is the fundamental unit of phonological encoding in spoken word production among native Mandarin speakers. ERPs are indicators of the electrical activities in the brain that are in response to specific events or stimuli (Blackwood & Muir, 1990). Native Mandarin-speaking adults were instructed to name colored line drawings of objects using color adjective-noun phrases. The color and object name either shared the same initial phoneme (e.g., *huang2-he2zi* "yellow box," the number denotes the tone) or were phonologically unrelated (e.g., *lü4-he2zi* "green box"). Compared with the phonologically unrelated condition, when the color and object name shared the initial phoneme, participants showed more positive ERPs in the posterior regions 200

300 ms and more negative ERPs in the anterior regions 300–400 ms after the picture appeared in the phonologically related condition. Previous research suggests that the phonological encoding stage is estimated to take place at a 275–400-ms time window after a picture had appeared (see Indefrey & Levelt, 2004). Qu et al. (2012) proposed that the posterior ERP amplitude in the 200–300-ms time window might be a result of facilitation due to phoneme repetition during phonological encoding, and that the anterior ERP effect in the 300–400-ms time window was a result of internal speech monitoring that aims to avoid speech errors. Participants did not show significantly faster naming latency in the behavioral data in the phonologically related condition. This was explained as resulting from cancellation of the facilitative effect due to phoneme repetition by the negative effect in the internal speech-monitoring stage. Repeated phonemes may have made it easier for participants to exchange the adjacent speech sounds so that the internal speech monitoring could be under higher load in the phonologically related condition. In alphabetic languages such as English, phoneme-based facilitation

may be very pervasive and much stronger than the inhibitory effect due to internal speech monitoring, such that speakers may still show facilitation as an overall effect. In summary, Qu et al. (2012) suggested that the phonemic segment may play a fundamental role in phonological retrieval and encoding when preparing spoken word production in Mandarin, although this does not imply that the phoneme plays exactly the same role in Mandarin as in English.

Previous research on different languages has not reached a consensus about the exact preparation unit during speech planning in various languages and whether orthography might play a role. Different tasks and paradigms may encourage participants to involve orthographic information to different degrees. In the present study, we investigated whether orthography might serve as a cue in spoken-word production by native Mandarin speakers by manipulating the orthographic form in the form-preparation task. Specifically, we studied whether the phonological preparation unit in spoken word production is influenced by the explicit representation of phonological information in the orthography.

Characteristics of the Chinese language and orthography

Unlike languages such as English or Dutch which are represented by only one writing system, there are two different writing systems – characters and Pinyin – used for Mandarin Chinese. As a result, Mandarin Chinese provides an excellent test case for the effect of orthographic form within the same language. The Chinese character is morphosyllabic, meaning that each character corresponds to a morpheme and a syllable. The Chinese character is also logographic, in that the phonological information is not explicitly represented in the orthography. Pinyin uses the Roman alphabet to transcribe the pronunciation of Chinese characters. Pinyin is a transparent system in which phonological information including onsets, vowels, codas, and tones are explicitly represented. In addition, Pinyin has a strict one-to-one letter-sound correspondence.

Pinyin is taught to all children in Mainland China in the first ten weeks of Grade 1 (6–7 years old) before they learn to read Chinese characters, and children use this system to help them learn the pronunciation of characters (Hanley, 2005). For example, the character "早" (early) is represented by Pinyin with the spelling "zǎo." Its onset is "z," the vowel is "ao," and the tone is marked above the vowel "ǎo." Children are instructed to articulate a syllable by pronouncing the onset and vowel or rime separately, and then combine them together (Wang & Gao, 2011). For example, children are taught to pronounce the syllable *mā* by repeatedly saying it as: *m-ā-mā*. After acquiring Pinyin knowledge, children then receive instruction in characters with Pinyin printed above them, such as 早. In Pinyin instruction, when learning rimes with a nasal coda such as "ang," children are not told that these sounds can

be further segmented into a vowel and a final consonant (e.g., "a" and "ng"). Therefore, Pinyin instruction encourages children to segment a CVC syllable into an onset and a rime instead of an onset, a vowel, and a coda. Native Mandarin-speaking children's awareness of the onset phoneme was significantly better at Grade 1 than that at the third level (i.e., the highest level) of kindergarten, largely due to their exposure to Pinyin instruction (Shu, Peng, & McBride-Chang, 2008). As a result, the acquisition of Pinyin knowledge may encourage Mandarin speakers to develop the ability to attend to the phonological information that comprises the onset and rime units. Note that even though Pinyin may not be a fully-fledged writing system and is not being read as commonly as Chinese characters in skilled readers' daily lives, most native Chinese-speakers do use Pinyin as an input system to type characters on a computer.

Mandarin Chinese has a simple syllable structure. There are only two legal codas, /n/ and /ŋ/, and consonant clusters are not allowed. Re-syllabification does not occur in Mandarin Chinese. For example, in English the sound /s/ changes from a coda to the onset of the next syllable when the word *mess* changes to its adjective form *messy*, but this phenomenon does not occur in Mandarin Chinese. According to the WEAVER++ model of spoken word production (Levelt et al., 1999), re-syllabification might be one reason that native English speakers retrieve and encode phonemes instead of syllables at the beginning of phonological encoding—only after all the phonemes are retrieved can the speakers determine the arrangement of syllables in a word. By the same token, the lack of re-syllabification in Mandarin Chinese may allow Mandarin speakers to retrieve the syllable as an integral unit at the beginning of phonological encoding. Another feature of the Chinese language is that consonant clusters (e.g., *sk-*, *gl-*) are illegal, which makes it difficult to differentiate the role of onset versus initial phoneme. However, considering that the focus of the present study is to investigate whether Pinyin and character orthographies provide cues for different preparation units, and the procedure of Pinyin instruction encourages onset-rime division instead of the separation among phonemes, we believe that the preparation unit should be the onset instead of the initial phoneme if it turns out that Pinyin promotes a smaller unit for phonological retrieval and encoding.

The present study

The present study investigated the way native Mandarin speakers retrieve and encode phonological units during spoken-word production, and, in particular, whether a certain orthographic form cues for a particular unit. Two form-preparation tasks were used when the items in each task were

presented in Chinese characters (Experiment 1) and Pinyin (Experiment 2). In both experiments, in order to investigate whether the onset is selected as the preparation unit, an onset session was included in which the critical variable was Context, namely whether a set of response spoken words shared the same onset (e.g., shéng, shǔ, shào, meaning *rope*, *rat*, and *whistle*, respectively) or did not (i.e., control condition, e.g., péng, tǔ, shào, meaning *awning*, *soil*, and *whistle*). In order to investigate the role of rime, a rime session was included in which the critical conditions were the same rime (e.g., shéng, péng, téng, meaning *rope*, *awning*, and *pain*, respectively) as the control condition (e.g., shéng, tǔ, pào, meaning *rope*, *soil*, and *bubble*, respectively). We hypothesized that an onset facilitation effect should be shown only for Pinyin stimuli (Experiment 2) but not for character stimuli (Experiment 1). In terms of the role of rime, Meyer (1991) showed that fore-knowledge of later components of a word did not lead to a form-preparation effect in native Dutch speakers, presumably because of the sequential processing of phonological units. If a similar procedure in Dutch applies to Chinese, then the same rime condition should not lead to a form-preparation effect when the materials are presented in Pinyin. For characters, shared rime alone may also fail to elicit any effect, considering that the characters would cue participants to retrieve and encode syllable as an integral unit.

Experiment 1: Implicit priming using characters

We used an associative naming task with the form-preparation paradigm. Participants memorized nine pairs of prompt-response words presented in Chinese characters. The critical variable was Context, that is, whether the response words shared the same onset, rime, or neither (heterogeneous). The same items were used in the same onset and same rime conditions. There were three lists in the same-onset, same-rime, and the heterogeneous conditions, respectively. The heterogeneous condition served as a control condition (see Table 1 for the lists in the onset session and Table 2 for the lists in the rime session). We expected to replicate the results found in previous literature (O'Seaghdha, et al., 2010) where participants failed to show a facilitative effect in the same-onset condition. We also expected to show a null effect in the same-rime condition if the orthographic form of Chinese characters cues speakers to retrieve and encode the syllable as an integral unit.

Participants

Participants were 16 native Mandarin-speaking students with normal or corrected-to-normal vision and without speech impairment from a mid-Atlantic University in the USA. There were eight females and eight males, whose ages ranged from 21 to 28 years (*M*=23.8, *SD*=1.87). All participants were paid for their participation.

Design and materials

In the form-preparation task, participants were asked to memorize nine pairs of monosyllabic characters. For the onset session, the response words consisted of three sets of three monosyllables with the onsets, /p/, /ʂ/, or /t/. For example, in the /p/ set, the three monosyllables were /pʰəŋ2/, /pu3/, and /pa 4/ (the number denotes tone) that only shared the same onset but different rime segment and tone. Each prompt-response pair was composed of semantically related monosyllabic characters (e.g., /ma 1/- /u3/, *cat-mouse*), and no characters were polyphonic, such that each character had only one pronunciation. Furthermore, the prompt and its response character did not share any pronunciation characteristics. By shuffling the combinations of the nine items, three sets of three monosyllabic characters with the rimes /ʂəŋ2/, /u3/, and /a 4/ were formed for the rime session. For example, in the /əŋ2/ set, the three monosyllabic characters were /pʰəŋ2/, /ʂəŋ2/ and /tʰəŋ2/. These items shared the same-rime segment and tone but differed in onset.

For the onset session, each participant was required to complete the prompt-response associative naming task for six lists. The primary manipulation, Context, was whether the response words were homogeneous (sharing the same onset) or heterogeneous (sharing nothing systematically in common; the control condition). Three lists were homogeneous, and the other three were heterogeneous. Each homogeneous list had its own corresponding heterogeneous list, so the six lists made up three homogeneous-heterogeneous pairs. See Table 1 for all stimuli for each list. Each list had three presentation Blocks, and in each block, each item was presented four times (Repetitions) randomly. Therefore, in each list, each prompt was presented 12 times in total. The Sequence of contexts (i.e., whether participants received the homogeneous list or heterogeneous list first from each pair of lists) was counterbalanced across participants. The order of the three list-pairs was counterbalanced among participants as well. Each participant received 216 trials (3 blocks ×2 contexts× 3 onsets ×3 items ×4 repetitions). Sequence was treated as a between-subjects factor. Another counterbalanced variable was the sequence of onset and rime sessions. Half of the participants received the onset session first and the other half received the rime session first.

For the rime session, the design was the same as that for the onset session, except that the items in a homogeneous list shared the same rime rather than the same onset. See Table 2 for all stimuli for each homogeneous list and heterogeneous list. All participants were asked to finish both the onset and rime sessions, and the order to finish these two sessions was counterbalanced among participants. The interval between the

Table 1   The six lists of all items used for the form-preparation task in the onset session in both Experiments

	Homogeneous Condition	Heterogeneous Condition
List 1	钩(gōu, *hook*) —— 绳 (shéng, *rope*) 猫 (māo, *cat*) —— 鼠(shǔ, *rat*) 笛 (dí, *flute*) —— 哨 (shào, *whistle*)	盖 (gài, *cover*) —— 篷 (péng, *awning*) 石(shí, *stone*) —— 土(tǔ, *soil*) 笛 (dí, *flute*) —— 哨 (shào, *whistle*)
List 2	盖 (gài, *cover*) —— 篷 (péng, *awning*) 表(biǎo, *table*) —— 谱(pǔ, *music score*) 球 (qiú, *ball*) —— 泡(pào, *bubble*)	钩(gōu, *hook*) —— 绳 (shéng, *rope*) 表(biǎo, *table*) —— 谱(pǔ, *music score*) 膜 (mó, *membrane*) —— 套(tào, *cover*)
List 3	困(kùn, *sleepy*) —— 疼(téng, *pain*) 石(shí, *stone*) —— 土(tǔ, *soil*) 膜 (mó, *membrane*) —— 套(tào, *cover*)	困(kùn, *sleepy*) —— 疼(téng, *pain*) 猫 (māo, *cat*) —— 鼠(shǔ, *rat*) 球 (qiú, *ball*) —— 泡(pào, *bubble*)

Note. In each cell, the left-hand monosyllabic word is the prompt and the response word is to the right. After each character, its Pinyin transcription followed by its English meaning appear in parentheses

two sessions was about five minutes, and participants were informed at the end of the first session that another session was to follow and that all the requirements for the second session were same as the first session. The sequence of onset and rime session was coded as OR in the subsequent analyses and was treated as a between-subjects factor.

Procedure

In the associative-learning session, each participant had nine cards, each of which had one prompt-response pair printed on it. After participants indicated that they had memorized all of the pairs (memorization took five minutes on average), a practice test session was administered to help participants become familiar with the paradigm and to confirm that they had memorized all pairs. In the practice session, the nine prompt characters were presented on a computer screen in a random order, and the participants were asked to say the corresponding response character as quickly and accurately as possible. Only after participants' performance indicated that they were familiar with the procedure and the materials (i.e., being able to provide the correct response for each prompt within 1,000 ms) did the formal testing session begin. Both the practice and formal tests were implemented using DMDX software (Forster & Forster, 2003). For each list, the cards of the three item pairs were shown to the participants, and none of the participants needed extra time to memorize the pairs before the testing. During the test session, each trial began with a 200-ms, 1,000-HZ warning tone and a cross ("+") fixation presented at the center of the screen. 600 ms after the offset of the tone, a prompt in size 48 font appeared at the center of the screen for 150 ms. Participants were instructed to say the corresponding response character aloud, as quickly and accurately as possible. An AUDIO TECHNICA ATR-20 microphone was used as the voice key of the DMDX program to

Table 2   The six lists of all items used for the form-preparation task in the rime session in both Experiments

	Homogeneous Condition	Heterogeneous Condition
List 1	钩(gōu, *hook*) —— 绳 (shéng, *rope*) 盖 (gài, *cover*) —— 篷 (péng, *awning*) 困(kùn, *sleepy*) —— 疼(téng, *pain*)	钩(gōu, *hook*) —— 绳 (shéng, *rope*) 石(shí, *stone*) —— 土(tǔ, *soil*) 球 (qiú, *ball*) —— 泡(pào, *bubble*)
List 2	猫 (māo, *cat*) —— 鼠(shǔ, *rat*) 表(biǎo, *table*) —— 谱(pǔ, *music score*) 石(shí, *stone*) —— 土(tǔ, *soil*)	困(kùn, *sleepy*) —— 疼(téng, *pain*) 表(biǎo, *table*) —— 谱(pǔ, *music score*) 笛 (dí, *flute*) —— 哨 (shào, *whistle*)
List 3	笛 (dí, *flute*) —— 哨 (shào, *whistle*) 球 (qiú, *ball*) —— 泡(pào, *bubble*) 膜 (mó, *membrane*) —— 套(tào, *cover*)	盖 (gài, *cover*) —— 篷 (péng, *awning*) 猫 (māo, *cat*) —— 鼠(shǔ, *rat*) 膜 (mó, *membrane*) —— 套(tào, *cover*)

Note. In each cell, the left-hand monosyllabic word is the prompt and the response word is to the right. After each character, its Pinyin transcription followed by its English meaning appear in parentheses

record participants' response times. The next trial began 200 ms after a response was given, or after 1,500 ms if no response was given. During the experiment, the first author sat behind the participants and scored their naming accuracy, and a voice recorder was used to record participants' responses so that the experimenter was able to re-listen to the responses later if needed.

Results

Response-time (RT) analyses were based on correct trials only (approximately 1 % of the trials were incorrect and were deleted). Approximately 3 % of the RT data was removed because the response failed to trigger the voice key or because hesitations or disfluencies occurred. Analyses were carried out in R, an open-source programming environment for statistical computing (R Development Core Team, 2008) with the lme4 package (Bates, Maechler, Bolker, & Walker, 2013) and lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2013) for linear mixed effects modeling (LLM, GLMM). For the onset session, Context, Block, OR (i.e., the sequence of onset and rime session) and Sequence were entered in the model as fixed effects whereas Participant and Item were entered as random slopes where Context served as the intercept. The mixed effects model entered in R for the analysis of RT was "RT ~ Block* Sequence * OR * Context+ (1 + Context|Participant) + (1 + Context|Item)." Onset was not included as a fixed effect because it had already been embedded in items (i.e., each item had its own onset) and Item has been included as a random slope. Therefore, we decided to collapse onsets and to investigate the overall Context effect.

Likewise, for the rime session, Context, Block, OR, and Sequence were entered in the model as fixed effects whereas Participant and Item were entered as random slopes with Context serving as the random intercept. An analysis of accuracy rate in both sessions followed a similar procedure with the generalized linear (binomial) mixed-effects regression, but failed to show any significant results. For the RT data, we reported analyses of the onset session and rime session separately in order to make the model concise. See the upper part (i.e., Orthographic Type is character) of Table 3 for summaries of the descriptive data of both the onset and rime sessions.

For the onset session, the main effect of Context was not significant ($F (1, 10.2) = .014, p = .908$). Context did not show significant interaction with any other variables ($F (2, 3262.5) =1.023, p = .360$ for Context and Block; $F (1, 11.9) =1.575, p = .233$ for Context and Sequence; $F (1, 11.9) =1.216, p = .292$ for Context and OR, respectively). A significant Block main effect was shown ($F (2, 3262.4) =10.332, p < .001$), probably due to a practice effect given that the RT showed a trend of being faster in the later blocks (Block 1: 749 ms; Block 2: 731 ms; Block 3: 721 ms). There was also a significant Block × OR interaction ($F (2, 3262.5) =5.6116, p = .004$), likely due to the fact that participants showed a more salient Block effect when they received the onset session first ($p < .001$) compared to when they received the rime session first ($p = .178$). See Table 4 for the full results.

For the rime session, the main effect of Context did not reach significance ($F (1, 15.6) = .348, p = .564$), and a main effect of Block was shown ($F (2, 3283.6) =6.186, p = .002$). The participants' RT tended to be faster in the later blocks (Block 1: 751 ms; Block 2: 737 ms; Block 3: 728 ms). Context did not show significant interaction with any other variables

Table 3 Descriptive data of participants' performance in both orthographic types (i.e., both experiments) with mean reaction time (M), error rates (E%), standard errors (SE), and preparation effects

Orthographic Type	Session		Homogeneous	Heterogeneous (Control)	Preparation Effect (ms)
Character (Experiment 1)	Onset	M	734	735	1
		E%	.97	1.04	
		SE	4.40	4.20	
	Rime	M	744	735	−9
		E%	.91	.97	
		SE	4.25	4.26	
Pinyin (Experiment 2)	Onset	M	663	690	27*
		E%	1.15	1.04	
		SE	3.01	3.71	
	Rime	M	714	686	−28*
		E%	1.20	1.16	
		SE	3.82	3.41	

* $p < .05$

Table 4 Results of the ANOVA approach to linear mixed-effect model analysis of reaction times of the onset session in Experiment 1

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
Block	419108	209554	2	3262.4	10.332	0.000***
Sequence	133228	133228	1	12	3.872	0.073
OR	188083	188083	1	12	4.191	0.063
Context	5	5	1	10.2	0.014	0.908
Block × Sequence	42529	21264	2	3262.4	1.327	0.265
Block × OR	249970	124985	2	3262.5	5.612	0.004*
Sequence × OR	37894	37894	1	12	0.611	0.449
Block × Context	32244	16122	2	3262.5	1.023	0.360
Sequence × Context	23738	23738	1	11.9	1.575	0.233
OR × Context	25263	25263	1	11.9	1.216	0.292
Block × Sequence × OR	29130	14565	2	3262.4	0.662	0.516
Block × Sequence × Context	38785	19393	2	3262.4	1.090	0.336
Block × OR × Context	26206	13103	2	3262.5	0.622	0.537
Sequence × OR × Context	12754	12754	1	11.9	0.571	0.464
Block × Sequence × OR × Context	11999	5999	2	3262.5	0.269	0.764

Note. $* p < .05$; $*** p < .001$

$(F (2, 3283.6) = .229, p = .795$ for Context and Block; $F (1, 12.0) = 3.135, p = .102$ for Context and Sequence; $F (1, 12.0) = 2.162, p = .167$ for Context and OR, respectively). There were also some complicated three-way interactions. See Table 5 for the full results.

Discussion

The critical results of Experiment 1 were: (1) participants failed to show an onset or rime effect; and (2)

Context did not show significant interaction with any of the other variables. The absence of onset facilitation was consistent with previous literature (e.g., Chen, et al., 2002; O'Seaghdha, et al., 2010) that native Mandarin speakers failed to benefit from foreknowledge of the onset during the associative naming of materials written in Chinese characters. The results suggested that presenting materials in characters did not cue onset to serve as the preparation unit among native Mandarin-speaking adults. Participants also

Table 5 Results of the ANOVA approach to linear mixed-effect model analysis of reaction times of the rime session in Experiment 1

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
Block	265974	132987	2	3283.6	6.186	0.002**
Sequence	23196	23196	1	12	0.773	0.397
OR	92377	92377	1	12	3.706	0.078
Context	7534	7534	1	15.6	0.348	0.564
Block × Sequence	69531	34766	2	3283.7	1.591	0.204
Block × OR	19900	9950	2	3283.6	0.462	0.630
Sequence × OR	66698	66698	1	12	3.689	0.079
Block × Context	8754	4377	2	3283.6	0.229	0.795
Sequence × Context	69290	69290	1	12	3.135	0.102
OR × Context	47534	47534	1	12	2.162	0.167
Block × Sequence × OR	41604	20802	2	3283.5	0.946	0.388
Block × Sequence × Context	906	453	2	3283.7	0.019	0.981
Block × OR × Context	161276	80638	2	3283.6	3.685	0.025*
Sequence × OR × Context	114589	114589	1	12	5.226	0.041*
Block × Sequence × OR × Context	675	337	2	3283.5	0.015	0.985

Note. $* p < .05$; $** p < .01$

failed to show any significant rime effects, suggesting that rime may not be the preparation unit either. In addition, Context did not show significant interaction with Block, Sequence, or OR, suggesting that even practice or increased familiarity with the paradigm failed to cue participants for the onset or rime unit.

Experiment 2: Implicit priming using Pinyin symbols

Experiment 2 aimed to investigate the preparation unit speakers select when they were asked to memorize and name words presented in Pinyin. If the absence of onset facilitation in Experiment 1 is due to the influence of the morphosyllabic Chinese characters, participants should show an onset-facilitative effect in Experiment 2 considering that onset is represented explicitly in Pinyin and children are taught to read Pinyin in the sequence of onset-rime-whole syllable assembly. In terms of the rime session, if a similar language production procedure in Dutch applies to Chinese, then the same rime condition should not lead to a form-preparation effect when the materials are presented in Pinyin (see Meyer, 1991). If Pinyin does serve as a cue for the rime unit as the preparation unit, a rime facilitative effect should be expected.

Participants

Participants consisted of 16 native Mandarin-speaking students with normal or corrected-to-normal vision and without speech impairment from the same subject pool in Experiment 1. None of these participants participated in Experiment 1. There were nine females and seven males, whose ages ranged from 21 to 29 years (M=24.3, SD=1.91). All of the participants were paid for participation.

Methods

The items, design, and procedure were the same as those in Experiment 1. The only difference was that the materials were written in Pinyin in Experiment 2. In other words, both the associative pairs in the learning session and the prompts for the retrieval during the naming session were given in Pinyin.

Results

The data-cleaning procedure and data analysis were the same as in Experiment 1. In total, 4.67 % of the data were removed, among which about 1.8 % of the data

were incorrect trials, and the remaining data loss was because the response failed to trigger the voice key or because hesitations or disfluencies occurred. For both the onset and rime sessions, the analysis of accuracy rate again did not show any significant results. For the RT data, we reported analyses of the onset session and rime session separately again to avoid a lengthy model that would include non-critical interactions. See the bottom part (i.e., Orthographic Type is Pinyin) of Table 3 for summaries of the descriptive data of both the onset and rime sessions.

For the onset session, there was a significant Context main effect (i.e., onset facilitation) (F (1, 10.0) =5.673, p = .039). A Block main effect was shown (F (2, 3240.0) =10.921, p < .001), and, similar to Experiment 1, this block effect may be due to a practice effect since faster RTs were shown in the later blocks (Block 1: 690 ms; Block 2:669 ms; Block 3: 670 ms). Context did not show significant interaction with any other factors (F (2, 3239.9) = 1.066, p = .345 for Context and Block; F (1, 12.0) = .095, p = .763 for Context and Sequence; F (1, 12.0) = .120, p = .735 for Context and OR, respectively). See Table 6 for the full results.

For the rime session, the main effect of Context reached significance (i.e., rime inhibition) (F (1, 9.2) =7.850, p = .020), and there was a significant interaction between Context and Sequence (F (1, 12.1) =7.626, p = .017). Participants showed a 13-ms interference effect when they received a heterogeneous list first, whereas they showed a 42-ms interference effect when a homogeneous list was given first. A main effect of Block was shown (F (2, 3222.4) =26.683, p < .001). The RT in Block 1 (723 ms) was higher than that in Block 2 (692 ms) and Block 3 (686 ms), suggesting a practice effect. However, Context did not show a significant interaction with Block (F (2, 3222.4) = .421, p = .657) or OR (F (1, 12.1) =1.275, p = .281). There were also some complicated three-way interactions. See Table 7 for the full results.

To further examine the effect of orthographic form-cuing on the preparation unit, we combined the onset sessions in Experiments 1 and 2 to examine the interaction between Orthographic Type (i.e., character vs. Pinyin) and Context (same onset vs. control). In the linear mixed effects model, Context, Block, OR, Orthographic Type (labeled "Ortho" in R for brevity) and Sequence were entered into the model as fixed effects whereas Participant and Item were entered as random slopes with Context serving as the intercept.

The critical result was a significant Ortho × Context interaction (F (1, 23.9) =9.197, p = .006). When the orthographic type was Pinyin, the mean RT in the homogeneous condition was faster than that in the heterogeneous condition (27 ms). In contrast, when the

Table 6 Results of the ANOVA approach to linear mixed-effect model analysis of reaction times of the onset session in Experiment 2

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
Block	301703	150851	2	3240	10.921	0.000***
Sequence	1544	1544	1	12	0.210	0.655
OR	7441	7441	1	12	0.140	0.715
Context	78127	78127	1	10	5.673	0.039*
Block × Sequence	36571	18286	2	3239.9	1.288	0.276
Block × OR	1306	653	2	3240	0.042	0.959
Sequence × OR	33150	33150	1	12	0.458	0.511
Block × Context	28805	14403	2	3239.9	1.066	0.345
Sequence × Context	1330	1330	1	12	0.095	0.763
OR × Context	1543	1543	1	12	0.120	0.735
Block × Sequence × OR	16927	8463	2	3240.1	0.602	0.548
Block × Sequence × Context	68181	34090	2	3239.8	2.486	0.083
Block × OR × Context	19508	9754	2	3239.9	0.733	0.480
Sequence × OR × Context	10831	10831	1	12	0.797	0.390
Block × Sequence × OR × Context	68265	34133	2	3240	2.478	0.084

Note. * $p < .05$; *** $p < .001$

orthographic type was Chinese character, the mean RTs in different contexts did not show a significant difference (only a 1 ms difference). Context did not show significant interaction with any other factors ($ps > .10$). The Block main effect was significant ($F (2, 6518.3) = 19.270$, $p < .001$). Orthographic Type showed a significant main effect ($F (1, 24.0) = 6.116$, $p = .021$). The mean RT in the Pinyin condition (677 ms) was faster than the mean RT in character condition (734 ms), most likely because Pinyin facilitated the encoding of phonological information in the task. See Appendix A for the full results.

Likewise, we also combined the rime sessions in the two experiments and included Orthographic Type as a new fixed effect in the linear mixed effects model.

Table 7 Results of the ANOVA approach to linear mixed-effect model analysis of reaction times of the rime session in Experiment 2

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
Block	871661	435831	2	3222.4	26.6828	0.000***
Sequence	89093	89093	1	12	0.8958	0.363
OR	12983	12983	1	12	0.1173	0.738
Context	126251	126251	1	9.2	7.8504	0.020*
Block × Sequence	19668	9834	2	3222.5	0.5974	0.550
Block × OR	28843	14421	2	3222.4	0.8772	0.416
Sequence × OR	13529	13529	1	12	0.5012	0.492
Block × Context	13277	6639	2	3222.4	0.4208	0.657
Sequence × Context	120396	120396	1	12.1	7.6261	0.017*
OR × Context	20100	20100	1	12.1	1.2745	0.281
Block × Sequence × OR	120815	60408	2	3222.4	3.6401	0.026*
Block × Sequence × Context	157323	78661	2	3222.5	4.8172	0.008**
Block × OR × Context	2261	1130	2	3222.4	0.0687	0.934
Sequence × OR × Context	1303	1303	1	12.1	0.0772	0.786
Block × Sequence × OR × Context	16205	8102	2	3222.4	0.4988	0.607

Note. * $p < .05$; ** $p < .01$; *** $p < .001$

Orthographic Type and Context did not show a significant interaction in the new model (*F* (1, 24.0) =2.123, *p* = .158). However, separate analyses had shown that a significant rime interference effect was only present in the Pinyin condition but not the character condition. When the orthographic type was pinyin, a 28 ms rime interference effect was presented; when the orthographic type was a character, participants' RT was also slower in the homogeneous condition than in the heterogeneous condition (9 ms), and this small trend of rime inhibition may have contributed to the non-significant Ortho × Context interaction. See Appendix B for the full results.

Discussion

For the onset session, the critical results in Experiment 2 were: (1) participants showed significant onset facilitation when the materials were written in Pinyin; and (2) there was no significant interaction between Context and any other variables. In addition, combining the onset session of Experiments 1 and 2, a significant Orthographic Type × Context interaction was shown (*p* = .006). This interaction and the significant onset facilitation in Experiment 2 were consistent with our prediction that participants tended to be more sensitive to the smaller unit (i.e., onset) when Chinese words were written in Pinyin, given the alphabetic nature of the Pinyin system and that presenting the materials in Pinyin may have cued onset-rime division.

For the rime session, inconsistent with our prediction, the current study showed a significant rime interference effect in Experiment 2. We speculate that this might be a result of lexical competition since the repetition of the target rime may activate the rime neighborhood of the target word. After the participants said the response word of a trial, the lexical information of its rime neighbors, including the response word of the next trial, might be activated, but the lexical representation of the current trial is in opposition to that of the next trial, thus inhibiting the pre-activation of the next trial and leading to an interference effect. This explanation was built upon Lukatela and Turvey (1996), in which a rime inhibitory effect was shown among native English speakers in a word-naming task with a priming paradigm. In Lukatela and Turvey (1996), when the stimulus onset asynchrony (SOA) was 36 ms or 70 ms, participants named a target word (e.g., *nose*) more slowly when its prime shared the target's rime (e.g., *hose*). Their explanation was that the prime (e.g., *hose*) helped the lexical activation of both the prime and its rime neighbors, including the target (e.g., *nose*). The lexical

activation of the prime was in competition with that of the target, and thus inhibited the pre-activation of the target. This inhibition effect occurred only in the same rime but not in same onset session, probably due to the fact that rime is the later component of a monosyllabic word. By the time speakers access the rime unit, they have also activated the lexical information; in contrast, by the time they have accessed the onset unit, only the facilitative effect occurred while the lexical information may not yet have been retrieved.

Another significant interaction that involved Context was the Sequence × Context interaction, which might be a result of the practice effect, considering that participants showed larger interference effects when they received a homogeneous list before its corresponding heterogeneous list. This was also in line with the main effect of Block, which could be due to a practice effect as well.

When the rime sessions of Experiments 1 and 2 were combined, the interaction between Orthographic Type and Context did not reach significance (*p*= .158). However, a post-hoc analysis and separate analysis in Experiments 1 and 2 showed that the interference effect was larger in Experiment 2 than in Experiment 1, and only Experiment 2 showed a significant rime effect. An explanation is that the rime unit is explicitly represented in Pinyin (Experiment 2), thus making the rime unit more salient to participants in Experiment 2. As a result, the lexical competition between the rime neighbors was stronger in Experiment 2.

General discussion

We used two form-preparation tasks to investigate the effect of orthographic form-cuing on the preparation unit in planning spoken-word production. The findings in the onset sessions of the two experiments suggest that when the target words in a production task are visually presented in advance (i.e., the orthographic information is explicitly provided in the associative-learning session), the preparation unit is cued by the orthographic form of the presented visual words. When the visual words are presented in the form of an alphabetic orthographic system (e.g., Pinyin), the phonological encoding is done in the smaller unit – onset; when the visual words are presented in the form of a morphosyllabic orthographic system (e.g., Chinese characters), phonological information may be encoded in a larger unit. Although it is difficult to tease apart the effect of phoneme and onset because there are no

consonant clusters in Mandarin, the simple syllabic structure of Chinese and the experience of Pinyin in early primary grades suggest that in practice Pinyin encourages phonological coding in onsets and rimes but not in phonemes (see Ziegler & Goswami, 2005, for review). Thus, it is more likely that the onset instead of the phoneme is the functional preparation unit that is responsible for the facilitation in the homogenous context in Experiment 2.

Onset facilitation

To the best of our knowledge, the present study is among the first that investigated the influence of orthographic form (e.g., alphabetic vs. morphosyllabic) on spoken word production. Our findings suggest that there is some flexibility in native Mandarin speakers' selection of the preparation unit. The written words may cue participants to encode the phonological information in a way that is consistent with its orthographic form since readers are flexible in selecting the preparation unit in different contexts when cued by different orthographic forms. Previous literature suggested that skilled readers retrieve and manipulate the phonological unit of visual words when comprehending and memorizing print materials (see Crowder, 1982; Mann, 1986; Perfetti & McCutchen, 1982; Stanovich, 1982, for reviews), and this explains why the adult participants in the current study tended to select the onset when memorizing and associatively naming materials written in Pinyin but not characters—it is more consistent with the alphabetic nature of the Pinyin system. The phonological organization of the Pinyin system, which is highly correlated to its orthographic form, encourages an onset-rime division. In contrast, the orthographic form of the Chinese character does not cue the onset-rime division. Chen and Chen (2013) argued that the syllable segment is the preparation unit in Chinese and that this is an intrinsic property of the production system. The researchers showed consistent results in simple picture-naming tasks with the form-preparation paradigm, in which no associative-learning session was involved. Taking all these findings together, we suggest that it is likely that native Mandarin speakers prefer to encode the phonological information in syllables in spoken-word production; however, we further suggest that skilled readers are flexible in encoding phonological information according to the cue of the context (e.g., the orthographic form of the visually presented stimuli), since they have acquired knowledge of multiple orthographic forms (i.e., Pinyin and characters). The explicit phonological information in Pinyin cues the readers to shift their preference from a large unit (syllable) to a smaller unit (onset) in planning spoken words.

For the general architecture of production, we propose that after lexical selection at the lemma level, orthographic form may be involved prior to phonological encoding. If the selected lexical items have been visually presented in advance, the orthographic form of the visual words cued speakers to encode a certain phonological unit. In the present study, Pinyin promotes the smaller phonological unit, onset, and Chinese characters may have promoted the larger unit, syllable. In addition, we propose that orthographic cues may be used to facilitate information retrieval as well. When the items in a production task are visually presented as input in advance, literate speakers rely on orthographic cues to decode the phonological information; in the production task carried out later, using the same orthographic cues may facilitate expedient retrieval of phonological information.

The present study suggested that the orthographic input cues phonological retrieval and encoding in a production system. Likewise, Chen and Li (2011) suggested that the output format in a production system (e.g., word typing vs. word naming) also cues word-form encoding among native Mandarin speakers. When the target words shared the onset consonants with the prime words, participants showed faster RTs compared to when targets and primes did not share onsets when typing the target words. However, this facilitation was not shown in a word-naming task. Both typing and naming tasks require accessing the phonological codes, but in word typing, participants need to use the Pinyin system to type character words and they need to anticipate the segment-based finger movements, compared to naming where participants are not required to access the segmental information. This contrast promotes the encoding of the smaller unit, the onset, as the preparation unit. The Pinyin input cue in the present study may have functioned similarly to the word-typing output cue in Chen & Li (2011), and both types of cues may have encouraged participants to attend to sub-syllabic units.

The rime interference effect

The present study showed a significant interference effect when the response words had the same rime when the orthographic type was Pinyin. As discussed earlier, this unpredicted interference effect may arise as a result of lexical competition. Unfortunately, the current experiment is not able to address the underlying mechanism

of the rime interference effect directly and future research is needed to pin down the exact factors that contribute to this effect. Meyer (1991) showed that the fore-knowledge of the rime unit did not have an effect in Dutch. A possible explanation for the inconsistent findings regarding the shared rime is cross-linguistic differences. Compared to Chinese, which does not allow consonant clusters within a syllable, Dutch has a much richer set of onsets. In fact, words with consonant clusters as the onsets were included in Meyer (1991), such as *snoek* (pike) and *vloek* (curse). However, only single consonants are legal as onsets in Chinese. Therefore, a shared rime in Dutch may not be enough to activate rime neighbors to lead to an inhibition effect. In contrast, in Chinese with a much simpler syllable structure, a shared rime may be sufficient to lead to activation of rime neighbors, resulting in lexical competition. Future research is needed to directly address the question of the conditions under which the rime interference effect occurs.

Limitations and future directions

The present study has some limitations. First, a possible factor that may have contributed to onset facilitation is participants' English exposure and proficiency. All participants were undergraduate or graduate students at a mid-Atlantic University. Compared with native Mandarin speakers in China, extensive exposure to English may have resulted in higher English language proficiency and may have facilitated the participants in the current study to attend more to a small phonological unit such as the onset. This proficiency effect would be consistent with previous research that suggests that bilingual speakers' experience with their second language (L2) may influence the preparation unit in their native language (L1) (Verdonschot, Nakayama, Zhang, Tamaoka, & Schiller 2013). In a masked priming experiment, native Mandarin Chinese speakers with high English proficiency were asked to read a series of Chinese characters as quickly and accurately as possible. One target character was shown at one time, and the prime prior to the target may share the same onset, same initial syllable, or nothing systematically with the target. Only when the syllable structure of the prime and the target overlapped (i.e., CV-CV or CVC-CVC), did participants show significant onset facilitation. The present study did not show onset facilitation in the character condition, which is probably related to the fact that we included different syllable structures in the three items of a homogeneous list (i.e., C+ Simple

V, C+ Simple V+C, and C+ Diphthong). Nevertheless, we showed a significant onset-facilitative effect in the Pinyin condition even when the syllable structure remained different across the items. This finding suggests that Pinyin did indeed cue the onset unit for encoding. In summary, the onset facilitation in Experiment 2 is more likely due to the phonological organization of Pinyin, or at least due to both the alphabetic nature of the Pinyin form and participants' English proficiency. In future research, participants' English proficiency level could be measured to be included as a covariate in the linear mixed effects model to investigate its potential influence.

Second, the absence of onset facilitation in Experiment 1 only suggested that onset is not the preparation unit when the orthographic type is the character; however, it does not address the exact preparation unit selected by speakers. A follow-up experiment with two-character words that share the same initial syllable segment in the homogeneous context should be able to help determine whether it is indeed the syllable segment that speakers select for planning spoken words when cued by a morphosyllabic orthographic form. Third, in the same rime session, words in the same rime context did not only share the same rime segment, but also the same tone. The interference effect in Experiment 2 may be a result of both rime segment and tone or only one of them. Chen et al. (2002) showed that tone alone did not play a significant role in preparing for spoken word production in Chinese; however, mismatched tone did decrease the size of the preparation effect. Thus, future research is needed to tease apart whether rime segment alone or rime segment plus tone could lead to the inhibition. Fourth, the mechanism underlying the rime inhibition effect merits further investigation. We speculated that the different syllable structures may play a role in different patterns of the rime effect between Dutch and Mandarin speakers. A possible direction in future research is to vary the syllable structure of the words in the same language and to compare the rime effect on items with complicated versus simple syllable structures. Finally, orthographic information was explicitly represented in both experiments, suggesting that the preparation unit in Mandarin can be cued by the orthographic form that is visually presented. However, with regard to the preparation unit, it remains unclear when orthographic cue is not visually presented (e.g., in a simple picture-naming task). Although Chen and Chen (2013) adopted a simple picture-naming task, their participants from Taiwan have different literacy experiences from our participants from Mainland China (e.g., Mandarin speakers in Taiwan do not receive exposure to

Pinyin), and so future research is needed for a direct comparison.

word production, and orthographic form may serve as a cue to promote a certain phonological unit for encoding. We suggest that a revised spoken word production model take orthographic cues into consideration.

Conclusion

The present study showed that literate speakers are flexible in selecting the preparation unit in spoken

Appendix A

Table 8 Results of the ANOVA approach to linear mixed-effects model analysis of reaction times when the onset session of Experiments 1 and 2 are combined

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
Block	681558	340779	2	6518.3	19.270	0.000***
Sequence	51085	51085	1	24	1.266	0.272
OR	35310	35310	1	24	1.544	0.226
Ortho	380886	380886	1	24	6.116	0.021*
Context	34166	34166	1	10	1.595	0.235
Block × Sequence	19734	9867	2	6518.3	0.962	0.382
Block × OR	102501	51251	2	6518.3	3.278	0.038*
Sequence× OR	54677	54677	1	24	1.072	0.311
Block × Ortho	39205	19602	2	6518.2	1.367	0.255
Sequence × Ortho	104614	104614	1	24	3.066	0.093
OR × Ortho	138077	138077	1	24	3.079	0.092
Block × Context	58446	29223	2	6518.2	1.945	0.143
Sequence × Context	18927	18927	1	23.9	1.452	0.240
OR × Context	3065	3065	1	23.9	0.467	0.501
Ortho × Context	153778	153778	1	23.9	9.197	0.006**
Block × Sequence × OR	25523	12762	2	6518.3	0.507	0.603
Block × Sequence × Ortho	57328	28664	2	6518.2	1.613	0.199
Block × OR × Ortho	131829	65914	2	6518.3	3.628	0.027*
Sequence × OR × Ortho	495	495	1	24	0.011	0.916
Block × Sequence × Context	1962	981	2	6518.1	0.016	0.984
Block × OR × Context	118	59	2	6518.2	0.037	0.964
Sequence × OR × Context	20567	20567	1	23.9	1.282	0.269
Block × Ortho × Context	1560	780	2	6518.1	0.072	0.931
Sequence × Ortho × Context	9547	9547	1	23.9	0.632	0.435
OR × Ortho × Context	20486	20486	1	23.9	1.147	0.295
Block × Sequence × OR × Ortho	27346	13673	2	6518.3	0.748	0.473
Block × Sequence × OR × Context	55503	27752	2	6518.1	1.666	0.189
Block × Sequence × Ortho × Context	108225	54113	2	6518	3.130	0.044*
Block × OR × Ortho × Context	45822	22911	2	6518.2	1.235	0.291
Sequence × OR × Ortho × Context	71	71	1	23.9	0.004	0.951
Block × Sequence × OR × Ortho × Context	15952	7976	2	6518.2	0.435	0.647

Note.:*p< .05; ** p< .01; *** p< .001. The table shows effects for the fixed-effect factors (Context [homogeneous vs. heterogeneous], Orthographic Type [character vs. Pinyin], OR [onset first vs. rime first], Block [1, 2, 3], and Sequence [homogeneous first vs. heterogeneous first])

Appendix B

Table 9 Results *of the ANOVA approach to linear mixed-effects model analysis of RT when the rime session of Experiments 1 & 2 are combined*

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
Block	1032765	516382	2	6520.8	27.144	0.000***
Sequence	1035	1035	1	24	0.002	0.963
OR	45906	45906	1	24	2.573	0.122
Ortho	34330	34330	1	24	2.048	0.165
Context	66745	66745	1	16.3	3.503	0.079
Block × Sequence	18006	9003	2	6520.8	0.445	0.641
Block × OR	29996	14998	2	6520.6	0.801	0.449
Sequence× OR	19209	19209	1	24	0.738	0.399
Block × Ortho	98945	49472	2	6520.8	2.499	0.082
Sequence × Ortho	30220	30220	1	24	1.657	0.210
OR × Ortho	19483	19483	1	24	1.267	0.272
Block × Context	2964	1482	2	6520.7	0.087	0.917
Sequence × Context	144100	144100	1	24	7.543	0.011*
OR × Context	15857	15857	1	24	0.793	0.382
Ortho × Context	40797	40797	1	24	2.123	0.158
Block × Sequence × OR	18828	9414	2	6520.7	0.517	0.597
Block × Sequence × Ortho	72598	36299	2	6520.8	1.860	0.156
Block × OR × Ortho	19596	9798	2	6520.9	0.504	0.604
Sequence × OR × Ortho	78594	78594	1	24	3.458	0.075
Block × Sequence × Context	85305	42652	2	6520.7	2.250	0.105
Block × OR × Context	88763	44381	2	6520.6	2.228	0.108
Sequence × OR × Context	75941	75941	1	24	3.917	0.059
Block × Ortho × Context	18912	9456	2	6520.8	0.523	0.593
Sequence × Ortho × Context	5228	5228	1	24	0.252	0.620
OR × Ortho × Context	61973	61973	1	24	3.257	0.084
Block × Sequence × OR × Ortho	143562	71781	2	6520.8	3.707	0.025*
Block × Sequence × OR × Context	9787	4893	2	6520.7	0.255	0.775
Block × Sequence × Ortho × Context	72452	36226	2	6520.8	1.877	0.153
Block × OR × Ortho × Context	77887	38944	2	6520.8	2.039	0.130
Sequence × OR × Ortho × Context	92161	92161	1	24	4.781	0.039*
Block × Sequence × OR × Ortho × Context	6842	3421	2	6520.7	0.178	0.837

Note.: $p < .05$; **: $p < .01$; ***: $p < .001$. *The table shows effects for the fixed-effect factors (Context [homogeneous vs. heterogeneous], Orthographic Type [character vs. Pinyin], OR [onset first vs. rime first], Block [1, 2, 3] and Sequence [homogeneous first vs. heterogeneous first])*

References

Alario, F.-X., Perre, L., Castel, C., & Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition, 102,* 464–475.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.0-5. http://CRAN.R-project.org/package=lme4.

Blackwood, D. H., and W. J. Muir. "Cognitive Brain Potentials and Their Application." *The British Journal of Psychiatry. Supplement*, no. 9 (1990): 96–101.

Chen, T.-M., & Chen, J.-Y. (2013). The syllable as the proximate unit in Mandarin Chinese word production: An intrinsic or accidental property of the production system. *Psychonomic Bulletin & Review, 20,* 154–162. doi:10.3758/s13423-012-0326-7

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*(4), 751–781.

Chen, J. Y., & Li, C. Y. (2011). Word form encoding in Chinese word naming and word typing. *Cognition, 121*(1), 140–146.

Cholin, J., Schiller, N. O., & Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language, 50,* 47–61.

Crowder, R. G. (1982). *The psychology of reading.* New York: Cambridge University Press.

Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language, 49,* 119–132.

Ferreira, V. S. (2010). Language production. *Wiley Interdisciplinary Reviews:Cognitive Science, 1*(6), 834–844.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116–124.

Hanley, J. R. (2005). Learning to read in Chinese. In *The Science of Reading,* 296–315.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1), 101–144.

Jacobs, C. L., & Dell, G. S. (2014). 'hotdog', not 'hot''dog': the phonological planning of compound words. *Language, Cognition and Neuroscience, 29*(4), 512–523.

Kureta, Y., Fushimi, T., & Tatsumi, I. F. (2006). The functional unit of phonological encoding: Evidence for moraic representation in native Japanese speakers. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 32,* 1102–1119.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). *lmerTest: Tests for random and fixed effects for linear mixed effect models* (lmer objects of lme4 package). R package version 2.0-0. http://CRAN.R-project.org/package=lmerTest.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences, 22,* 1–75.

Lukatela, G., & Turvey, M. T. (1996). Inhibition of naming by rhyming primes. *Perception & Psychophysics, 58*(6), 823–835.

Mann, V. A. (1986). Phonological awareness: The role of reading experience. *Cognition, 24*(1), 65–92.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29,* 524–545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30,* 69–89.

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115,* 282–302.

Perfetti, C. A., & McCutchen, D. (1982). Speech processes in reading. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 7, pp. 237–269). New York: Academic Press.

Qu, Q., Damian, M. F., & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences, 109*(35), 14265–14270.

R Development Core Team. (2008). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes, 14*(2), 173–200.

Roelofs, A. (2006). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin & Review, 13*(1), 33–37.

Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 922–939. doi:10.1037/0278-7393.24.4.922.

Stanovich, K. E. (1982). Individual differences in the cognitive processes of reading: I. *Word decoding. Journal of Learning Disabilities, 15,* 485–493.

Shu, H., Peng, H., & McBride-Chang, C. (2008). Phonological awareness in young Chinese children. *Developmental Science, 11*(1), 171–181.

Verdonschot, R. G., Nakayama, M., Zhang, Q., Tamaoka, K., & Schiller, N. O. (2013). The proximate phonological unit of Chinese-English bilinguals: Proficiency matters. *PloS One, 8*(4), e61454.

Wang, M., & Gao, W. (2011). Subsyllabic unit preference in learning to read pinyin syllables. *Contemporary Educational Psychology, 36,* 142–151.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3–29.