



OPEN

DATA DESCRIPTOR

# A dataset of behavioral measures on Chinese word production in picture naming

Chen Feng<sup>1,2</sup> , Markus F. Damian<sup>3</sup> & Qingqing Qu<sup>1,2</sup> 

Most studies of language production have been conducted with speakers of alphabetic languages, but relatively little research has examined languages with non-alphabetic scripts, such as Chinese. Moreover, most work on language word production has investigated phonological output processing (i.e., speaking), whereas comparatively little research has focused on orthographic output, such as writing and typing. Work on non-alphabetic languages and/or written production is particularly promising, given that it speaks to universalities vs. specificity in terms of architectures and mechanisms underlying language processing across all world languages and modalities. The current article reports a dataset of word production in Chinese with spoken and written responses, which includes reaction times of 193,851 trials of naming 403 pictures obtained from 667 participants across 23 Chinese word production experiments. All data were collected in the same experimental environment and from participants with relatively homogenous characteristics, using the same protocols and parameters. The dataset enables researchers to explore how Chinese speakers produce spoken and/or written words, and to identify language-specific features underlying word production.

## Background & Summary

Language production involves successive planning stages, starting with conceptual preparation, followed by a stage of grammatical and lexical selection during which entries are retrieved from the mental lexicon and slotted into a grammatical structure. After selection of a word, a target lexical item is then given its sound form, which finally enables its articulation. Over the past decades, many studies have investigated these stages with the aim of identifying the architecture and processing dynamics that underlie these stages. Existing theories<sup>1–5</sup> differ considerably on various aspects but they also converge on a few critical assumptions, for instance, that planning to produce a word leads to not just the activation of the representations of the target word, but also elicits co-activation of representations of related words. Moreover, current models of language production generally assume that phonological encoding involves retrieval of abstract sound-sized segmental units or phonemes; however, these existing frameworks are largely based on Indo-European languages, and despite some universalities it cannot be assumed that all aspects generalize across all world languages. Representations and processing dynamics of word-form encoding are very likely not uniform across languages. This is because languages differ substantially in the phonological and/or orthographic systems, and the way in which phonologic and orthographic codes are mapped to each other. Indeed, several studies have demonstrated that syllables rather than phonemes are the primary unit of phonological encoding in Chinese<sup>6–8</sup> (although phonemes are also significant for Chinese speakers<sup>8</sup>), in contrast to what is generally assumed for alphabetic languages.

Most research on language production has been conducted in European languages, and relatively little research has examined non-European languages such as Chinese. In contrast to languages with alphabetic scripts, Chinese has a logographic writing system in which writing is represented by box-like characters and each character maps onto a syllable. Moreover, in alphabetic languages, orthographic symbols express speech sounds (notwithstanding considerably variability among languages regarding the transparency of this mapping); by contrast, the relation between spelling and sound in Chinese is largely opaque (e.g., the sound /shu/ can be written as the orthographically dissimilar words 书, 梳, 鼠, 树, or 薯). Research on Chinese is therefore particularly informative as the findings speak to cross-linguistic universalities and specificity in terms of architectures

<sup>1</sup>Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China.

<sup>2</sup>Department of Psychology, University of Chinese Academy of Sciences, Beijing, China. <sup>3</sup>School of Psychological Science, University of Bristol, Bristol, United Kingdom. ✉e-mail: [quqq@psych.ac.cn](mailto:quqq@psych.ac.cn)

and processing dynamics underlying spoken planning. In addition, most work on language word production has investigated spoken production (i.e., speaking), whereas comparatively little research has focused on the generation of orthographic output such as in writing and typing. Alongside the increasing need for written (rather than, or in addition to, spoken) language skills for successful e-communication in the modern world, written production as the most under-researched domain in the psycholinguistic field requires more investigation.

Experiments which involve a specific task and a limited set of items in which one or more factors are manipulated have been informative in revealing representations and processing dynamics underlying word production. A large-scale dataset would be beneficial by providing a rich source of information which takes into account differences of modalities, paradigms, and stimulus types, and hence allows to evaluate the stability and generalization of the conclusions. Moreover, such a dataset can be used to provide further constraints on models of language production by allowing to computationally simulate the production of a large set of words. Here, we introduce a large-scale dataset of word production in Chinese which is assembled from 193,851 trials of 23 experiments with 667 participants and 403 targets under different modalities, paradigms, and stimulus types. These data are derived from a series of published<sup>8–18</sup> and unpublished studies. Below we report the procedure of data acquisition, measurement computation, and data organization.

## Methods

**Data Acquisition.** Data were obtained from 667 participants across 23 experiments. All participants were native Chinese speakers who were recruited from nearby universities. They were all right-handed, with normal or corrected-to-normal vision and no reported language disorders. All experiments were approved by the Institutional Ethics Committee at the Institute of Psychology of the Chinese Academy of Sciences (No. H17030). Informed consent was obtained from all participants before the experiment, with their agreement for data sharing, as well as consent for the open publication of their data. We ensured that all necessary ethical considerations and informed consent procedures were followed throughout the studies.

Of all 23 experiments included in the dataset, 17 experiments have been published, whereas the remaining 6 have not. Details of the publication state (including the DOIs of all published experiments) of all experiments can be found in the Supplementary Table file and the ‘Metadata.csv’ file in our OSF repository at <https://doi.org/10.17605/OSF.IO/6GTZH><sup>19</sup>. All experiments involved the production of responses consisting of single words (with the exception of the “Color picture naming” task in which responses consisted of adjective-noun phrases; see “Description of Paradigm” in the Supplementary Information file), either in spoken or in written format, and implemented a very similar procedure. For exact procedural details, the reader is referred to a template paper from our group which compares spoken and written responses<sup>18</sup>. Characteristics of each experiment are described in the Supplementary Information file (see the subsections ‘Description of Paradigm’ and ‘Description of Experimental Design’) and the ‘Metadata.csv’ in the OSF repository. Dependent on the experimental paradigm, responses were elicited either via pictorial stimuli, or via associations to a cue word. Pictorial stimuli across all experiments included 403 distinct black-white line drawings. Most of these pictures were selected from standardized picture databases<sup>20–26</sup>, while a few were downloaded from open-access sources on the internet. All pictures were normalized to a size of 350 × 350 pixels. The pictures from the standardized databases can be found as a compressed archive named ‘Pictures.zip’ in the OSF repository.

In all experiments, native Chinese speakers were seated in a quiet room, around 60 cm away from a 21-inch CRT monitor (Sony G520; resolution: 1,024 × 768 pixels; refresh rate: 60 Hz). For each experiment using pictorial materials, participants were required to familiarize themselves with all pictures before tasks. Thereafter, they were instructed to speak overtly or write down the names of objects in black-and-white line drawings. During each experiment, the order of stimuli was pseudo-randomized so that stimuli were not repeated on consecutive trials. Experiments were conducted with E-prime (Psychology Software Tools, Inc. Pittsburgh, PA, USA) or DMDX<sup>27</sup>. In experiments which involved spoken word production, participants’ vocal responses were recorded via a microphone, and response latencies of recorded vocal responses were checked by Checkvocal<sup>28</sup> manually or Chronset<sup>29</sup> automatically. In experiments which involved written responses, these were collected using a graphic tablet and inking pen (Wacom Intuo 4), and written response latencies were measured as the interval between onset of the picture presentation and initial contact of the pen with the tablet. For all experiments, data files consisted of a spoken or written response latency (in milliseconds) and accuracy of a response on each given experimental trial. For both spoken and written modality, responses other than the expected ones were marked as errors. For spoken responses, trials were additionally defined as errors when they involved dysfluencies, stutters, repairs, or incomplete or missing responses, whereas for written responses, incomplete/missing or corrected responses were counted as errors. Data from all 23 experiments were collated into a single data file which can be found in the OSF repository as file “Raw data.csv”.

**Word-level and Character-level Measures.** Stimuli used in the experiments were monosyllabic, disyllabic or trisyllabic words which consist of one, two, or three Chinese characters. To aid in further exploration of the dataset, we provide several character-level and word-level measures of the 403 target words. For word-level measures, word length, word pronunciation, word frequency, age of acquisition, and phonological neighborhood density are provided. Word frequency values were obtained from the Chinese Linguistic Data Consortium (CLDC)<sup>30</sup> and SUBTLEX-CH<sup>31</sup>. Word-level age of acquisition (AoA) was obtained from a large-scale collection of age of acquisition ratings of simplified Chinese words<sup>32</sup>, and character-level AoA was obtained from a database of objective ages of acquisition of simplified Chinese characters<sup>33</sup>. For character-level measures, we provided stroke number, character pronunciation, character frequency and syllable frequency of each character in each word from the CLDC<sup>30</sup>, phonological neighborhood density of words and characters, and regularity of each character, from a large-scale Chinese lexical database<sup>34</sup>. For words do not exist in the lexical database, values of corresponding measurements were set to NA.

## Data Records

The current dataset is available on an open-access repository under the CC BY 4.0 License and can be accessed via the OSF repository at <https://doi.org/10.17605/OSF.IO/6GTZH><sup>19</sup>. All relevant raw datasets are provided in the csv files ‘Raw data.csv’. Measurements of words’ linguistic variables are provided in separate sheets of the file named ‘Linguistic measurements.csv’. The structure of these data files is listed below.

### Structure of the Raw data.

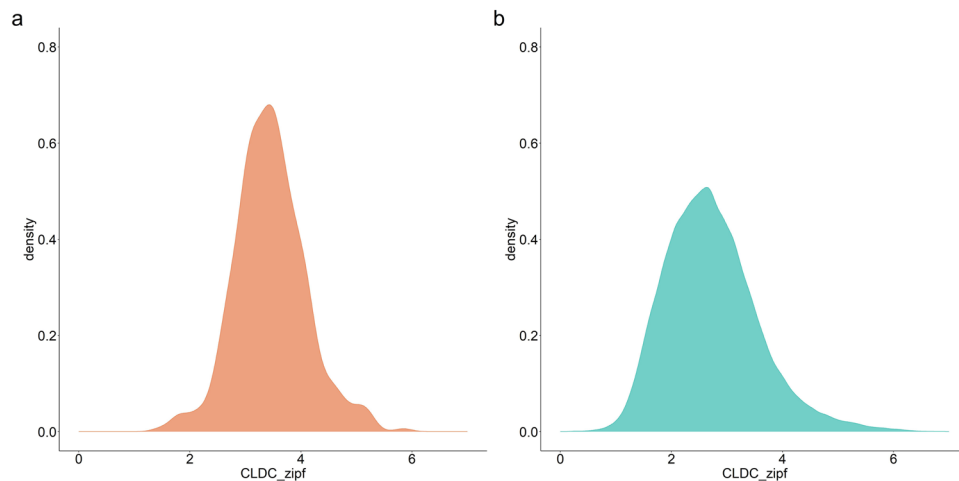
1. The column named ‘Experiment\_name’ indicates the experiment in which the current dataset was collected. The information of published studies is listed in the Supplementary Table file, and design of all experiments is described in the section of “Description of Experimental Design” in the Supplementary Information file.
2. The column named ‘Paradigm’ indicates the specific paradigm used to collect the data. The detailed paradigms are described in the section of “Description of Paradigm”.
3. The column named ‘Modality’ indicates the modality of response (speaking or writing) which participants were required to respond during that experiment.
4. The column named ‘Subject’ represents the ID of participants of that experiment.
5. The three columns starting with ‘Object\_’ show the names of the to-be-produced object in the form of Chinese Characters, Chinese Pinyin, and English.
6. The column named ‘RT’ shows the response latency of each trial. All RTs provided in the “Raw data.csv” document and used for quantitative validation below are raw, rather than trimmed. The column named ‘ACC’ shows the response accuracy of each trial, with 1 indicating a correct answer, 0 indicating a wrong answer, and –1 indicating no response was made.
7. The column named ‘Condition’ represents the conditions to which that trial belongs. For experiments with multiple independent variables, levels of all variables are combined with underscores ‘\_’ between them. For example, the condition ‘Homogeneous\_Cycle1’ indicates a trial which was produced in the first cycle of a homogeneous block.

### Structure of the linguistic measurements.

1. The column named ‘Words’ represents the words for which these variables were provided.
2. The column named ‘Pinyin’ shows the pronunciation of each word.
3. The column named ‘En\_translation’ shows the translation from the Chinese word into English.
4. The column named ‘Word\_length’ indicates how many characters/syllables each word has.
5. The three columns named ‘CLDC\_wfpm’, ‘CLDC\_logfpm’ and ‘CLDC\_zipf’ show three measurements of word frequency which are obtained from the Chinese Linguistic Data Consortium (2003). CLDC\_wfpm indicates word frequency per million words, CLDC\_logfpm is the results of a Log transformation of wfpm + 1, and CLDC\_zipf is the Laplace transformation of wfpm.
6. The column named ‘SUBTLEX\_freq’ shows the frequency of each word obtained from SUBTLEX-CH. This frequency is the log<sub>10</sub> of word count based on total of 33.5 million words.
7. The column named ‘AoA’ shows the value of age (in years) of acquisition of each word. The AoA measure indicates the age (in years) at which participants had learned a word.
8. The column named ‘PhonologicalN’ shows the value of phonological neighborhood density of each word, which is calculated as Coltheart’s N<sup>35</sup> that measures the number of words that differ from the target word by one phoneme or by one tone.
9. The three columns ending with ‘\_Strokes’ show the number of strokes of each character of each word.
10. The three columns ending with ‘\_Pinyin’ show the pronunciation of each character of each word.
11. The three columns ending with ‘\_PN’ show the phonological neighborhood density of each character in each word. The character-level phonological neighborhood density measures the number of characters that differ from the target syllable by one phoneme or by its tone.
12. The three columns ending with ‘Characterfrequency’ show the frequency per million of each character in each word, which is calculated from the Chinese Linguistic Data Consortium (2003).
13. The three columns ending with ‘Syllablefrequency’ show the frequency per million of each syllable, which is calculated from the Chinese Linguistic Data Consortium (2003).
14. The three columns ending with ‘\_AoA’ show the age of acquisition of each character of each word, which is obtained from a database of objective ages of acquisition of simplified Chinese characters<sup>29</sup>.

## Technical Validation

**Descriptive validation.** The raw data available in the OSF repository<sup>19</sup> are provided in untrimmed format. For the analyses below, we deleted response latencies from error trials, latencies smaller than 200 or larger than 2,000 ms, and response latencies above 3 standard deviations from a participant’s mean. Figure 1 visualizes the distribution of word frequency (CLDC\_zipf) of words in the current dataset (Fig. 1a), as well as of all words in the CLDC dataset (Fig. 1b). It can be observed that the word frequency values in this dataset cover a wide range, rather than being concentrated at one end of the frequency distribution. The range and pattern of the word frequency distribution is similar to the overall word frequency distribution of CLDC. Also, the midpoint of the scale distinguishes between low-frequency words and high-frequency words, aligning with the typical recommendation<sup>36</sup>.



**Fig. 1** Density plot of word frequency. **(a)** word frequency distribution of the current dataset. **(b)** word frequency distribution of the CLDC dataset.

Figure 2 visualizes the distribution of mean response times of participants for each experiment separately. Overall, mean response time ranges from ~0.45 s to ~1.5 s. It can be observed that there is considerable heterogeneity in the distribution of response times across experiments. Experiments involving multiple repetitions of small set of items have faster and less variable reaction time distributions, whereas experiments involving more items and fewer or no repetitions have longer and more dispersed response time distributions. Experiments using the same paradigm show similar ranges of response time distributions.

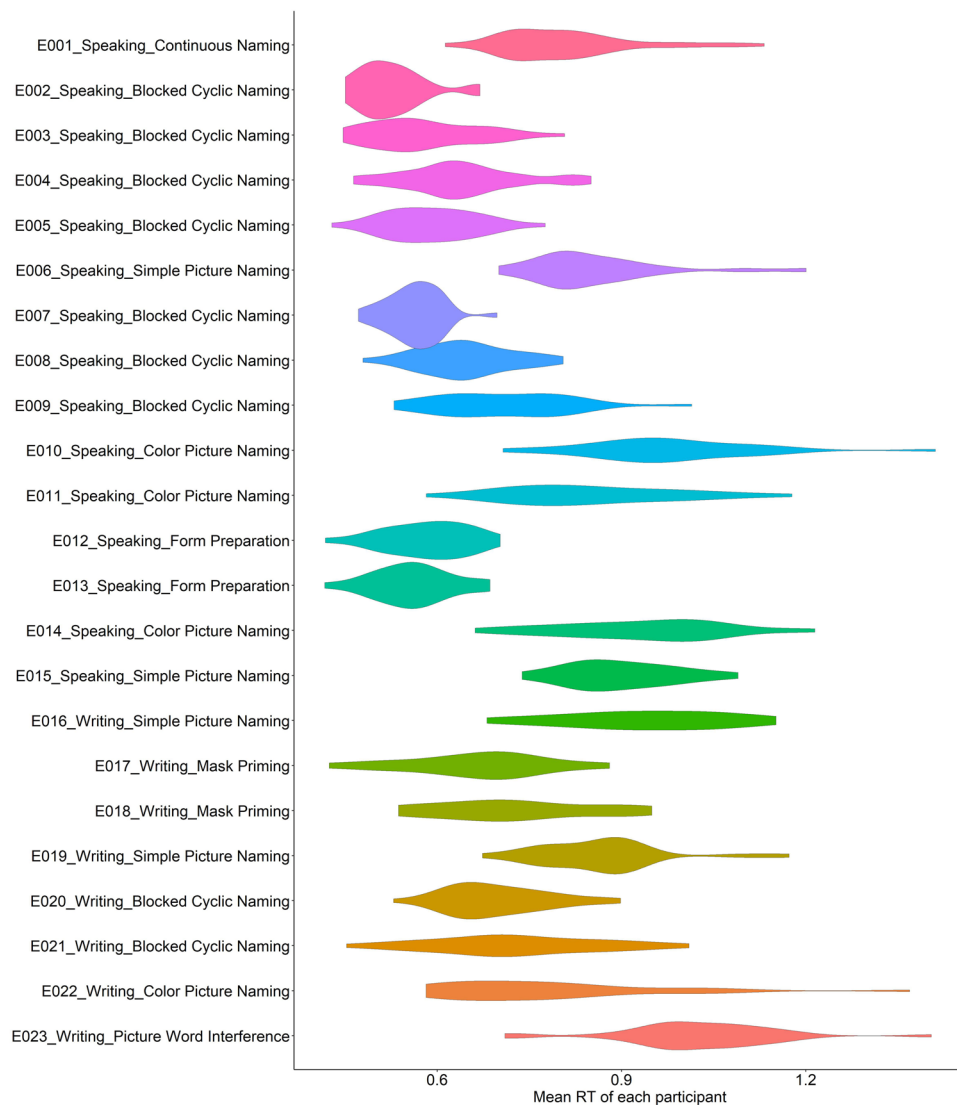
**Quantitative validation.** We quantitatively validated our dataset via a robust word frequency effect on naming latencies which has been observed in many previous studies<sup>14,37,38</sup>. To address the issue of non-negative and right-skewed distribution of reaction latencies, we performed generalized linear mixed model (GLMM) analyses with word frequency as an independent continuous variable, and response latencies as the dependent variable. GLMM allows fitting data without the normality assumptions and the need for transformation. It allows differences between individuals to be properly assessed<sup>39</sup>. To provide robust estimation, we included by-subject and by-experiment random slopes of frequency in the random structure of each model. The analysis was performed separately on the speaking modality and the writing modality. We first modeled word frequency effect on the whole dataset from each modality. Additional analyses were performed on response latencies from only unrelated trials. Analysis was performed using the *lme4* package<sup>40</sup> in the R environment<sup>41</sup>. To ensure that the estimated fixed effects could be accurately interpreted as population-level means, we checked whether or not the mean random effects were close to zero. For these and the following analysis, the unit of response latencies was changed from milliseconds to seconds, which prevented a number of non-convergence or singularity warnings which arose from analyses based on milliseconds.

Table 1 shows the results of the generalized linear mixed model analysis on the speaking and the writing datasets. No word frequency effect was found for the speaking modality. Only the word frequency effect of the writing modality reached significance. However, when analyzing only unrelated trials, a significant word frequency effect was obtained in both speaking and writing modality (see Table 2), suggesting that words with higher frequency were associated with shorter response latencies than words with lower frequency. The results replicate previous findings of word frequency effects in production<sup>14,37,38</sup>.

Word frequency effects were further validated using word frequency statistics from the SUBTLEX-CH. As shown in Table 3, word frequency effects were replicated, and the magnitude of the effects was comparable to that reported in Table 1. Additional analyses on response latencies of unrelated trials revealed word frequency effects, and thus also validated the current dataset (Table 4).

Word frequency effects were additionally validated at the level of single experiments. To achieve this, we extracted experimental level estimations from the fitting model. Thereafter, we computed the 95% confidence interval used the *sim()* function in the *arm* package<sup>42</sup>. The results validated word frequency effects at the level of each single experiment. Detailed results are listed in Table S1 of the Supplementary Information file.

To provide further validation of the current dataset, we evaluated AoA effects. To this end, a generalized linear mixed model was built, including word frequency, AoA and the interaction between AoA and modality as fixed factors, and random effects of AoA on the experiment level. As shown in Table 5, the result revealed only a significant word frequency effect. There was no significant AoA effect, and only a marginally significant interaction between AoA and Modality.



**Fig. 2** Density plot of participants' mean response time (in second) of each experiment.

	Predictors	RT				
		Estimates	SE	<i>t</i>	CI	<i>p</i>
Speaking	(Intercept)	0.772	0.060	12.74	0.64–0.89	<0.001
	frequency	−0.008	0.008	−1.01	−0.02 – −0.01	0.314
Writing	(Intercept)	0.927	0.055	16.73	0.82–1.04	<0.001
	frequency	−0.023	0.006	−3.99	−0.03 – −0.01	<0.001

**Table 1.** Generalized linear mixed model results of the whole dataset for speaking and writing. Word frequency was obtained from Chinese Linguistic Data Consortium (2003). Units for response latencies are seconds (this also applies to Tables 2–5 and Table S1 in the Supplementary Information file). *Note.* Model =  $glmer(RT \sim frequency\_cldc + (1 + frequency\_cldc | Subject) + (1 + frequency\_cldc | Experiment))$ .

### Usage Notes

The current dataset is available at OSF repository<sup>19</sup> under the CC BY 4.0 License. The dataset can be used to apply further analysis to investigate hypotheses and models of speech production. First, as it contains behavioral response measurements of a large set of words presented in various modalities and paradigms, the current dataset makes it possible to compare responses among different modalities and tasks. Second, the current dataset could be further analyzed by incorporating a wider range of linguistic variables of interest, including but not limited to those outlined in the ‘Linguistic measurements’ document (detailed information about the ‘Linguistic measurements’ file is provided in the metadata file at the OSF repository<sup>19</sup>). Moreover, it can be used to explore

	Predictors	RT				
		Estimates	SE	<i>t</i>	CI	<i>p</i>
Speaking	(Intercept)	0.805	0.069	11.70	0.67–0.94	<0.001
	frequency	−0.018	0.008	−2.29	−0.03 – −0.00	0.022
Writing	(Intercept)	0.938	0.040	23.10	−0.86–1.02	<0.001
	frequency	−0.018	0.002	−9.70	−0.02 – −0.01	<0.001

**Table 2.** Generalized linear mixed model results of unrelated trials for speaking and writing. Word frequency was obtained from Chinese Linguistic Data Consortium (2003). *Note.* Model = glmer(RT ~ frequency\_cldc + (1 + frequency\_cldc | Subject) + (1 + frequency\_cldc | Experiment)).

	Predictors	RT				
		Estimates	SE	<i>t</i>	CI	<i>p</i>
Speaking	(Intercept)	0.785	0.058	13.55	0.67–0.90	<0.001
	frequency	−0.018	0.008	−2.14	−0.03 – −0.00	0.033
Writing	(Intercept)	0.899	0.050	18.08	0.08–1.00	<0.001
	frequency	−0.024	0.005	−4.68	−0.03 – −0.01	<0.001

**Table 3.** Generalized linear mixed model results of the whole dataset for speaking and writing. Word frequency was obtained from SUBTLEX-CH. *Note.* Model = glmer(RT ~ frequency\_subtlex + (1 + frequency\_subtlex | Subject) + (1 + frequency\_subtlex | Experiment)).

	Predictors	RT				
		Estimates	SE	<i>t</i>	CI	<i>p</i>
Speaking	(Intercept)	0.808	0.068	11.91	0.68–0.94	<0.001
	frequency	−0.029	0.010	−2.89	−0.05 – −0.01	0.004
Writing	(Intercept)	0.933	0.047	19.91	0.84–1.02	<0.001
	frequency	−0.025	0.007	−3.38	−0.04 – −0.01	<0.001

**Table 4.** Generalized linear mixed model results of unrelated trials for speaking and writing. Word frequency was obtained from SUBTLEX-CH. *Note.* Model = glmer(RT ~ frequency\_subtlex + (1 + frequency\_subtlex | Subject) + (1 + frequency\_subtlex | Experiment)).

Predictors	Estimates	SE	<i>t</i>	CI	<i>p</i>
(Intercept)	1.385	0.074	18.74	1.24–1.53	<0.001
frequency	0.025	0.002	13.00	−0.02 – −0.03	<0.001
AoA	−0.0001	0.003	−0.03	−0.017 – −0.006	0.979
Modality	−0.111	0.124	−0.90	−0.354–0.131	0.368
AoA:Modality	−0.010	0.006	−1.76	−0.021– −0.001	0.079

**Table 5.** Generalized linear mixed model results for word-frequency and AoA effects. Word frequency was obtained from Chinese Linguistic Data Consortium (2003). Word-level AoA was obtained from the large-scale collection of age of acquisition ratings of simplified Chinese words<sup>32</sup>. *Note.* Model = glmer(RT ~ frequency + AoA x Modality + (1 + AoA | Experiment), family = Gamma(link = 'inverse')).

the issue of language-general vs language-specific mechanisms of language production when compared with results from other languages. Finally, we believe that this dataset can provide insights for computational models of language word production.

### Code availability

The codes for measurements, descriptive statistics and quantitative validation are available in an OSF repository<sup>19</sup>

Received: 5 May 2023; Accepted: 30 January 2024;

Published online: 10 February 2024

### References

1. Levelt, W. J., Roelofs, A. & Meyer, A. S. A theory of lexical access in speech production. *Behav Brain Sci.* **22**, 1–38 (1999).
2. Dell, G. S. A spreading-activation theory of retrieval in sentence production. *Psychol Rev.* **93**, 283–321 (1986).
3. Rapp, B. & Goldrick, M. Discreteness and interactivity in spoken word production. *Psychol Rev.* **107**, 460–499 (2000).
4. Caramazza, A. How many levels of processing are there in lexical access? *Cogn Neuropsychol.* **14**, 177–208 (1997).

5. Dell, G. S. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *J Mem Lang*. **27**, 124–142 (1988).
6. Chen, J.-Y. & Chen, T.-M. Word form encoding in mandarin Chinese typewritten word production: evidence from the implicit priming task. *Acta Psychol (Amst)*. **142**, 148–153 (2013).
7. O’Seaghdha, P. G., Chen, J.-Y. & Chen, T.-M. Proximate units in word production: phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*. **115**, 282–302 (2010).
8. Qu, Q., Damian, M. F. & Kazanina, N. Sound-sized segments are significant for Mandarin speakers. *Proc Natl Acad Sci USA* **109**, 14265–14270 (2012).
9. Feng, C. & Qu, Q. Phonological inhibition in written production. *Psychol Res*. **85**, 2271–2278 (2021).
10. Feng, C., Damian, M. F. & Qu, Q. Parallel processing of semantics and phonology in spoken production: Evidence from Blocked Cyclic Picture Naming and EEG. *J Cogn Neurosci*. **33**, 725–738 (2021).
11. Feng, C., Damian, M. F. & Qu, Q. A joint investigation of facilitation and interference effects of semantic and phonological similarity in a continuous naming task. *J Exp Psychol Learn Mem Cogn*. **48**, 1193–1201 (2022).
12. Qu, Q., Damian, M. F., Zhang, Q. & Zhu, X. Phonology contributes to writing: Evidence from written word production in a nonalphabetic script. *Psychol Sci*. **22**, 1107–1112 (2011).
13. Qu, Q., Damian, M. F. & Li, X. Phonology contributes to writing: evidence from a masked priming task. *Lang Cogn Neurosci*. **31**, 251–264 (2016).
14. Qu, Q., Zhang, Q. & Damian, M. F. Tracking the time course of lexical access in orthographic production: An event-related potential study of word frequency effects in written picture naming. *Brain Lang*. **159**, 118–126 (2016).
15. Qu, Q. & Damian, M. F. Orthographic effects in Mandarin spoken language production. *Mem Cognit*. **47**, 326–334 (2019).
16. Qu, Q. & Damian, M. F. An electrophysiological analysis of the time course of phonological and orthographic encoding in written word production. *Lang Cogn Neurosci*. **35**, 360–373 (2020).
17. Qu, Q., Feng, C., Hou, F. & Damian, M. F. Syllables and phonemes as planning units in Mandarin Chinese spoken word production: Evidence from ERPs. *Neuropsychologia*. **146**, 107559 (2020).
18. Qu, Q., Feng, C. & Damian, M. F. Interference effects of phonological similarity in word production arise from competitive incremental learning. *Cognition*. **212**, 104738 (2021).
19. Feng, C., Damian, M. F. & Qu, Q. A dataset of behavioral measures on Chinese word production in picture naming. *Open Science Framework*, <https://doi.org/10.17605/OSF.IO/6GTZH> (2023).
20. Cycowicz, Y. M., Friedman, D., Rothstein, M. & Snodgrass, J. G. Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *J Exp Child Psychol*. **65**(2), 171–237 (1997).
21. Bonin, P., Peereman, R., Malardier, N., Méot, A. & Chalard, M. A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behav Res Meth Instrum Comput*. **35**, 158–167 (2003).
22. Duñabeitia, J. A. et al. MultiPic: A standardized set of 750 drawings with norms for six European languages. *Q J Exp Psychol*. **71**(4), 808–816 (2018).
23. Liu, Y., Hao, M., Li, P. & Shu, H. Timed Picture Naming Norms for Mandarin Chinese. *PLoS ONE*. **6**(1), e16505 (2011).
24. Severens, E., Lommel, S. V., Ratinckx, E. & Hartsuiker, R. J. Timed picture naming norms for 590 pictures in Dutch. *Acta Psychol*. **119**(2), 159–187 (2005).
25. Snodgrass, J. G. & Vanderwart, M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp Psychol Hum Learn*. **6**, 174–215 (1980).
26. Zhang, Q. & Yang, Y. The determiners of picture-naming latency. *Acta Psychol Sin*. **35**, 447 (2003).
27. Forster, K. I. & Forster, J. C. DMDX: a windows display program with millisecond accuracy. *Behav Res Meth Instrum Comput*. **35**, 116–124 (2003).
28. Protopapas, A. CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behav Res Methods*. **39**, 859–862 (2007).
29. Roux, F., Armstrong, B. C. & Carreiras, M. Chronset: An automated tool for detecting speech onset. *Behav Res Methods*. **49**, 1864–1881 (2017).
30. Chinese Linguistic Data Consortium. 现代汉语通用词表 [Chinese lexicon] (CLDC-LAC-2003-001). (2003).
31. Cai, Q. & Brysbaert, M. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One*. **5**, e10729 (2010).
32. Xu, X., Li, J. & Guo, S. Age of acquisition ratings for 19,716 simplified Chinese words. *Behav Res Methods*. **53**, 558–573 (2021).
33. Cai, Z. G., Huang, S., Xu, Z. & Zhao, N. Objective ages of acquisition for 3300+ simplified Chinese characters. *Behav Res Methods*. **54**, 311–323 (2022).
34. Sun, C. C., Hendrix, P., Ma, J. & Baayen, R. H. Chinese lexical database (CLD): A large-scale lexical database for simplified Mandarin Chinese. *Behav Res Methods*. **50**, 2606–2629 (2018).
35. Coltheart, M., Davelaar, E., Jonasson, J. T. & Besner, D. in *Attention and Performance VI* (ed. Stanislav Dornić) Ch. 2. (Routledge, 1977).
36. van Heuven, W. J. B., Mandera, P., Keuleers, E. & Brysbaert, M. SUBTLEX-UK: a new and improved word frequency database for British English. *Q J Exp Psychol*. **67**, 1176–1190 (2014).
37. Jescheniak, J. D. & Levelt, W. J. M. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *J Exp Psychol Learn Mem Cogn*. **20**, 824–843 (1994).
38. Zhang, Q. & Wang, C. Syllable frequency and word frequency effects in spoken and written word production in a non-alphabetic script. *Front Psychol*. **5**, 120 (2014).
39. Lo, S. & Andrews, S. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front Psychol*. **6**, (2015).
40. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J Stat Soft*. **67** (2015).
41. R Core Team. R: A language and environment for statistical computing. (2014).
42. Gelman A, Su Y. arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. <https://CRAN.R-project.org/package=arm> (2022).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 32171058, No. 31771212 and No. 62061136001), Youth Innovation Promotion Association (Chinese Academy of Sciences), and Youth Elite Scientist Sponsorship Program (No. YESS20200138, China Association for Science and Technology), Beijing Nova Program, the Scientific Foundation of Institute of Psychology (No. E2CX3625CX, Chinese Academy of Sciences) to Qingqing Qu, and the CAS Key Laboratory of Behavioral Science, Institute of Psychology (No. Y5CX052003). We are also grateful to Dr. Luke Maurits for suggestions on data analysis in the Quantitative Validation section.

### Author contributions

Conceiving and writing: C. Feng, M. Damian, QQ. Qu; Data collection and inspection: C. Feng, QQ. Qu; Statistical validation: C. Feng, QQ. Qu; Overall supervision: QQ. Qu, M. Damian.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03022-8>.

**Correspondence** and requests for materials should be addressed to Q.Q.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024