Picture naming in bilingual and monolingual Chinese speakers: Capturing similarity and variability

Mohammad Momenian[1,2,3] · Mehdi Bakhtiar[3] · Yu Kei Chan[1] · Suet Lin Cheung[1] · Brendan Stuart Weekes[1,4]

Abstract

Picture-naming latency differs across languages in bilingual speakers. We compared the effects of key psycholinguistic variables on picture naming among two groups of Chinese bilingual speakers and Mandarin monolingual speakers. First, we asked bilingual and monolingual speakers to estimate the age of acquisition, familiarity, visual complexity, name agreement, and imageability of a set of object and action pictures in Mandarin and Cantonese. Next, we recruited 60 Cantonese-English speakers, 50 Mandarin-Cantonese bilingual speakers, and 30 monolingual speakers who named the object and action pictures in Cantonese and Mandarin, respectively. We observed variability in the effects of item-level characteristics among groups, suggesting an interaction between item-level and individual-level characteristics as predicted. This variability was higher in bilingual speakers who spoke similar languages (Mandarin-Cantonese) in comparison to those speaking more distant languages (Cantonese-English). Our results suggest that monolingual norms and bilingual norms capture the same amount of variability; however, grammatical class interactions with other variables are explained differentially by the bilingual and monolingual norms. We discuss the implications of our findings in terms of norming studies for timed picture naming and effects of bilingualism on language processing.

Keywords Action naming · Bilingual · Cantonese · LMEM · Mandarin · Monolingual · Object naming

Introduction

Timed picture naming is an established method for studying lexical retrieval and word production in both healthy and impaired speakers. Several different models of timed picture naming have been proposed. All agree that picture naming consists of three stages: (1) visual recognition and conceptual identification, where features of the depicted object or event are extracted and matched with semantic knowledge; (2) lexical selection, in which suitable lexical representation of words along with their syntactic properties (lemmas) are accessed and selected; and (3) articulation of the word (Bonin, Meot, Lagarrigue, & Roux, 2015; Glaser, 1992; Johnson, Paivio, & Clark, 1996; Levelt, 1999; Levelt, Roelofs, & Meyer, 1999; Perret & Bonin, 2018; Rapp & Goldrick, 2000). These stages are assumed to be universal across both languages and speakers.

Multiple studies in many languages confirm a number of key variables that predict timed picture naming. In a recent Bayesian meta-analysis, Perret and Bonin (2018) confirmed the effects of these key variables which include rated age of acquisition (AoA), familiarity, imageability, and name agreement (NA) in monolingual speakers. These variables are assumed to exert their influence on at least one stage of timed picture naming (Alario et al., 2004; Humphreys, Riddoch, & Quinlan, 1988), but may also have multiple loci. For example, NA, imageability, and visual complexity (VC) are assumed to arise during the visual recognition of objects and actions (Alario et al., 2004; Humphreys et al., 1988), but NA is also assumed to have an impact on spoken word retrieval. By contrast, rated familiarity and imageability are assumed to influence semantic and conceptual identification only (Barry, Morrison, & Ellis, 1997; Ellis & Morrison, 1998; Weekes, Shu, Hao, Liu, & Tan, 2007), and word frequency is assumed

✉ Mohammad Momenian
mohammad.momenian@polyu.edu.hk

[1] Laboratory for Communication Science, Faculty of Education, University of Hong Kong, Hong Kong, Hong Kong

[2] CF705, Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hong Kong, Hong Kong

[3] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, Hong Kong

[4] School of Psychological Sciences, University of Melbourne, Melbourne, Australia

to reflect processes at the lexical selection and encoding stage of speech production (Alario et al., 2004) as does AoA, although AoA may also have an impact on semantic and conceptual identification. Other studied variables have less robust effects on timed picture naming, including word length, morphological complexity, argument structure, and verb instrumentality (Barbieri, Basso, Frustaci, & Luzzatti, 2010; Crepaldi, Che, Su, & Luzzatti, 2012; Cuetos, Ellis, & Alvarez, 1999; Jonkers & Bastiaanse, 2007; Kambanaros, 2009; Parris & Weekes, 2001; Thompson, 2003). A majority of studies use objects in timed picture naming studies. However, studies of action naming reveal similar effects across languages in both typical and atypical monolingual speakers (Alyahya & Druks, 2016; Bird, Franklin, & Howard, 2001; Druks et al., 2006; Edmonds & Donovan, 2012; Khwaileh, Mustafawi, Herbert, & Howard, 2018; Morrison, Hirsh, & Duggan, 2003; Nilipour, Bakhtiar, Momenian, & Weekes, 2017; Szekely et al., 2005). These studies suggest that the item-level characteristics of actions and objects constrain typical and atypical picture naming across languages for monolingual speakers. Recently, however, the study of bilingual language processing has come into focus.

One question is whether the established effects of psycholinguistic variables on timed picture naming in monolingual speakers are observed for bilingual speakers (Lisa A. Edmonds & Donovan, 2014; Ramanujan & Weekes, 2019). A related question is whether the item-level characteristics taken from the ratings of monolingual speakers are valid for testing bilingual speakers. Bilingual speakers are very different in terms of several linguistic and non-linguistic factors. Moreover, unlike monolingual speakers, bilingual speakers are highly diverse in their language proficiency, age of exposure to each language, amount of exposure on a daily basis, and relative linguistic distance between the languages spoken (Abutalebi et al., 2013; Chee, Soon, Lee, & Pallier, 2004; Degani, Prior, & Hajajra, 2018; Golestani et al., 2006; Kuzmina, Goral, Norvik, & Weekes, 2019; Lutz, Lee, & Weekes, 2018; Momenian, Nilipour, Samar, Cappa, & Golestani, 2018; Newman, Tremblay, Nichols, Neville, & Ullman, 2012; Ramanujan, 2019; Sörman, Hansson, & Ljungberg, 2019; Suh et al., 2007; Yan, Zhang, Xu, Chen, & Wang, 2016). Overall, it can be assumed that bilingual speakers are a more heterogeneous group compared to monolinguals; i.e. they have a wider range of participant level variability in language experience and language processing. This offers a unique way to test models of timed picture naming given that participant-level variability in monolingual speakers is typically highly controlled under the assumption that speakers are not variable in processing.

In this study, we test whether psycholinguistic variables which are assumed to impact picture naming in monolingual speakers are different in bilingual speakers. We first adapted

Druks and Masterson's object and action naming battery (OANB) (Druks & Masterson, 2000) for bilingual speakers. To make a direct comparison between bilingual and monolingual speakers, we then asked Cantonese-English (C-E) and Mandarin-Cantonese (M-C) bilingual speakers to name pictures in 'monolingual mode', a term borrowed from Grosjean (2001), referring to sustained maintenance of the first language while the second language is inhibited, and compared the data to monolingual Chinese speakers as a baseline. We measured the effects of individual differences on timed picture naming in both groups such as age, education level, and amount of exposure to spoken and written language in first (L1) and second (L2) languages for bilinguals.

Two contrasting hypotheses will be tested. If the effects of item-level and participant-level variability are similar in bilingual and monolingual speakers, then the variables which have an effect on picture naming in monolinguals should impact naming of bilingual speakers. Under this hypothesis, both monolingual and bilingual speakers should respond similarly to the key psycholinguistic variables. The alternative hypothesis is that key factors will have different effects in bilingual and monolingual speakers. We predict that we will see more variability in the bilingual speakers' responses based on the argument that bilingual speakers' individual differences and diverse language learning experiences could modulate the effects of the item-level variables (Edmonds & Donovan, 2012). Based on this hypothesis, we expect to see differences between monolingual and bilingual speakers in the effects of psycholinguistic variables such as AoA and frequency which are modulated more by bilingual experiences, whereas the effects of variables that are related to the conceptual level of processing such as imageability might be more consistent across both populations of monolingual and bilingual speakers. This hypothesis is based on the weaker links hypothesis (WLH; Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Gollan, Montoya, Cera, & Sandoval, 2008; Kroll & Gollan, 2013). The WLH predicts different effects for frequency due to weaker links between semantics and phonology in the bilingual lexicon in comparison with the monolingual lexicon.

Methods

Preparatory study

Forty early C-E and 24 early M-C bilingual speakers were presented with 162 pictures of objects and 100 pictures of actions from OANB (Druks & Masterson, 2000). They were asked to write down the names of all pictures via an online self-paced Google docs form. If they did not know the name of the picture, they could select an option which read 'not familiar/do not recognize the item'. Name agreement (NA)

was derived for each picture by selecting the most dominant Mandarin or Cantonese name (the name with the highest token frequency count among competitors). The cut-off score for objects was 40% and for actions was 30%. Items below the cut-off were removed. Linguistically unfamiliar and culturally inappropriate items were also removed, e.g. baseball bat. For items that had the same modal name in Cantonese or Mandarin (i.e. 服務員 = waiter/waitress), one was excluded from the study. Hence, the number of items for rating was 144 objects and 86 actions in Cantonese, and 145 objects and 84 actions in Mandarin.

A total of 54 early C-E bilingual speakers and 56 early M-C bilingual speakers participated in the rating studies. Rating data was collected in two sessions either individually or in small groups with a 1-week interval between the sessions. In the first session, ratings were produced for NA and VC. In the second session, ratings were produced for AoA, imageability, and familiarity of the names with the highest agreement produced by the whole group in the previous week. The order of rating for each variable was randomized across participants in each session in addition to item randomization across participants. Recruiting age- and education-matched monolingual speakers of Mandarin and Cantonese in Hong Kong is impossible because everyone speaks at least two languages. We therefore collected ratings from 37 monolingual Mandarin speakers for 145 objects and 84 actions in Mainland China. The monolingual and bilingual participants were all age- and education-matched.

We followed Druks and Masterson's (2000) rating instructions for all variables across all groups. For NA, participants were asked to write down the name of the target line drawing in their native language. They were instructed to produce only one name per picture. For rating of VC, participants were asked to rate the VC of each drawing in terms of the existence of visual details and lines of the drawing. A value of 1 indicated the lowest VC of the drawing, and 7 indicated the highest VC. In addition to subjective VC, we established objective VC measures using computational algorithms (Donderi, 2006; Machado et al., 2015; Székely & Bates, 2000). Objective VC measures have shown correlation with subjective VC and are also not confounded with variables such as familiarity and frequency (Bates et al., 2003). We will use the objective measure for our modelling of data in this study.

For the AoA rating, the participants were asked to judge the age at which they thought they had acquired each word, using the scale of 1 = 0–2 years old; 2 = 3–4 years old; 3 = 5–6 years old; 4 = 7–8 years old; 5 = 9–10 years old; 6 = 11–12 years old; 7 = 13 years old or above. We also reviewed AoA of words from Mandarin and Cantonese parent-report norms such as MacArthur-Bates Communicative Development Inventories (CDIs) (Frank, Braginsky, Yurovsky, & Marchman, 2017). However, many of the object and action

names we used in this study were missing in the parent-report norms which made it difficult to look at the correlation between rated AoA and parent-report AoA. For the Imageability rating, participants were asked to rate how readily each word could arouse a mental image. A value of 1 indicated that the word evoked its mental image with the greatest difficulty, and 7 indicated that it evoked the image most readily. For the familiarity rating, the participants were asked to rate the extent to which they came into contact with or thought about the object or action represented by a word. A value of 1 indicated the target was very unfamiliar, and 7 indicated very familiar. Cantonese spoken words were depicted in traditional format in print, whereas Mandarin spoken words were depicted in simplified form due to writing reform of the PRC in the mid-20th century.

We calculated the reliability of ratings for familiarity, subjective VC, imageability, and AoA using intraclass correlation (ICC). ICC is an index to assess the reliability of human ratings (Shrout & Fleiss, 1979). Koo and Li (2016) suggest the following ICC rubric for assessing the reliability of a measurement: below 0.50: poor reliability; between 0.50 and 0.75: moderate reliability; between 0.75 and 0.90: good reliability; and above 0.90: excellent reliability. We calculated the corrected correlation following the procedure suggested by Nicewander (2018) (See Tables 1, 2, and 3).

The values for word frequency in Cantonese were obtained from the Sketch Engine database, computed from a corpus named Cantonese Web Corpus (CantoneseWaC) of over a million tokens. The Sketch Engine database is made up of texts collected from the internet for several languages (Kilgarriff, Reddy, Pomikálek, & Avinesh, 2010). For the values in Mandarin text, the SUBTLEX-CH corpus (Cai & Brysbaert, 2010) was used to establish the value for each item. This corpus is a spoken one consisting of film subtitles.

The ratings derived are available online as psycholinguistic norms for both Cantonese and Mandarin objects and actions in the following link (DOI 10.17605/OSF.IO/8HVWR).

Picture-naming experiments

Participants

Sixty C-E speakers (28 male, average age = 22.4, ranging from 18 to 34 years), 50 M-C bilingual speakers (18 male, average age = 21.8, ranging from 18 to 33 years), and 30 monolingual Mandarin speakers (12 male, average age = 20.6, ranging from 19 to 25) participated in the timed picture-naming task (see Tables 4 and 5). None of them had been tested in the preparatory study. They all had normal or correct-to-normal vision. All participants completed a language background questionnaire and a consent form prior to the experiment. The language background questionnaire included questions about the AoA of each language, reported proficiency for each

Table 1 The corrected correlation matrix for the key variables in the C-E group

	VC	Familiarity	AoA	Imageability	NA	Frequency (log)
VC	0.79*					
Familiarity	−0.20	0.86*				
AoA	0.31	−0.72	0.92*			
Imageability	−0.62	0.75	−0.75	0.82*		
NA	−0.51	0.23	−0.29	0.44	1.00	
Frequency	0.00	0.41	−0.58	0.23	0.05	1.00

*ICC

language, amount of exposure to the written and spoken language, demographic information, and other questions about how they learned each of the languages (i.e. through TV, interaction, reading books, etc.). All of the bilinguals were early bilinguals based on the data from the questionnaire. They all reported they could communicate in both of their languages fluently. All of them had received formal education in their first language for at least 15 years. The monolingual Mandarin speakers had been taught English in school but was basic written English. They all reported they could not communicate in English. They had exposure to written and spoken Mandarin on a daily basis 85% and 87% of the time, respectively. Native language (L1) was defined according to the first acquired and most commonly spoken language on a daily basis based on answers from the questionnaire. Native language is denoted first, and non-native language is denoted second hereafter. All the procedures in this study were approved by the Ethical Committee at the University of Hong Kong (Tables 4 and 5).

Procedure

Timed picture naming was recorded in sound-proof rooms. Two blocks of stimuli (objects and actions) were designed using DMDX (Forster & Forster, 2003). Items were all randomized within each block. The block order was counterbalanced across participants. Responses were captured

by a microphone with calibration completed before each session to ensure background noise was not recorded as a response. The input threshold level for recording was in fact adjusted to match the natural speaking volume of each participant. Participants were familiarized with the experiment format through practice trials before testing commenced. They were instructed to name pictures as quickly and accurately as possible in their first language. They were instructed not to cough, breathe loudly, move their heads, and produce starters or fillers, e.g. 'um', during or before each response. Each trial began with the presentation of a cross or fixation point at the centre of the monitor for 500 ms. After that, each picture was presented in the middle of the screen and remained until a response was detected, with 2000 ms as the time-out period. An error was recorded by DMDX if the participant was unable to respond within the time period. Participants' errors including production of wrong names, nontarget sounds, hesitations, and voice-key failures were all recorded for off-line analysis.

Analysis plan of the latency data

We used linear mixed effects modeling (LMEM) using the lme4 package with R software to analyze the data (see Baayen, Davidson, & Bates, 2008, for a good introduction to LMEM). The practice suggested by Barr, Levy, Scheepers, and Tily (2013) and Bates, Kliegl, Vasishth, and Baayen (2015) was followed to

Table 2 The corrected correlation matrix for the key variables in the M-C group

	VC	Familiarity	AoA	Imageability	NA	Frequency (log)
VC	0.96*					
Familiarity	−0.18	0.95*				
AoA	0.20	−0.79	0.96*			
Imageability	−0.46	0.78	−0.73	0.77*		
NA	−0.11	0.18	−0.33	0.28	1.00	
Frequency	−0.03	0.20	−0.22	0.13	0.08	1.00

*ICC

Table 3 The corrected correlation matrix for the key variables in the monolingual mandarin group

	VC	Familiarity	AoA	Imageability	NA	Frequency (log)
VC	0.95*					
Familiarity	−0.24	0.92*				
AoA	0.25	−0.72	0.95*			
Imageability	−0.43	0.76	−0.64	0.73*		
NA	−0.24	0.24	−0.33	0.37	1.00	
Frequency	0.04	0.20	−0.20	0.18	0.05	1.00

*ICC

fit the models. Our dependent variable was transformed naming latency recorded as reaction time (RT) using common log transformation. The missing and incorrect responses (C-E group: 9.8%; M-C group: 11.2%; monolingual group: 14.2%) were removed from further analysis. Every response which was different from the dominant name was considered incorrect. Before modelling, we standardized the continuous independent variables. To check the collinearity among the variables, we used a variance inflation factor (VIF). Variables with a VIF value above 5 should be removed from the analysis based on the recommendation by Craney and Surles (2002).

As in previous RT studies, we began with a maximal model (Barr et al., 2013), entering all the fixed variables of interest including objective VC, imageability, AoA, log frequency, NA, familiarity, and grammatical class (GC). We added variables such as participants' age, education level (in months), reported L1 spoken and written use, and reported L2 spoken and written use as control fixed variables only for the bilingual analyses. The interaction between GC and other fixed effects (except for age and education) was also included in the model. For the random-effects structure of the model, random intercepts of items and subjects together with by-subject random slopes for all fixed effects of interest including VC, imageability, AoA, log frequency, NA, familiarity, and GC were added to the model.

In order to find the most parsimonious random effects structure for our data, we performed a singular value decomposition (SVD) on the covariance matrix of the maximal model using principal component analysis (PCA) (Bates et al.,

2015). PCA could tell which of the random effects structures were not contributing significantly to the model. This further helped resolve unnecessary complications such as over-specification and convergence problems in the model. PCA was accompanied by likelihood ratio tests (LRT) for both statistical significance and model evaluation. Random effects correlation parameters were not included in the maximal model at first to prevent convergence problems (Bates et al., 2015). However, once the most appropriate random effects structure was determined, correlation parameters were added to the model and compared with the model without correlation parameters using LRT.

Having dealt with the random effects structure of the model, we then turned to the fixed effects part. Determining which variables have a significant effect is controversial in LME models. Many approaches have been proposed, including Wald tests, LRT, Markov-chain Monte Carlo (MCMC), Kenward-Roger, Satterthwaite, and parametric bootstrapping (Baayen et al., 2008; Barr et al., 2013; Luke, 2017; Pinheiro & Bates, 2000). Wald and LRT are less computationally demanding (Luke, 2017). In order to judge which fixed effects were significant, we used conditional F-tests because doing LRT on the fixed effects is anti-conservative and could result in misleading findings (Halekoh & Højsgaard, 2014; Luke, 2017; Pinheiro & Bates, 2000). We used Kenward-Roger approximations to calculate denominator degrees of freedom which have shown more acceptable type 1 error rates in comparison with LRT and Wald tests (Kuznetsova, Brockhoff, & Christensen, 2017; Luke, 2017). All analysis procedures

Table 4 C-E participants' language profile in the timed picture-naming experiment

Highest academic qualifications	No. of years of formal Cantonese study	Self-rated percentage of L1 spoken use	Self-rated percentage of L2 spoken use	Self-rated percentage of L1 reading use	Self-rated percentage of L2 reading use
High school: 20%	<15 years: 0%	<50%: 0%	<50%: 82%	<50%: 10%	<50%: 80%
Bachelor: 65%	15–20 years: 88.3%	50–70%: 35%	50–70%: 13%	50–70%: 48.33%	50–70%: 15%
Masters or above: 15%	>20 years: 11.67%	70–90%: 65%	70–90%: 5%	70–90%: 25%	70–90%: 5%
			>90%: 0%	>90%: 16.67%	>90%: 0%

Table 5 M-C participants' language profile in the timed picture-naming experiment

Highest academic qualifications	No. of years of formal Mandarin study	Self-rated percentage of L1 spoken use	Self-rated percentage of L2 spoken use	Self-rated percentage of L1 reading use	Self-rated percentage of L2 reading use
High school: 1.7%	<15 years: 0%	<30%: 23.3%	<30%: 31.7%	<30%: 18.3%	<30%: 68.3%
Bachelor: 66.7%	15-20 years: 88.3%	30–50%: 33.3%	30–50%: 46.7%	30–50%: 28.3%	30–50%: 23.3%
Masters or above: 31.7%	>20 years: 11.7%	50–70%: 33.3%	50–70%: 13.3%	50–70%: 33.3%	50–70%: 6.7%
		70–90%: 6.7%	70–90%: 6.7%	70–90%: 15%	70–90%: 1.7%
		≥90%: 3.3%	≥90%: 1.7%	≥90%: 5%	≥90%: 0%

described above were applied to the data from all timed picture-naming experiments. The above-mentioned analysis pipeline was applied to both the bilingual and monolingual data. To determine whether the experiments had enough power, we consulted Brysbaert and Stevens (2018). They recommend a properly powered repeated-measures study should have at least 40 participants with 40 stimuli (1600 observations per condition). The number of participants and stimuli in our study exceeded this threshold well enough. The data for both bilingual experiments and R codes are available online at the following link (DOI 10.17605/OSF.IO/8HVWR).

Results

Monolingual group

We first checked for multicollinearity by applying VIF. VIF values were below the cut-off, except for age of participants. We therefore removed age from further models. We standardized all predictor variables before the analysis. The removal of the frequency $x^2(1) = 1.05$, $p = 0.30$, familiarity $x^2(1) = 0.0$, $p = 1$, and AoA $x^2(1) = 0.0$, $p = 1$ from the random effects structure did not have any effects on the model fit. However, when the following random effects were removed, a significant change happened in the model fit: VC $x^2(1) = 4.60$, $p < 0.05$, imageability $x^2(1) = 3.91$, $p < 0.05$, NA $x^2(1) = 16.29$, $p < 0.001$, and GC $x^2(3) = 133.57$, $p < 0.001$. The removal of zero correlation parameters did not change the model fit $x^2(8) = 11.84$, $p = 0.15$. The only significant fixed effects were GC, imageability, and NA. The AoA effect was marginally significant (See Table 6 for a summary of the model).

Mandarin-Cantonese group

We did two separate analyses for this group. In the first analysis, we used normed ratings for psycholinguistic variables from the bilingual M-C population to predict reaction time. In the second analysis, we used ratings from a monolingual Mandarin population. VIF results suggested age should be

removed for both analyses for its high collinearity with education level. As in the previous analyses, we used PCA accompanied by LRT. In the former analysis, by-subject random slopes of frequency $x^2(1) = 1.96$, $p = 0.16$ and imageability $x^2(1) = 2.02$, $p = 0.15$ were not significant. PCA showed that frequency and imageability accounted for less than 1% of the whole variance in the random effects structure. However, the exclusion of familiarity $x^2(1) = 14.17$, $p < 0.001$, AoA $x^2(1) = 14.33$, $p < 0.001$, NA $x^2(1) = 10.97$, $p < 0.001$, VC $x^2(1) = 10.65$, $p < 0.01$ and GC $x^2(3) = 115.92$, $p < 0.001$ changed the model fit significantly. This means that M-C bilingual speakers showed different degrees of effect when it comes to AoA, GC, imageability, NA, familiarity, and VC. We then removed the zero correlation parameters and performed another LRT. The results showed that removing the correlation parameter had no significant effect $x^2(13) = 17.70$, $p = 0.16$. Therefore, we decided to remove correlation parameters in the random effects model (see Table 7 for a summary of the LMEM). Finally, we tested the fixed effects and their interactions in the model. The results are summarized in Table 8.

Table 6 A summary of the significant effects in the monolingual group

Fixed effects	t value	Std. error	p value	95% CI
Intercept	363.66	0.008	0.001	3.03, 3.06
NA	−7.47	0.002	0.001	−0.02, −0.01
Imageability	−4.18	0.004	0.001	−0.02, −0.008
GC	−2.22	0.007	0.05	−0.03, 0.002
AoA	1.89	0.004	0.059	−0.0002, 0.01
Random effects	Variance		SD	
Item (intercept)	0.001		0.03	
Subject (intercept)	0.001		0.03	
GC-action	0.0007		0.02	
GC-object	0.001		0.03	
NA	0.00003		0.005	
Imageability	0.00001		0.004	
VC	0.00001		0.003	
Residual	0.004		0.06	

Table 7 A summary of the significant effects in the M-C bilingual group with the bilingual ratings

Fixed effects	t value	Std. error	p value	95% CI
Intercept	374.06	0.008	0.001	2.98, 3.01
Imageability	−4.11	0.004	0.001	−0.02, −0.003
NA	−5.82	0.002	0.001	−0.02, −0.01
GC	−2.82	0.006	0.01	−0.03, −0.007
GC: AoA	2.38	0.01	0.05	0.004, 0.04
GC: Imageability	2.12	0.008	0.01	0.008, 0.04
GC: VC	2.77	0.007	0.01	-−0.003, 0.03
Random effects	Variance		SD	
Item (intercept)	0.001		0.03	
Subject (intercept)	0.001		0.04	
GC-action	0.001		0.03	
GC-object	0.001		0.03	
NA	0.00002		0.005	
AoA	0.00005		0.007	
VC	0.00002		0.005	
Familiarity	0.00005		0.007	
Residual	0.006		0.07	

Figure 1 shows the GC interactions with AoA, VC, and imageability.

In the second analysis, we wanted to see how ratings of psycholinguistic variables collected from monolingual Mandarin speakers would predict picture naming in bilingual M-C speakers. The random effects of frequency $x^2(1) = 1.79$, $p = 0.18$ and imageability $x^2(1) = 0.11$, $p = 0.73$ were not significant. However, the random effects of familiarity $x^2(1) =$

Table 8 A summary of the significant effects in the M-C bilingual group with the monolingual ratings

Fixed effects	t value	Std. error	p value	95% CI
Intercept	372.48	0.008	0.001	2.98, 3.01
AoA	2.66	0.004	0.01	0.003, 0.02
Imageability	−4.46	0.003	0.001	−0.02, −0.007
NA	−7.34	0.002	0.001	−0.02, −0.01
GC	−3.12	0.006	0.01	−0.03, −0.004
Random effects	Variance		SD	
Item (intercept)	0.001		0.03	
Subject (intercept)	0.001		0.04	
GC-action	0.001		0.03	
GC-object	0.001		0.03	
NA	0.00005		0.007	
AoA	0.00004		0.006	
VC	0.00001		0.004	
Familiarity	0.00002		0.005	
Residual	0.006		0.07	

6.22, $p < 0.05$, AoA $x^2(1) = 12.96$, $p < 0.001$, NA $x^2(1) = 31.30$, $p < 0.001$, VC $x^2(1) = 10.38$, $p < 0.01$ and GC $x^2(3) = 124.56$, $p < 0.001$ were all significant. The addition of the correlation parameters to the random effects structure had no significant effects on the model fit $x^2(13) = 17.62$, $p = 0.17$. The significant fixed effects, plus two marginally significant interactions, are summarized in Table 8.

Cantonese-English group

We followed exactly the analysis pipeline outlined above. VIF results suggested that L2 spoken and written use should be removed from the analysis for high collinearity with L1 spoken and written use. Based on PCA results and the variance-covariance matrix, we started removing random effects with the lowest variance followed by LRTs. The removal of familiarity $x^2(1) = 0.17$, $p = 0.67$, NA $x^2(1) = 3.42$, $p = 0.06$, frequency $x^2(1) = 1.12$, $p = 0.28$, and VC $x^2(1) = 0.001$, $p = 0.97$ did not have any significant effects on the model since they accounted for less than 3% of the variance in the random effects structure. The exclusion of AoA $x^2(1) = 13.76$, $p < 0.001$, imageability $x^2(1) = 9.32$, $p < 0.01$, and GC $x^2(3) = 108.24$, $p < 0.001$, however, changed the model fit significantly. In other words, the effects of AoA, imageability, and GC were not exactly the same across all participants. Finally, we removed all of the zero correlation parameters and performed an LRT. The results showed that a model with correlation parameters included was not better $x^2(4) = 0.0$, $p = 1$ (see Table 6 for a summary of the LMEM).

Regarding the fixed effects, only familiarity, imageability, and NA significantly predicted RT in C-E bilingual speakers (See Table 9 for further information).

Discussion and conclusion

In this study, we first established normative data for object and action pictures for both monolingual and bilingual Chinese speakers. We then tested influences of key psycholinguistic variables on timed picture naming across bilingual and monolingual speakers. NA and imageability, as we had predicted, were the only variables to show robust effects across all groups, suggesting that these variables were independent of individual differences and language learning experiences. However, substantial variability was observed in the bilingual groups in terms of interactions with GC. The differences in the results among the groups clearly suggest that bilingual experiences and language-specific properties can modulate the effects.

NA is often found to influence lexical access across many studies (Alario & Ferrand, 1999; Bates et al., 2003; Szekely et al., 2005). Research in several languages shows that the number of alternative names associated with an object or
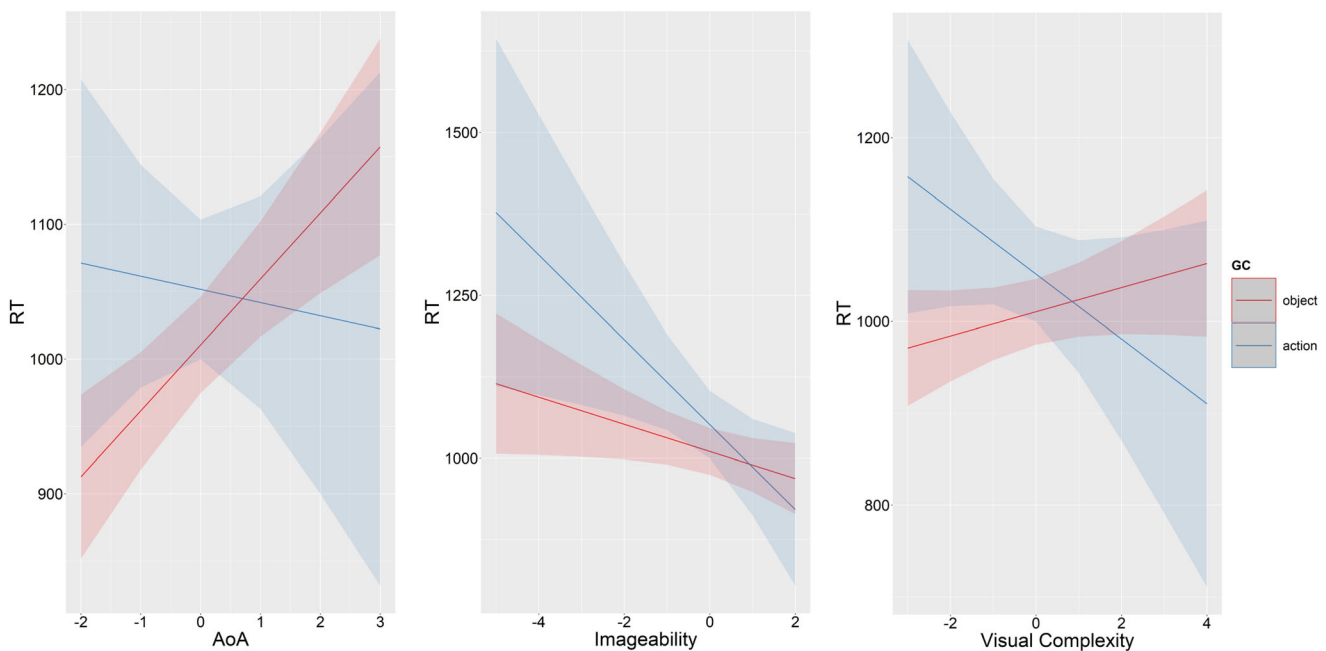
Fig. 1 GC interactions with AoA, imageability, and VC in the M-C group with a 0.95 confidence interval (CI). RT is on original scale for display purposes in this figure

action has an effect on how quickly it is named in monolinguals (Alario et al., 2004; E. Bates et al., 2003); the less alternative names an object or action has, the faster it should be named. According to the codability hypothesis (Ramanujan & Weekes, 2019), the same effect of NA is expected for bilingual speakers, but speaking two languages may give rise to cross-linguistic interference and thus within language effects. This is because in bilingual speakers, every concept has at least two language name associations in the lexicon, leading to higher amounts of competition and, more critically, interference compared with monolingual speakers according to models of bilingual language processing (Abutalebi, & Green, 2008; Green, 1998; though see Gollan, Montoya,

Cera, & Sandoval, 2008, for a different explanation but the same effect).

We observed a continuum in the strength of NA effects. NA had the strongest effects for the monolingual speakers, followed by the C-E bilingual speakers, and finally the M-C bilingual speakers. We suggest the largest effect was observed in monolingual speakers because there is no between-language competition. However, bilingual speakers have more alternative names generated from the conceptual system (at least two names for each concept), and this will create interference at the level of speech production even in monolingual mode (see (Abutalebi, & Green, 2008; Green, 1998). The inhibitory role of higher linguistic similarity was also evident because we only observed a significant by-participant variability in codability in the M-C group. One may contend that when the languages spoken are more similar, e.g. Cantonese and Mandarin, there is a trend toward an interaction with by-participant variability. By-participant variability for codability was also reported in Ramanujan and Weekes (2019) where bilinguals spoke unrelated languages (Hindi and English). Taken together, we contend that this effect cannot be attributed to linguistic similarity only.

We found that NA was a strong predictor of naming latency in all groups of bilingual and monolingual speakers. However, a more important finding was that this effect was not uniform across all bilingual speakers. This could mean that codability is not only a property of the item, but it is also a product of individual differences, i.e. within-participant variability. Modelling this variability is only possible using sophisticated statistical techniques such as LMEM and ideally

Table 9 A summary of the significant effects in the C-E bilingual group

Fixed effects	*t* value	Std. error	*p* value	95% CI
Intercept	515.135	0.005	0.001	3.01, 3.04
Familiarity	2.78	0.004	0.01	0.003, 0.02
Imageability	−7.19	0.004	0.001	−0.04, −0.02
NA	−6.45	0.003	0.001	−0.02, −0.01
Random effects	Variance	SD		
Item (intercept)	0.001	0.03		
Subject (intercept)	0.0007	0.02		
GC-action	0.0007	0.02		
GC-object	0.001	0.03		
AoA	0.00003	0.006		
Imageability	0.00003	0.005		
Residual	0.006	0.08		

tested with groups that are heterogeneous. What is the source of this variability? It cannot be attributed solely to the bilingual lexicon that can be studied in an experimental setting, but rather to the evolving context of learning languages. It is the context and learning history that defines the vocabulary of a bilingual speaker as well as naming patterns, degree and patterns of usage, and sociolinguistic prestige earned for each of the languages (see (Malt, Sloman, & Gennari, 2003). This may explain the higher degree of variability in picture naming of bilingual speakers.

The effect of AoA was not similar across the groups. It is typically assumed in the monolingual literature that AoA is a reliable and robust predictor of timed picture naming. An interaction between AoA and GC was observed only for the M-C group where only object naming was affected. AoA was partially significant in the monolingual group, with no effect in the C-E group. The significant effect of AoA on object naming is consistent with previous studies (Bakhtiar, Nilipour, & Weekes, 2013; Bakhtiar, Su, Lee, & Weekes, 2016; Patrick Bonin, Guillemard-Tsaparina, & Méot, 2013; Liu, Hao, Li, & Shu, 2011; Perret & Bonin, 2018). However, the null effect of AoA in the C-E bilingual speakers contrasts with results from other studies (Khwaileh et al., 2018; Schwitter, Boyer, Méot, Bonin, & Laganaro, 2004; Shao, Roelofs, & Meyer, 2014). One reason that AoA did not have a significant effect may be a high correlation between AoA and familiarity. This multi-collinearity could diminish the effects. However, this possibility was ruled out as the VIF analysis showed a safe degree of collinearity between AoA, imageability, and familiarity despite rather high correlation values. In addition, we ran another LMEM without including familiarity on the C-E data, and AoA was not still significant. One other study on Chinese action naming also reported a borderline effect of AoA (Chen & Zhu, 2015).

We believe that individual level differences in bilingual speakers such as the amount of L1 and L2 written and spoken usage and the amount of proficiency in each language could modulate AoA effects. The evidence to support this claim comes from the random effects structure of the models. While AoA was not a significant random effect in the monolingual group, in both bilingual groups, the random effect of AoA contributed significantly to the model. This means that the AoA effect was uniform across monolingual speakers, which makes sense given the lack of inter-individual variability in this group. However, in the bilingual speakers, the effect was not uniform, suggesting that in some participants, it was a significant predictor and in some it was not. This finding is in line with studies which show bilingual speakers' individual experiences interact with the item-level properties (Edmonds & Donovan, 2012; Kroll & Gollan, 2013; Ramanujan & Weekes, 2019).

In line with several previous studies, rated imageability was a strong predictor of picture naming (Akinina et al.,

2015; Bakhtiar, Jafary, & Weekes, 2017; Bird et al., 2001; Ramanujan & Weekes, 2019). It thus seems likely that imageability has a more central role than AoA in picture naming for Chinese speakers. Several linguists have argued that Chinese words carry richer and more fine-grained semantic and sensory features compared to English (Ma, Golinkoff, Hirsh-Pasek, McDonough, & Tardif, 2009). For instance, in English, one plays the piano, violin, or flute, but in Chinese (literally speaking), one 'plucks piano with fingers', 'pulls violin', and 'blows flute'. Given the additional semantic load that is carried by verbs (and actions) in Chinese, we contend that the semantic representations of words could play a more concrete role in Chinese processing than in languages such as English wherein the morphosyntax lessens the semantic load on word structure allowing them to take a more abstract role in naming, therefore diminishing the impact of imageability on picture naming. Our results seem to suggest that imageability plays a more central role in picture naming across Mandarin and Cantonese than AoA because of the importance of conceptual-semantic information.

Word frequency was not significant in any analysis. This finding is in line with Ramanujan and Weekes' study (2019) and other studies of timed picture naming (Bastiaanse, Wieling, & Wolthuis, 2016; Patrick Bonin, Chalard, Méot, & Fayol, 2002; Nishimoto, Miyawaki, Ueda, Une, & Takahashi, 2005). These researchers contend that AoA is a proxy for word frequency, and therefore once AoA is taken into consideration, the effect of frequency will not be explained (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Zevin & Seidenberg, 2002). However, another explanation for the null effect of frequency could be related to the type of corpora used to extract the frequency values in this study. The corpora available are usually created based on monolingual language use which only reflects cumulative frequency of the words (Zevin & Seidenberg, 2002). This is most relevant in the context of bilingual speakers whereby the patterns and degree of L1 and L2 usage are different compared to monolingual speakers (Edmonds & Donovan, 2012; Kroll & Gollan, 2013; Ramanujan & Weekes, 2019). Although we tried to mitigate this problem using a spoken corpus for Mandarin groups, we think that even this corpus could not adequately capture the variability of L1 and L2 language use in bilingual speakers.

An important outcome of this study is that key psycholinguistic variables predicted the response similarly across all groups of monolingual and bilingual speakers. However, bilingual speakers showed more variability in variables such as AoA. Moreover, we witnessed a difference between monolingual and bilingual norms in their capacity to explain GC interactions. This pattern was more evident when the bilinguals' languages were more similar to each other. We think that collecting normative data from bilingual participants is essential work. Only when such data is available will we be able to

properly document the diversity associated with bilingual language processing.

We revealed the similarities and differences between bilingual and monolingual lexical access. Variables like AoA and frequency that are assumed to impact on lexical selection and encoding (Alario et al., 2004) in all models of (monolingual) timed picture naming did not have significant effects in bilingual speakers. On the other hand, we identified that the effects of psycholinguistic variables are not completely uniform among all bilingual speakers. We witnessed a substantial amount of individual variability among different groups of bilingual speakers. Specifically, the bilingual speakers who spoke similar languages (Mandarin and Cantonese) showed extensively higher variability in the effects of psycholinguistic variables on timed picture naming. This variability can be attributed to linguistic and non-linguistic factors as outlined above. However, by using LMEM, we could capture and explain this variability as much as possible taking into account individual differences including demographic information such as age and education and patterns of language use. We acknowledge that other participant-level properties could be added to the model. An important new insight from the present study is that by testing the heterogeneity of bilingual language experience at the participant level, it is possible to explain more variability in timed picture naming. Thus, it is recommended that future studies include participant-level characteristics in their analyses, including even in studies of monolingual speakers. Such variables may include exposure to print, vocabulary size, spelling abilities, and other individual differences that are seen in the general population but typically controlled.

We did not have data for monolingual Cantonese speakers. Finding monolingual Cantonese speakers is not possible in Hong Kong or in Mainland China to the best of our knowledge. There might still be some monolingual Cantonese speakers available, but they are not suitable for testing predictions about picture naming because they are illiterate or aged. Moreover, they are not representative of the population. This makes it difficult to compare a matched monolingual baseline for the C-E bilingual group which is ideal.

We suggest there is a greater need to develop psycholinguistic norms based on data from multilingual speakers as they form the larger portion of the global population. Furthermore, replication studies are needed from other multilingual populations living in diverse sociolinguistic contexts with differing amounts of language use and linguistic distance.

References

Abutalebi, J., Della Rosa, P. A., Gonzaga, A. K., Keim, R., Costa, A., & Perani, D. (2013). The role of the left putamen in multilingual language production. *Brain Lang, 125*(3), 307-315.

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and cognitive processes, 23*(4), 557-582.

Akinina, Y., Malyutina, S., Ivanova, M., Iskra, E., Mannova, E., & Dragoy, O. (2015). Russian normative data for 375 action pictures and verbs. *47*(3), 691-707. https://doi.org/10.3758/s13428-014-0492-9

Alario, F. X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers, 31*(3), 531-552. https://doi.org/10.3758/bf03200732

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36*(1), 140-155. https://doi.org/10.3758/bf03195559

Alyahya, R. S. W., & Druks, J. (2016). The adaptation of the Object and Action Naming Battery into Saudi Arabic. *Aphasiology, 30*(4), 463-482. https://doi.org/10.1080/02687038.2015.1070947

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language, 59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Bakhtiar, M., Jafary, R., & Weekes, B. S. (2017). Aphasia in Persian: Implications for cognitive models of lexical processing. *J Neuropsychol, 11*(3), 414-435. https://doi.org/10.1111/jnp.12095

Bakhtiar, M., Nilipour, R., & Weekes, B. S. (2013). Predictors of timed picture naming in Persian. *Behavior Research Methods, 45*(3), 834-841. Retrieved from https://link.springer.com/content/pdf/10.3758%2Fs13428-012-0298-6.pdf

Bakhtiar, M., Su, I. F., Lee, H. K., & Weekes, B. S. (2016). Neural correlates of age of acquisition on visual word recognition in Persian. *Journal of Neurolinguistics, 39*, 1-9. https://doi.org/10.1016/j.jneuroling.2015.12.001

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283.

Barbieri, E., Basso, A., Frustaci, M., & Luzzatti, C. (2010). Argument structure deficits in aphasia: New perspective on models of lexical production. *Aphasiology, 24*(11), 1400-1423. https://doi.org/10.1080/02687030903580325

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of Experimental Psychology: Section A, 50*(3), 560-585.

Bastiaanse, R., Wieling, M., & Wolthuis, N. (2016). The role of frequency in the retrieval of nouns and verbs in aphasia. *Aphasiology,*

30(11), 1221-1239. https://doi.org/10.1080/02687038.2015.1100709

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., … Tzeng, O. (2003). Timed picture naming in seven languages. *10*(2), 344-380. https://doi.org/10.3758/bf03196494

Bates, Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967.*

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *33*(1), 73-79. https://doi.org/10.3758/bf03195349

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology, 93*(1), 89-114.

Bonin, P., Guillemard-Tsaparina, D., & Méot, A. (2013). Determinants of naming latencies, object comprehension times, and new norms for the Russian standardized set of the colorized version of the Snodgrass and Vanderwart pictures. *45*(3), 731-745. https://doi.org/10.3758/s13428-012-0279-9

Bonin, P., Meot, A., Lagarrigue, A., & Roux, S. (2015). Written object naming, spelling to dictation, and immediate copying: Different tasks, different pathways? *Q J Exp Psychol (Hove), 68*(7), 1268-1294. https://doi.org/10.1080/17470218.2014.978877

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of cognition, 1*(1), 9-9. https://doi.org/10.5334/joc.10

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLoS One, 5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Chee, M.W., Soon, C. S., Lee, H. L., & Pallier, C. (2004). Left insula activation: a marker for language attainment in bilinguals. *Proc Natl Acad Sci U S A, 101*(42), 15265-15270. https://doi.org/10.1073/pnas.0403703101

Chen, Y., & Zhu, L. (2015). Predictors of Action Picture Naming in Mandarin Chinese. *Acta Psychologica Sinica, 47*(1), 11-18. https://doi.org/10.3724/sp.J.1041.2015.00011

Craney, T. A., & Surles, J. G. (2002). Model-Dependent Variance Inflation Factor Cutoff Values. *Quality Engineering, 14*(3), 391-403. https://doi.org/10.1081/qen-120001878

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

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *31*(4), 650-658. https://doi.org/10.3758/bf03200741

Degani, T., Prior, A., & Hajajra, W. (2018). Cross-language semantic influences in different script bilinguals. *Bilingualism: Language and cognition, 21*(4), 782-804. https://doi.org/10.1017/S1366728917000311

Donderi, D. C. (2006). Visual complexity: a review. *Psychol Bull, 132*(1), 73-97. https://doi.org/10.1037/0033-2909.132.1.73

Druks, J., & Masterson, J. (2000). An object and action naming battery. *Philadelphia, PA: Psychology Press.*

Druks, J., Masterson, J., Kopelman, M., Clare, L., Rose, A., & Rai, G. (2006). Is action naming better preserved (than object naming) in Alzheimer's disease and why should we ask? , *98*(3), 332-340. https://doi.org/10.1016/j.bandl.2006.06.003

Edmonds, L. A., & Donovan, N. J. (2012). Item-level psychometrics and predictors of performance for Spanish/English bilingual speakers on an object and action naming battery. *J Speech Lang Hear Res, 55*(2), 359-381. https://doi.org/10.1044/1092-4388(2011/10-0307)

Edmonds, L. A., & Donovan, N. J. (2014). Research applications for An Object and Action Naming Battery to assess naming skills in adult Spanish–English bilingual speakers. *Behavior Research Methods, 46*(2), 456-471. https://doi.org/10.3758/s13428-013-0381-7

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(2), 515.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116-124.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *J Child Lang, 44*(3), 677-694. https://doi.org/10.1017/s0305000916000209

Glaser, W. R. (1992). Picture naming. *Cognition, 42*(1-3), 61-105.

Golestani, N., Alario, F. X., Meriaux, S., Le Bihan, D., Dehaene, S., & Pallier, C. (2006). Syntax production in bilinguals. *Neuropsychologia, 44*(7), 1029-1040.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of memory and language, 58*(3), 787-814. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2409197/pdf/nihms43338.pdf

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & cognition, 33*(7), 1220-1234. https://doi.org/10.3758/bf03193224

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition, 1*(2), 67-81.

Grosjean F. (2001). The bilingual's language modes. In Nicol J (Ed.), *One Mind,Two Languages: Bilingual Language Processing* (pp. 1-22). Oxford: Blackwell.

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models – The R Package pbkrtest. *2014, 59*(9), 32. https://doi.org/10.18637/jss.v059.i09

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5*(1), 67-104. https://doi.org/10.1080/02643298808252927

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin, 120*(1), 113.

Jonkers, R., & Bastiaanse, R. (2007). Action naming in anomic aphasic speakers: Effects of instrumentality and name relation. *Brain and Language, 102*(3), 262-272. https://doi.org/10.1016/j.bandl.2007.01.002

Kambanaros, M. (2009). Group effects of instrumentality and name relation on action naming in bilingual anomic aphasia. *110*(1), 29-37. https://doi.org/10.1016/j.bandl.2009.01.004

Khwaileh, T., Mustafawi, E., Herbert, R., & Howard, D. (2018). Gulf Arabic nouns and verbs: A standardized set of 319 object pictures and 141 action pictures, with predictors of naming latencies. *Behavior Research Methods, 50*(6), 2408-2425. Retrieved from https://link.springer.com/content/pdf/10.3758%2Fs13428-018-1019-6.pdf

Kilgarriff, A., Reddy, S., Pomikálek, J., & Avinesh, P. V. S. (2010). A Corpus Factory for Many Languages. *LREC.*

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kroll, J. F., & Gollan, T. H. (2013). Speech Planning in Two Languages: What Bilinguals Tell Us about Language Production. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford Hndbook of Language Production* (pp. 165-181). USA: Oxford University Press.

Kuzmina, E., Goral, M., Norvik, M., & Weekes, B. S. (2019). What Influences Language Impairment in Bilingual Aphasia? A Meta-Analytic Review. *Frontiers in psychology, 10*, 445-445. https://doi.org/10.3389/fpsyg.2019.00445

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *2017, 82*(13), 26. https://doi.org/10.18637/jss.v082.i13

Levelt, W. J. (1999). Models of word production. *Trends in cognitive sciences, 3*(6), 223-232. Retrieved from https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(99)01319-4?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1364661399013194%3Fshowall%3Dtrue

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav Brain Sci, 22*(1), 1-38; discussion 38-75.

Liu, Y., Hao, M., Li, P., & Shu, H. (2011). Timed picture naming norms for Mandarin Chinese. *PLoS One, 6*(1), e16505. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3027682/pdf/pone.0016505.pdf

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494-1502. https://doi.org/10.3758/s13428-016-0809-y

Lutz, C., Lee, L. Y., & Weekes, B. (2018). Wor(l)ds apart ? does the N400 reflect bilingual language distance and meaning in translation? An ERP study of the effects of L1-L2 distance and translation direction in German-English and Cantonese-English bilinguals. *Human Neuroscience Archive.* https://doi.org/10.3389/conf.fnhum.2018.228.00080

Ma, W., Golinkoff, R. M., Hirsh-Pasek, K., McDonough, C., & Tardif, T. (2009). Imageability predicts the age of acquisition of verbs in Chinese children. *Journal of child language, 36*(2), 405-423. https://doi.org/10.1017/s0305000908009008

Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J., & Carballal, A. (2015). Computerized measures of visual complexity. *Acta psychologica, 160*, 43-57. https://doi.org/10.1016/j.actpsy.2015.06.005

Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Universality and language specificity in object naming. *49*(1), 20-42. https://doi.org/10.1016/s0749-596x(03)00021-4

Momenian, M., Nilipour, R., Samar, R. G., Cappa, S. F., & Golestani, N. (2018). Morpho-syntactic complexity modulates brain activation in Persian-English bilinguals: An fMRI study. *Brain and Language, 185*, 9-18. https://doi.org/10.1016/j.bandl.2018.07.001

Morrison, C. M., Hirsh, K. W., & Duggan, G. B. (2003). Age of Acquisition, Ageing, and Verb Production: Normative and Experimental Data. *The Quarterly Journal of Experimental Psychology Section A, 56*(4), 1-26. https://doi.org/10.1080/02724980244000594

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The Influence of Language Proficiency on Lexical Semantic Processing in Native and Late Learners of English. *Journal of Cognitive Neuroscience, 25*(4), 1205 –1223.

Nicewander, W. A. (2018). Modifying Spearman's Attenuation Equation to Yield Partial Corrections for Measurement Error-With Application to Sample Size Calculations. *Educational and psychological measurement, 78*(1), 70-79. https://doi.org/10.1177/0013164417713571

Nilipour, R., Bakhtiar, M., Momenian, M., & Weekes, B. S. (2017). Object and action picture naming in brain-damaged Persian speakers with aphasia. *Aphasiology, 31*(4), 388-405.

Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. *37*(3), 398-416. https://doi.org/10.3758/bf03192709

Parris, B., & Weekes, B. (2001). Action Naming in Dementia. *Neurocase, 7*(6), 459-471. https://doi.org/10.1093/neucas/7.6.459

Perret, C., & Bonin, P. (2018). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods.* https://doi.org/10.3758/s13428-018-1100-1

Pinheiro, J., & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS.* USA: Springer.

Ramanujan, K. (2019). The impact of Relative Language Distance on Bilingual Language Control – a functional imaging study. *bioRxiv*, 771212. https://doi.org/10.1101/771212

Ramanujan, K., & Weekes, B. S. (2019). Predictors of lexical retrieval in Hindi–English bilingual speakers. *Bilingualism: Language and cognition*, 1-9. https://doi.org/10.1017/s1366728918001177

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review, 107*(3), 460.

Schwitter, V., Boyer, B., Méot, A., Bonin, P., & Laganaro, M. (2004). French normative data and naming times for action pictures. *Behavior Research Methods, Instruments, & Computers, 36*(3), 564-576. https://doi.org/10.3758/bf03195603

Shao, Z., Roelofs, A., & Meyer, A. S. (2014). Predicting naming latencies for action pictures: Dutch norms. *46*(1), 274-283. https://doi.org/10.3758/s13428-013-0358-6

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol Bull, 86*(2), 420-428. https://doi.org/10.1037//0033-2909.86.2.420

Sörman, D. E., Hansson, P., & Ljungberg, J. K. (2019). Different Features of Bilingualism in Relation to Executive Functioning. *Frontiers in psychology, 10*(269). https://doi.org/10.3389/fpsyg.2019.00269

Suh, S., Yoon, H. W., Lee, S., Chung, J. Y., Cho, Z. H., & Park, H. (2007). Effects of syntactic complexity in L1 and L2; an fMRI study of Korean-English bilinguals. *Brain Res, 1136*(1), 178-189. https://doi.org/10.1016/j.brainres.2006.12.043

Székely, A., & Bates, E. (2000). Objective visual complexity as a variable in studies of picture naming. *Center for Research inLanguageNewsletter, 12*(2).

Szekely, A., Damico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., … Bates, E. (2005). Timed Action and Object Naming. *Cortex, 41*(1), 7-25. https://doi.org/10.1016/s0010-9452(08)70174-6

Thompson, C. K. (2003). Unaccusative verb production in agrammatic aphasia: the argument structure complexity hypothesis. *Journal of Neurolinguistics, 16*(2), 151-167. https://doi.org/10.1016/S0911-6044(02)00014-3

Weekes, B. S., Shu, H., Hao, M., Liu, Y., & Tan, L. H. (2007). Predictors of timed picture naming in Chinese. *Behavior Research Methods, 39*(2), 335-342. https://doi.org/10.3758/bf03193165

Yan, H., Zhang, Y. M., Xu, M., Chen, H. Y., & Wang, Y. H. (2016). What to do if we have nothing to rely on: Late bilinguals process L2 grammatical features like L1 natives. *Journal of Neurolinguistics, 40*, 1-14. https://doi.org/10.1016/j.jneuroling.2016.04.002

Zevin, J. D., & Seidenberg, M. S. (2002). Age of Acquisition Effects in Word Reading and Other Tasks. *47*(1), 1-29. https://doi.org/10.1006/jmla.2001.2834