# Developing psycholinguistic norms for action pictures in Cantonese

Mohammad Momenian[a], Kai-Yan Dustin Lau[a], and Mehdi Bakhtiar[b]

[a]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR;

[b]Unit of Human Communication, Development, and Information Sciences, The University of Hong Kong, Hong Kong, Hong Kong SAR

## Abstract

The purpose of this study was to establish psycholinguistic norms for 249 action pictures in Cantonese, a language with few norms available. We provide normative data for rated visual complexity, rated age of acquisition, name agreement, word frequency and rated familiarity in this study. Forty participants were recruited to participate in both timed picture naming and rating experiments. The linear mixed effect analysis revealed that familiarity, visual complexity, and name agreement were significant predictors of action naming in Cantonese. However, AoA did not show any significant effect on action naming, which is consistently observed in previous studies of action picture naming in Chinese. The possible explanation for null effect of AoA on naming latency are discussed. This set of psycholinguistic norms in Cantonese could serve as a valuable resource for future psycholinguistic, neurolinguistic and clinical studies in Cantonese.

## Introduction

Picture naming is one of the most frequently used methods in psycholinguistics to study lexical access and retrieval (Perret & Bonin, 2018). There are several models for lexical access and retrieval such as Levelt's model (Levelt, 2001) and Cohort model (Marslen-Wilson & Welsh, 1978). These models agree that speakers go through mainly three stages during lexical retrieval in picture naming. The first stage is visual recognition and conceptual identification in which the speakers visually recognize an entity and match it with a concept in their mental representation. The next stage is lexical selection where a lexical entity is searched and retrieved through the mental lexicon, which is followed by the articulation of the retrieved word at the final stage (Bonin, Meot, Lagarrigue, & Roux, 2015; Glaser, 1992; Johnson, 1996; W. Levelt, 1999; Levelt, Roelofs, & Meyer, 1999; Perret & Bonin, 2018; Rapp & Goldrick, 2000).

It has been reported that different psycholinguistic properties such as visual complexity (VC) of a picture, name agreement (NA), imageability of a word, familiarity of a concept, frequency of occurrence, and age of acquisition (AoA) show effects across different stages of lexical access and retrieval in a picture naming task, which can be reflected in the speed of processing (Alario et al., 2004; Humphreys, Riddoch, & Quinlan, 1988; Perret & Bonin, 2018). For example, VC is supposed to exert its influence on visual recognition stage (Alario et al., 2004), whereas familiarity and imageability show their effects on conceptual identification (Barry, 1997; Ellis & Morrison, 1998; Weekes, Shu, Hao, Liu, & Tan, 2007). NA, AoA, and word frequency are assumed to have an influence on lexical retrieval (Alario et al., 2004). Although many of the previous studies have been done on object naming, there are still abundant studies showing the same effects on action naming (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).

The majority of the findings on picture naming comes mainly from Indo-European languages such as English (Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; Snodgrass & Vanderwart, 1980), Dutch (Shao & Stiegert, 2016), French (Alario & Ferrand, 1999; Bonin, Peereman, Malardier, Méot, & Chalard, 2003), Spanish (Cuetos, Ellis, & Alvarez, 1999; Manoiloff, Artstein, Canavoso, Fernández, & Segui, 2010) and Italian (Dell'Acqua, Lotto, & Job, 2000; Navarrete, Arcara, Mondini, & Penolazzi, 2019; Nisi, 2000). Only few other languages, for instance such as Persian, Czech and Arabic (Alyahya & Druks, 2016; Bartos, Hohinova, & Holla, 2020; Nilipour et al., 2017), have standard picture naming batteries which could take into account the unique linguistic features and cultural variations accompanied with each linguistic community. Moreover, the majority of available norms and ratings are established for object pictures. There is a lack of norms for action pictures in many languages (Bonin, Boyer, Méot, Fayol, & Droit, 2004) including Cantonese.

To the best of our knowledge, there is only one published study reporting psycholinguist norms for action picture naming in Cantonese (Momenian, Bakhtiar, Chan, Cheung, & Weekes, 2021), though there are few studies reporting these norms in Mandarin which belongs to the same language family as Cantonese (see Table 1), a dialect of the Yue family of Chinese that is spoken primarily in southern Chinese province of Guangdong and Guangxi, as well as Hong Kong and Macau. As a result of emigration from Hong Kong, Cantonese is also used in Chinese communities in Southeast Asia, such as Singapore and Malaysia, Australia, Europe, and North America (Matthews & Yip, 1994). There are certain variations regarding the Cantonese spoken in different places, due to the influence of the culture in different places. Yet, the influence from

Hong Kong Cantonese is considered strongest because of the popularity of the movies, television programmes, pop songs, and other pop culture produced in Hong Kong. Therefore, Hong Kong serves as the perfect location to obtain norm of Cantonese.

Momenian, Bakhtiar, et al. (2021) developed norms for 144 object and 86 action pictures for Cantonese-English bilingual speakers in Hong Kong. Their results showed that NA, familiarity, and imageability had significant effects on both object and action picture naming latency. The effects of AoA, VC and word frequency were not, however, significant.

<insert table 1 here>

The purpose of this study was to establish action picture norms for psycholinguistic variables including VC, NA, AoA, familiarity, and written word frequency in Cantonese. It is notable that the current study compared to the previous study in Cantonese (Momenian, Bakhtiar, et al., 2021) established psycholinguistic norms for a different and larger set of action pictures (249 pictures) . We used these norms to predict action picture naming latency in Cantonese speakers (Chen & Zhu, 2015; Crepaldi, Che, Su, & Luzzatti, 2012; Momenian, Bakhtiar, et al., 2021)

## Methods

This research includes a picture naming experiment followed by a rating study in which different psycholinguistic norms were provided by the same participants for the same set of pictures. The preparatory rating study is presented first here for the sake of consistency with similar studies in the literature.

### Rating study

### Participants

Forty native speakers of Cantonese who were mainly undergraduate students ranging in age between 19-26 years old (Mean=22.5, SD= 1.90) participated in the rating study. The education level of participants ranged from 13 to 16 years. The dominant language in all participants was Cantonese. However, they were able to communicate in at least one or two other languages such as English and Mandarin. The participants were recruited by posting flyers across the university campus or some relevant social media.

**Materials**

Two hundred and seventy five action pictures were retrieved from International Picture Naming Project (Szekely et al., 2004). IPNP is a database which provides ratings for 520 object and 275 action pictures in several language (Szekely et al., 2004). Twenty-six pictures were excluded from the picture set as they were not recognized correctly by 15 native Cantonese speakers who participated in the pilot study. The purpose of the pilot study was to exclude pictures which were either culturally unfamiliar or pictorially unclear or ambiguous. We removed 26 pictures based on the results from pilot study.

**Procedure**

A link was sent to all participants through Survey Monkey platform for rating the AoA, Familiarity and VC. Participants were required to complete the rating procedure for all pictures within the same day on a computer. As mentioned before participants took part in the picture naming experiment first followed by the rating study to prevent any repetition effects on naming latency. Among all participants three participants did not complete one of these three surveys. Around 97% of the participants finished each survey within one day, while about 3% finished the survey within two to six days.

For VC rating, participants were asked to rate the visual complexity of each picture based on the complexity of visual details and lines existing in the picture. A value of 1 indicated the lowest visual complexity , and 5 indicated the highest visual complexity of the picture.

For familiarity, the participants needed to rate each picture based on the amount they encountered or how often thought about the action using a 5-point rating scale from 1 defined as very unfamiliar to 5 defined as very familiar.

For AoA, we used a 7-point scale with 1 indicating 0-2 years old; 2 indicating 2-4 years old; 3 indicating 4-6 years old; 4 indicating 6-8 years old; 5 indicating 8-10 years old; 6 indicating 10-12 years old and 7 indicating 13 years old or older. Participants were presented with pictures for AoA rating. We also looked at the AoA of words from parent-report norms in Cantonese such as MacArthur-Bates Communicative Development Inventories (CDIs) (Frank, Braginsky, Yurovsky, & Marchman, 2017). More than 50 percent of the action names used in this study were not available in the parent-report norms. We could not, therefore, study the correlation between rated AoA and parent-report AoA.

In addition to the ratings, we also obtained the frequency of the action words from the CantoLexicon Project (Lau, Su, & Yum, 2019), which consists of texts in over 120,000 news pieces published by the eight most popular newspapers in Hong Kong. We need to mention that most of the actions conveyed with a simple word in English are usually expressed in Cantonese by the combination of at least two characters. These combinations are sometimes called verb compounds or light verb constructions (Momenian, Cham, Mohammad Amini, Radman, & Weekes, 2021). For example, the words /tiu6/ [jump] and /soey2/ [water] were used to form the compound /tiu6-soey2/ by our participants to name the action "dive". It is because the verb /tiu6/ alone refers to "jump", which is not sufficient to represent the meaning of "dive". Similarly, the words /zyu2/ [cook] and /je5sik6/ [food] were used to form the compound /zyu2-je5sik6/ by our participants to name the action "cook", because /zyu2/ is rarely used without an object. All these types of verb compounds are, therefore, intransitive because they already include an object. In our analysis, we will take this into account.

The percentage of name agreement for each picture was calculated based on the total number of common names given by all participants divided by the total number of accurate names during the picture naming study (Weekes et al., 2007). The ratings derived are available online as psycholinguistic norms for Cantonese in the following link at Open Science Framework (https://osf.io/jhgpb/). We used Intraclass Correlation (ICC) to find the reliability of ratings for familiarity, VC, and AoA (Shrout & Fleiss, 1979). Ratings with a value above 0.75 are supposed to have good reliability suggested by Koo and Li (2016). The corrected correlation in this study was calculated following the practice suggested by Nicewander (2018) (See Table 2).

<insert Table 2 about here>

**Picture naming Experiment**

**Participants**

The same 40 participants who took part in the rating study participated in the picture naming experiment. They all had normal or corrected-to-normal vision. All participants completed a consent form prior to the experiments. All the procedures in this study were approved by the Ethical Sub-Committee at Hong Kong Polytechnique University.

**Procedure**

There was a practice session before we started the main experiment so that participants were fully aware of the experiment requirements. They were told that they should not think too much about each picture and should produce the name which comes to their mind as quickly as possible upon seeing the picture. They were asked not to cough or produce any sounds such as starters or fillers e.g. [um] during or before each response. We presented fixation point at the center of the monitor for 500ms before each picture. After that, each picture was presented in the middle of the screen and the participants were required to name the picture within 3000ms as the time-out period. Intertrial interval was 500ms. Each of the participants' naming responses were recorded using a Shure SM58 dynamic microphone connected to a Roland Quad-Capture audio interface controlled using the E-prime software for off-line analysis. The whole session consisted of three blocks, and the participants had a short break between the blocks. All the pictures were randomly presented. No feedback on accuracy of naming was given. We were measuring the naming latency in this experiment.

**Analysis plan**

All naming responses were transcribed off-line for judgment of accuracy and naming agreement allowing us to find errors made by participants including production of nontarget words and unwanted sounds. RT of each trial, defined as the speech onset time of each naming attempt, was collected by submitting the normalized audio recordings to Chronset, an automatic tool for detecting speech onset (Roux, Armstrong, & Carreiras, 2017).

Generalized Linear Mixed Effects Modeling (GLMEM) was used to analyze the data using lme4 package with R software (Baayen, Davidson, & Bates, 2008). We follow a similar pipeline used by Momenian, Bakhtiar, et al. (2021). We started with a maximal model which was informed by our design (Barr, Levy, Scheepers, & Tily, 2013) . We excluded missing responses, incorrect responses, and responses that were above 2500ms (11.85%) from further analysis. Every naming response that did not correspond with the translation of the action name in English or content of the picture was considered as incorrect naming. For instance, if the picture of 'eating' was named as 'drinking', it was marked as incorrect naming.

All continuous independent variables were first standardized before the analysis. To examine whether our model would suffer from collinearity problems, variance inflation factor (VIF) was used. Not every high correlation between variables means a high VIF necessarily which makes it a suitable metric to be used in this study. Usually values larger than 5 or 10 could be excluded from the analysis based on the recommendation by Craney and Surles (2002). Following Lo and Andrews (2015), we did not transform our dependent variable. We used link functions assuming a Gaussian distribution. Gamma and Inverse Gaussian distributions are commonly used in psycholinguistic research (Lo & Andrews, 2015). We fitted six models including models on raw RT (DV = RT), inverse RT (DV = − 1000/RT), Gamma and Inverse Gaussian distributions with identity link function, and Gamma and Inverse Gaussian distributions with inverse link function. We used plots and fit indices such as AIC to decide which model best represented the distribution of data.

Our fixed variables were rated VC, rated AoA, log frequency, NA, and rated familiarity, in addition to control variables such as number of characters and type of the verb (compound or simple). The random effects included both intercepts and slopes. The model had item and participant intercepts, and only by-participants slopes for rated VC, rated AoA, log frequency, NA, and rated familiarity. This model was informed by our design and suggested by Bar et al. (2013).

After deciding on what the maximal model should look like in the study, we started modelling. We first fitted the maximal model and then did a Singular Value Decomposition (SVD) on the covariance matrix of the maximal model using principal component analysis (PCA) (Bates et al., 2015). One major problem with maximal models is that they sometimes suffer from convergence and over-specification problems. PCA analysis helps avoid these problems by telling which random effects could be removed from the model without having a significant effect on the results. After doing PCA, we compared the maximal model with the next model using Likelihood Ratio Tests (LRT). LRT could reveal which model fits the data better. To avoid convergence problems, we did not include correlation parameters for the random effects in the maximal model (Bates et al., 2015). However, after finding the most parsimonious random effects structure we added the correlation parameters. We used LRT again to compare the model with correlation parameters and the model without. If the result was significant, we added the correlation parameters to the model with the best random effects structure. Verb type was deviation coded (-.5 and .5) at this stage.

Finally, to determine which of the variables in our study was a significant predictor, conditional F-tests were used. F-tests are preferred over LRT because using LRT to examine fixed effects could be anti-conservative leading to unreliable findings (Halekoh & Højsgaard, 2014; Luke, 2017; Pinheiro & Bates, 2000). Kenward-Roger approximations were used to produce denominator degrees of freedom which have more promising Type 1 error rates compared with LRT and Wald tests (Kuznetsova, Brockhoff, & Christensen, 2017; Luke, 2017). To find out if the effect of verb type was significant, this variable was dummy coded (1 and 0) once we found the best model.

Our study had enough power based on the recommendation by Brysbaert and Stevens (2018). Based on their study at least 40 participants with 40 stimuli (1600 observations per condition) are needed for a reaction time experiment to have enough power. We had more than this threshold in our study. The data for the Experiment and R codes are available online at the following link (https://osf.io/jhgpb/).

## Results

The results of VIF analysis suggested that multicollinearity among the variables was not an issue. Among all 6 distributions we tested, Gamma Distribution with identity link function was the best based on AIC index and diagnostic plots. Based on PCA results and the variance-covariance matrix we started removing random effects with the lowest variance followed by LRTs. The removal of NA $x^2(1)= 19.73$, $p < 0.001$, AoA $x^2(1) = 12.09$, $p < 0.001$, frequency $x^2(1)= 18.59$, $p < 0.001$, familiarity $x^2(1)= 8.57$, $p < 0.01$, and VC $x^2(1)= 15.08$, $p < 0.001$ had significant effects on the model fit. For the final stage, all the zero correlation parameters were removed from the previous best model. The results of the LRT showed that including correlation parameters did not increase the model fit significantly $x^2(15) = 3.80$, $p = 0.99$ (See table 3 for a summary of the GLMEM). When it comes to the predictor variables in the model, only the effects of familiarity, VC and NA were significant (See table 3 for further information). There was no difference between verbs which were compound and verbs which were simple (t = -0.20, $p = 0.83$).

<Insert table 3 about here>

## Discussion

In the current study, the psycholinguistic norms of AoA, familiarity, VC, frequency, and NA for 249 action pictures were established in Cantonese. In addition, we examined which of these variables were able to predict naming latency. The results of GLMEM analysis showed that familiarity, VC, and NA were the strongest predictors of action picture naming in Cantonese. AoA, one of the most important predictors in prior studies, did not have significant effects on picture naming.

When it comes to name agreement, our findings are consistent with many previous studies in other languages such as French, English, German, and Chinese (Alario & Ferrand, 1999; Bates et al., 2003; Perret & Bonin, 2018; Szekely et al., 2005) and Chinese including Cantonese and Mandarin (Chen & Zhu, 2015; Momenian, Bakhtiar, et al., 2021; Weekes et al., 2007). A recent meta-analysis (Perret & Bonin, 2018) suggested that NA is an essential predictor of picture naming and should be controlled in all psycholinguistic studies. This is not new because studies in several other languages have shown that the number of alternative names associated with an object or action affects how quickly it could be named (Alario et al., 2004; Bates et al., 2003); the less competition in the selection of names for the pictures, the faster they should be retrieved (Ramanujan & Weekes, 2019).

Findings from previous studies are mixed regarding the effects of familiarity on picture naming latency. Although our findings are in line with the majority of prior studies on Chinese (Chen & Zhu, 2015; Weekes et al., 2007; Zhou & Chen, 2017), there are studies on other languages which have reported null effects (Alario et al., 2004; Bonin, Chalard, Méot, & Fayol, 2002; Bonin et al., 2003; Dell'Acqua et al., 2000; Ellis & Morrison, 1998; Nishimoto, Miyawaki, Ueda, Une, & Takahashi, 2005). It seems that familiarity is a more influential variable than frequency in picture naming at least in Chinese (Gernsbacher, 1984; Gordon, 1985). In a study on Arabic, Khwaileh et al. (2018) claimed that familiarity could be an alternative to spoken frequency. Given that our measure of frequency and that of similar studies in Chinese was based on the written language, it might be expected to regard familiarity as a stronger predictor of picture naming than written frequency in Cantonese. In line with Khwaileh et al. (2018), we assert that familiarity could be a proxy for frequency in Chinese.

Our finding that written frequency was not a strong predictor of picture naming is in line with several studies in other languages (Bastiaanse, Wieling, & Wolthuis, 2016; Bonin et al., 2002; Nishimoto et al., 2005) and Chinese (Chen & Zhu, 2015; Crepaldi et al., 2012; Momenian, Bakhtiar, et al., 2021). One explanation for why frequency did not show a significant effect could be attributed to the type of corpus used in our study. We used a written corpus which could represent the cumulative frequency of the words at best (see Zevin & Seidenberg, 2002). Unfortunately, there are no reported spoken corpora in Cantonese (for Mandarin Chinese see (Cai & Brysbaert, 2010). The availability of spoken corpus in Cantonese would allow researchers to test whether familiarity could still be a significant predictor in the presence of spoken frequency or not. A second explanation for lack of frequency effect could be related to the effect of AoA. Several studies suggest that once AoA is already modelled, frequency effect could disappear (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Zevin & Seidenberg, 2002). The latter explanation is less tenable in this study because AoA itself was not significant.

The fact that AoA did not have a significant effect on naming may seem surprising in this study because AoA has always been one of the strongest predictors of both object and action naming in several languages (Bennett, Burnett, Siakaluk, & Pexman, 2011; Khwaileh et al., 2018; Schwitter, Boyer, Méot, Bonin, & Laganaro, 2004; Shao & Stiegert, 2016). However, it is important to note that AoA was not significant in the only existing study in Cantonese

(Momenian, Bakhtiar, et al., 2021) and in any of the three studies available on action naming in Mandarin (Chen & Zhu, 2015; Crepaldi et al., 2012; Momenian, Bakhtiar, et al., 2021). If we think of familiarity as a proxy for spoken frequency, then it is possible that familiarity could be washing away any AoA effects. It is also notable that several studies have shown high correlations between AoA and familiarity (for instance see Bakhtiar & Weekes, 2015), and it has been argued by some authors (see Crepaldi et al., 2012) that subjective ratings of AoA are the result of rater's self-contemplation about a word representation which is similar to concept familiarity. Therefore, it is also possible that AoA effects might be exerted by the concept familiarity (at least partially) across different Chinese studies including the current study.

Our finding regarding VC in this study is not consistent with prior studies on action picture naming in Chinese (Chen & Zhu, 2015; Crepaldi et al., 2012; Momenian, Bakhtiar, et al., 2021). A recent meta-analysis (Perret & Bonin, 2018) also revealed a null effect for VC. One explanation could be that the visual properties of an action picture could be very important in Cantonese because verbs are supposed to carry richer semantic and sensory features in this language (see Ma, Golinkoff, Hirsh-Pasek, McDonough, & Tardif, 2009). Another explanation could be the absence of other variables such as imageability in this study which could possibly drive VC significant. It would be interesting to see if VC could still be significant in the presence of imageability or not. In this study we did not collect imageability ratings for the pictures because we used picturable verbs as the stimuli in this study. Even if we had tried to collect imageability ratings, it was expected that most of the items would be highly imageable. However, we believe future studies could obtain data on variables such as imageability and manipulability of the action verbs.

## Conclusion and limitations

This study provides new data available on Cantonese action picture norms. There are norms available in many languages of the world, and the norms in this study enable the researchers to compare findings across languages. The norms we established in Cantonese could be used for both research and practice on healthy populations and people with brain damage. The normed pictures could be used in a variety of psycholinguistic and neurolinguistic studies, particularly on semantic processing in Cantonese. The established norms are very useful for studying language impairment in brain damaged populations and determining language functions during awake brain surgery. One example of how action pictures could be used is when doing awake surgery on people who have frontal lesions (Rofes & Miceli, 2014). Numerous studies show that left inferior frontal regions are very important for sentence processing, particularly verb processing (Price, 2010; Vigliocco, Vinson, Druks, Barber, & Cappa, 2011).

The findings of this study pose several interesting questions for psycholinguistic theories of picture naming. An interesting question is that why AoA and frequency are not significant predictors in action naming in Chinese. These variables have always been very important in other languages. However, since AoA values for this study is limited to subjective AoA rated by adults, future studies might also examine the effects of objective AoA on action naming in Cantonese.  Lastly, we believe more cross-linguistic studies need to be done to further validate the observed effects in other languages.

The findings presented in this study must be interpreted in light of some limitations. The participants of this study were mainly recruited from an undergraduate population with an age range of 19-26 years old. This could possibly mean that the norms should be used cautiously with other populations such as children, teenagers, and older adults. Future studies could be designed to establish norms for the same set of action pictures for these populations. Another limitation of the study is that the same participants were recruited for both rating and naming experiments. We are aware that many studies recruited different participants for each experiment to avoid repetition effects. However, our ratings had significant positive correlation with those reported by Momenian, Bakhtiar, et al. (2021) in Cantonese that recruited different participants for naming and rating experiments, suggesting that our design did not have a significant effect on the quality of the ratings.

## References


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. doi: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. doi: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. doi: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. doi:10.1016/j.jml.2007.12.005

Bakhtiar, M., & Weekes, B. (2015). Lexico-semantic effects on word naming in Persian: Does age of acquisition have an effect? *Memory & cognition, 43*(2), 298-313. doi:10.3758/s13421-014-0472-4

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.

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. doi: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.

Bartos, A., Hohinova, M., & Holla, M. (2020). High electronic name agreement of 70 pictures in normative study of 5,290 Czechs for easy multicultural replication. *Appl Neuropsychol Adult*, 1-12. doi:10.1080/23279095.2020.1753744

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. doi:10.1080/02687038.2015.1100709

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

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

Bennett, S. D. R., Burnett, A. N., Siakaluk, P. D., & Pexman, P. M. (2011). Imageability and body–object interaction ratings for 599 multisyllabic nouns. *Behavior Research Methods, 43*(4), 1100-1109. doi:10.3758/s13428-011-0117-5

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

Bonin, P., Boyer, B., Méot, A., Fayol, M., & Droit, S. (2004). Psycholinguistic norms for action photographs in French and their relationships with spoken and written latencies. *Behavior Research Methods, Instruments, & Computers, 36*(1), 127-139. doi:10.3758/bf03195558

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., 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. doi:10.1080/17470218.2014.978877

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). 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. *Behavior Research Methods, Instruments, & Computers, 35*(1), 158-167. doi:10.3758/bf03195507

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of cognition, 1*(1), 9-9. doi: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. doi:10.1371/journal.pone.0010729

Chen, Y., & Zhu, L. (2015). Predictors of Action Picture Naming in Mandarin Chinese. *Acta Psychologica Sinica, 47*(1), 11-18. doi: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. doi: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. doi:10.3233/ben-2012-119002

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

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture Naming by Young Children: Norms for Name Agreement, Familiarity, and Visual Complexity. *Journal of experimental child psychology, 65*(2), 171-237. doi:10.1006/jecp.1996.2356

Dell'Acqua, R., Lotto, L., & Job, R. (2000). Naming times and standardized norms for the italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers, 32*(4), 588-615. doi:10.3758/bf03200832

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? *Brain and Language, 98*(3), 332-340. doi: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. doi:10.1044/1092-4388(2011/10-0307)

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.

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. doi:10.1017/s0305000916000209

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*(2), 256.

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

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of memory and language, 24*(6), 631-645. doi:https://doi.org/10.1016/0749-596X(85)90050-6

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models    The R Package pbkrtest. *Journal of Statistical Software, 59*(9), 32. doi: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. doi:10.1080/02643298808252927

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

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

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. doi:10.1016/j.jcm.2016.02.012

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

Lau, K., Su, I.-F., & Yum, Y.-N. C. (2019). *The Canto-Lexicon Project: A Preliminary Report*. Paper presented at the Academy of Aphasia 57th Annual Meeting, Macau.

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.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences, 98*(23), 13464. doi:10.1073/pnas.231459498

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in psychology, 6*. doi:10.3389/fpsyg.2015.01171

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

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. doi:10.1017/s0305000908009008

Manoiloff, L., Artstein, M., Canavoso, M. B., Fernández, L., & Segui, J. (2010). Expanded norms for 400 experimental pictures in an Argentinean Spanish-speaking population. *Behavior Research Methods, 42*(2), 452-460. doi:10.3758/brm.42.2.452

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29-63. doi:https://doi.org/10.1016/0010-0285(78)90018-X

Matthews, S. J., & Yip, V. (1994). *Cantonese: A Comprehensive Grammar*: Routledge.

Momenian, M., Bakhtiar, M., Chan, Y. K., Cheung, S. L., & Weekes, B. S. (2021). Picture naming in bilingual and monolingual Chinese speakers: Capturing similarity and variability. *Behavior Research Methods*. doi:10.3758/s13428-020-01521-1

Momenian, M., Cham, S. K., Mohammad Amini, J., Radman, N., & Weekes, B. (2021). Capturing the effects of semantic transparency in word recognition: a cross-linguistic study on Cantonese and Persian. *Language, Cognition and Neuroscience, 36*(5), 612-624. doi:10.1080/23273798.2020.1862878

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. doi:10.1080/02724980244000594

Navarrete, E., Arcara, G., Mondini, S., & Penolazzi, B. (2019). Italian norms and naming latencies for 357 high quality color images. *PLoS One, 14*(2), e0209524. doi:10.1371/journal.pone.0209524

Ni, L., Liu, Y., Yu, W., & Fu, X. (2019). The China Image Set (CIS): A New Set of 551 Colored Photos With Chinese Norms for 12 Psycholinguistic Variables. *Frontiers in psychology, 10*. doi:10.3389/fpsyg.2019.02631

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. doi: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. *Behavior Research Methods, 37*(3), 398-416. doi:10.3758/bf03192709

Nisi, M., & Longoni, A. M. (2000). Italian measurement on the relation of name, familiarity, and acquisition age for the 260 figures of Snodgrass and Vanderwart. *Giornale Italiano di Psicologia, 27*, 205-218.

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

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

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann N Y Acad Sci, 1191*, 62-88. doi:10.1111/j.1749-6632.2010.05444.x

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

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

Rofes, A., & Miceli, G. (2014). Language Mapping with Verbs and Sentences in Awake Surgery: A Review. *Neuropsychology Review, 24*(2), 185-199. doi:10.1007/s11065-014-9258-5

Roux, F., Armstrong, B. C., & Carreiras, M. (2017). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods, 49*(5), 1864-1881. doi:10.3758/s13428-016-0830-1

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. doi:10.3758/bf03195603

Shao, Z., & Stiegert, J. (2016). Predictors of photo naming: Dutch norms for 327 photos. *Behavior Research Methods, 48*(2), 577-584. doi:10.3758/s13428-015-0613-0

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

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory, 6*(2), 174.

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. doi:10.1016/s0010-9452(08)70174-6

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., . . . Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of memory and language, 51*(2), 247-250. doi:10.1016/j.jml.2004.03.002

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews, 35*(3), 407-426.

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. doi:10.3758/bf03193165

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

Zhou, D., & Chen, Q. (2017). Color Image Norms in Mandarin Chinese. *Frontiers in psychology, 8*(1880). doi:10.3389/fpsyg.2017.01880

**Table 1**

*A Summary of Significant Predictors Found in Picture Naming Studies in Mandarin*

| Study | Type of Pictures | Significant Predictors |
|---|---|---|
| Momenian, Bakhtiar, et al. [2021] | Object & Action | NA<br>Imageability |
| Chen & Zhu [2015] | Action | Imageability<br>Familiarity<br>NA |
| Crepaldi et al. [2012] | Object & Action | Imageability |
| Weekes et al. [2007] | Object | AoA<br>Familiarity<br>NA<br>Frequency |
| Zhou & Chen [2017] | Object | Imageability<br>AoA<br>Familiarity<br>NA<br>Frequency |
| Ni, Liu, Yu, & Fu [2019] | Object | VC<br>Imageability<br>AoA<br>Familiarity<br>NA |

**Table 2**

*The Corrected Correlation Matrix for the Key Variables in the Experiment*

| | RT | Familiarity | AoA | VC | NA | Frequency(log) |
|---|---|---|---|---|---|---|
| RT | 1.00 | | | | | |
| Familiarity | -0.34 | 0.86* | | | | |
| AoA | 0.20 | -0.58 | 0.92* | | | |
| VC | 0.36 | -0.68 | 0.38 | 0.79* | | |
| NA | -0.29 | 0.39 | -0.17 | -0.35 | 1.00 | |
| Frequency(log) | 0.01 | -0.06 | -0.18 | 0.12 | 0.19 | 1.00 |

* ICC

**Table 3**

*A Summary of the Significant Effects in the Naming Experiment*

| **Fixed effects** | *t* value | *Std. Error* | *p value* | *95% CI* |
|---|---|---|---|---|
| Intercept | 110.66 | 13.47 | **0.001** | [1464.84, 1517.67] |
| Frequency (log) | 0.81 | 10.53 | 0.42 | [-12.21, 29.10] |
| Familiarity | -4.12 | 11.87 | **0.001** | [-73.31, -26.74] |
| VC | 9.87 | 10.34 | **0.001** | [81.90, 122.45] |
| AoA | 1.09 | 9.47 | 0.27 | [-8.23, 28.91] |
| NA | -8.49 | 8.88 | **0.001** | [-92.87, -58.03] |
| Verb type | -0.20 | 19.32 | **0.83** | [-41.86, 33.88] |
| Number of characters | 0.002 | 9.43 | **0.99** | [-18.46,18.50] |
| **Random effects** | Variance | | *SD* | |
| Item (Intercept) | 7883.39 | | 88.78 | |
| Subject (Intercept) | 4733.93 | | 68.79 | |
| Frequency (log) | 437.73 | | 20.91 | |
| NA | 510.57 | | 22.58 | |
| AoA | 382.56 | | 19.56 | |
| VC | 554.97 | | 23.54 | |
| Familiarity | 440.59 | | 20.97 | |
| Residual | 0.06 | | 0.24 | |

CI: Confidence Intervals were calculated based on Wald method. Verb type is deviated coded here.