Norms for name agreement, familiarity, subjective frequency, and imageability for 348 object names in Tunisian Arabic

Mariem Boukadi[1] · Cirine Zouaidi[2] · Maximiliano A. Wilson[1]

Abstract Normative databases for pictorial stimuli are widely used in research on language processing in order to control for a number of psycholinguistic variables in the selected stimuli. Such resources are lacking for Arabic and its dialectal varieties. In the present study, we aimed to provide Tunisian Arabic (TA) normative data for 348 line drawings taken from Cycowicz, Friedman, Rothstein, and Snodgrass (1997), which include Snodgrass and Vanderwart's (1980) 260 pictures. Norms were collected for the following psycholinguistic variables: name agreement, familiarity, subjective frequency, and imageability. Word length data (in numbers of phonemes and syllables) are also listed in the database. We investigated the effects of these variables on word reading in TA. We found that word length and frequency were the best predictors of word-reading latencies in TA. Name agreement was also a significant predictor of word-reading latencies. A particularly interesting finding was that the semantic variables, imageability and familiarity, affected word-reading latencies in TA. Thus, it would seem that TA readers rely on semantics even when reading individual Arabic words that are transparent in terms of orthography-to-phonology mappings. This database represents a precious and much-needed psycholinguistic resource for researchers investigating language processing in Arabic-speaking populations.

It has long been established that standardized pictorial stimuli allow for more reliable comparisons between the results of different studies and for better control of psycholinguistic variables. As a result, their use has become common practice in experimental as well as clinical research on language. Indeed, the effects of several psycholinguistic variables on spoken and written word processing have been extensively documented among both healthy and language-impaired populations and in several languages, including English, French, Spanish, Italian, German, and Chinese (e.g., Alario et al., 2004; Barca, Burani, & Arduino, 2002; Barry, Morrison, & Ellis, 1997; Bates et al., 2003; Bonin, Boyer, Méot, Fayol, & Droit, 2004; Cortese & Schock, 2013; Cuetos, Ellis, & Alvarez, 1999). Therefore, it is important to carefully control for the effects of such factors when conducting psycholinguistic experiments investigating language production or comprehension processes.

Snodgrass and Vanderwart's (1980) pioneering set of 260 standardized pictures for American English has been extended to 400 pictures (Cycowicz, Friedman, Rothstein, & Snodgrass, 1997). Although the latter, extended set was originally developed for children, norms have been collected for adult speakers of different languages, including French (Alario & Ferrand, 1999), Italian (Nisi, Longoni, & Snodgrass, 2000), Greek (Dimitropoulou, Duñabeitia, Blitsas, & Carreiras, 2009), Spanish (Manoiloff, Artstein, Canavoso, Fernández, & Segui, 2010; Sanfeliu &

✉ Mariem Boukadi
mariem.boukadi@gmail.com

[1] Centre de recherche de l'Institut universitaire en santé mentale de Québec (CRIUSMQ), Université Laval, Quebec City, Canada

[2] Institut supérieur des langues de Tunis (ISLT), Université de Carthage, Tunis, Tunisia

Fernandez, 1996), Turkish (Raman, Raman, & Mertan, 2014), Japanese (Nishimoto, Miyawaki, Ueda, Une, & Takahashi, 2005), Persian (Bakhtiar, Nilipour, & Weekes, 2012), and Chinese (Weekes, Shu, Hao, Liu, & Tan, 2007). This picture set is one of the most widely used in research studies on adult picture naming as well as other tasks (e.g., object recognition, object decision, etc.). Several of these studies have shown that variables such as name agreement and familiarity (for which we provide norms in this study) are culturally specific and vary from one language community to another. This highlights the importance of obtaining norms for different languages and even distinct varieties of the same language (e.g., Argentine Spanish vs. the Spanish spoken in Spain).

The presence of norms for this specific set across many different languages makes it ideal for conducting cross-linguistic psycholinguistic studies. For example, Bates et al. (2003) investigated predictors of picture-naming latencies in seven languages, including Chinese, Bulgarian, and Hungarian, using a large set of 520 pictures that included 174 items from the Snodgrass and Vanderwart (1980) set. However, the fact that the set was originally developed for English may raise some concerns about its validity for other cultural contexts. Indeed, some items in the set may be familiar in one culture but not in another. For example, the *American football* and the *pretzel* are not familiar in Spanish-speaking settings (Manoiloff et al., 2010). With regard to the case of Tunisia, the sociolinguistic setting of the present study, its specific cultural profile makes the present picture set at least as appropriate for the Tunisian culture as it is for the French, Italian, Spanish, or Greek ones. Tunisia is a Mediterranean country whose strategic position (close to Europe, its main exchange partner) has given birth to a highly dynamic and multicultural society. Additionally, and also as a result of its geographical position, Tunisia shares many of the cultural values and traits common to neighboring Mediterranean societies (e.g., Italy or Greece). Therefore, we believe that this possible intercultural limitation does not undermine the usefulness of this set, nor does it outweigh its practicality, as long as the familiarity of the set as a whole remains high.

Psycholinguistic resources in Arabic for both pictorial and verbal stimuli are quite scarce, and no extensive normative database exists for this language. A few computerized databases containing information regarding word frequency are available for modern standard Arabic (MSA; e.g., Aralex: Boudelaa & Marslen-Wilson, 2010). However, the scope of their use is limited to the written variety of Arabic (i.e., MSA). The language situation in the Arab world is characterized by diglossia, a sociolinguistic condition in which two varieties of the same language are used by a speech community for different functions and contexts (Ferguson, 1959). Dialectal Arabic (DA) is the medium of oral communication, and MSA that of formal written communication, such as in mass

media (press, radio, and TV), textbooks, and official documents (Boudelaa & Marslen-Wilson, 2010, 2013; Daoud, 2001). Additionally, MSA and DA present some typological differences at the phonological, lexical, and morpho-syntactic levels (Boudelaa & Marslen-Wilson, 2013). DA itself is further subdivided into several, and sometimes mutually unintelligible, varieties across the Arab world, including Tunisian Arabic (TA), the variety spoken in Tunisia.

Another difference between MSA and DA (and more specifically, TA) is the manners of acquisition of these two varieties. While DA is acquired as a native language, MSA is acquired much later in a formal instruction context (viz. at school). In Tunisia, for example, TA is acquired as a first language, whereas instruction in MSA begins only at age six, when children start primary school. Concerns have been raised with regard to the impact of the difference in acquisition modes for the two varieties on the ways that they are processed during language production and comprehension (Boudelaa & Marslen-Wilson, 2013).

Therefore, research involving Arabic-speaking populations is in dire need of psycholinguistic databases for the different varieties of DA. Norms have recently been established for Levantine Arabic, one of the DA varieties spoken in the Middle East (Khwaileh, Body, & Herbert, 2014). However, the ratings were collected for a different and smaller set ($n = 186$ pictures) than the commonly used Snodgrass and Vanderwart (1980) pictorial database (e.g., Alario & Ferrand, 1999; Cycowicz et al., 1997; Dimitropoulou et al., 2009; Manoiloff et al., 2010; Nisi et al., 2000; Raman et al., 2014; Sanfeliu & Fernandez, 1996; Tsaparina, Bonin, & Méot, 2011). Additional norms are therefore needed in a spoken variety of Arabic for the extended and widely used Cycowicz et al. (1997) picture set, which includes Snodgrass and Vanderwart's (1980) 260 line drawings.

The language situation specific to each Arabic-speaking country is also an important factor to take into consideration. In Tunisia, for example, the language situation is a mixture of diglossia and societal bilingualism (Daoud, 2001). In addition to TA and MSA, the Tunisian sociolinguistic portrait is characterized by the marked presence of French in formal as well as informal written and spoken communication, and code-switching between TA and French is common in daily informal communication. TA itself is marked by numerous French lexical borrowings (e.g., /farʃitɑ/ in TA, from French *fourchette*). Recent years have also seen the rise of English, which is gaining influence in daily communication, especially among the youth and as the language of science (Daoud, 2001). Thus, we would expect culturally specific psycholinguistic variables to be influenced by and to reflect this specific language situation for TA.

The aim of the present study was to establish a normative database in TA for the 400 line drawings of objects taken from Cycowicz et al. (1997). Norms were collected for the name

agreement and familiarity of the pictures, as well as for the subjective frequency and imageability of their names. Values for word length (in numbers of phonemes and syllables) of the object names were also listed. A second aim was to study the influence of these newly obtained variables in a word-reading task by means of a regression design in TA.

Picture naming involves several processing stages. The first step is object recognition, which involves the activation of the visual or structural representation of the object, which in turn leads to the activation of the object's semantic representations. The next, and central, step in picture naming is lexical access. According to Levelt, Roelofs, and Meyer's (1999) influential model, lexical access takes place in two steps: lemma selection and phonological encoding. Levelt et al.'s model postulates three processing levels: conceptual or semantic representations, lemmas, and lexemes. Lemma selection involves the spreading of activation from the semantic representations of the object to be named to lemmas (nonphonological representations of words). The most highly activated lemma is selected. Phonological encoding consists in the activation and selection of the lexeme (or word phonological form) corresponding to the target lemma.

Each of the psycholinguistic variables for which we present norms in this database affects one (or more) of the abovementioned processing stages. In what follows, we describe each of these variables and identify the loci of their respective effects in the picture-naming process within Levelt et al.'s (1999) framework.

Name agreement (NA) refers to the degree of variability in the names given to the picture across participants. A picture that elicits the same name from most subjects is said to have high NA, and a picture that elicits several different names has low NA. This variable has been shown to be the most important predictor of latencies in picture naming (e.g., Alario et al., 2004; Ellis & Morrison, 1998; Gilhooly & Gilhooly, 1980; Lachman, Shaffer, & Hennrikus, 1974; Valente, Burki, & Laganaro, 2014; Vitkovitch & Tyrrell, 1995). Pictures that elicit different names take longer to be named because of the lexical competition that takes place between the different alternatives (Barry et al., 1997; Cuetos et al., 1999). Two possible loci of the NA effect have been identified, depending on the cause behind low NA. If low agreement is caused by the misidentification of pictures, then the locus is possibly at the level of structural encoding, namely in the object recognition stage. However, if the variance in NA is the result of the availability of various correct names for the same object, then low NA possibly exerts its influence at the level of lemma selection (Barry et al., 1997; Cuetos et al., 1999; Vitkovitch & Tyrrell, 1995). Pictures that have low NA will activate more lemmas than high-NA pictures, and thus will result in longer latencies for one of the possible candidates to be selected.

Familiarity (FAM) refers to how common an object is in the language speakers' realm of experience. FAM has been shown to influence naming speed among healthy and aphasic individuals, with pictures representing more familiar objects being named faster than those representing uncommon objects (Cuetos et al., 1999; Hirsh & Funnell, 1995; Kremin et al., 2001; Snodgrass & Yuditsky, 1996). Even though this effect has not been found consistently across all studies investigating the determinants of picture naming (Alario et al., 2004; Barry et al., 1997; Bonin, Chalard, Méot, & Fayol, 2002; Bonin, Peereman, Malardier, Méot, & Chalard, 2003; Dell'Acqua, Lotto, & Job, 2000; Valente et al., 2014), it is considered in the literature to be an important possible predictor of naming latencies that should be taken into account when conducting naming studies.

It has been suggested that FAM influences objects' recognition ease and speed, and that this variable has its effect at the level of the link between visual recognition and the conceptual level (Cuetos et al., 1999). Consistent with the semantic-level locus, Hirsh and Funnell (1995) found familiarity effects in patients suffering from the semantic variant of primary progressive aphasia, a disease in which semantic knowledge is impaired. They hypothesized that the structural/visual representations of highly familiar objects activate more quickly and easily their corresponding semantic representations than do low-familiarity objects.

Imageability (IMA) refers to the ease with which a given word evokes a mental image. This semantic variable influences performance on a number of tasks involving the naming or recognition of words, because the semantic representations of picture names that easily evoke a mental image are accessed more quickly (Ellis & Morrison, 1998). Highly imageable words elicit faster reaction times and fewer errors than do low-imageability words (Alario et al., 2004; Bonin et al., 2002; Cortese & Schock, 2013; Ellis & Morrison, 1998; but see Barry et al., 1997; Bonin et al., 2003; and Morrison, Ellis, & Quinlan, 1992, who failed to find an imageability effect in picture naming). IMA has been found to significantly affect naming latencies even when the stimulus set consisted solely of pictures representing imageable concrete objects (Alario et al., 2004).

Plaut and Shallice (1993) found imageability effects among patients with deep dyslexia, a type of patient that reads using a semantic pathway only. They concluded that imageability was related to the number of semantic features that form a concept. Accordingly, the concepts of highly imageable words are linked to many more semantic features than the concepts of low-imageability words. Imageability effects have also been found among healthy participants in semantic tasks such as word association (e.g., de Groot, 1989). These findings support the idea that the locus of the imageability effect is at the conceptual level of processing. Thus, activation of the semantic

representation of an object (i.e., its lexical concept) is faster for high- than for low-imageability words (Alario et al., 2004; Ellis & Morrison, 1998).

Subjective frequency (FREQ) refers to how often a word is used or heard in daily communication, as subjectively estimated by native speakers of a given language. Word frequency is estimated in two ways: objective or subjective. *Objective* word frequency refers to the sum of occurrences of a word in textual corpora, whereas the *subjective* frequency of a given word is estimated by the speakers of the language on a Likert scale, usually ranging from 1 to 7 (Desrochers & Thompson, 2009). Objective and subjective frequency measures have been shown to be strongly associated and to be robust predictors of the ease and speed of response in different types of tasks (Balota, Pilotti, & Cortese, 2001; Desrochers, Liceras, Fernández-Fuertes, & Thompson, 2010). In some studies, subjective frequency estimates were found to be a better predictor of visual and auditory word processing than objective frequency counts (Balota et al., 2001; Connine, Mullennix, Shernoff, & Yelen, 1990).

Oldfield and Wingfield (1965) were the first to report an effect of lexical frequency on naming speed, with words that are used or heard more frequently being more easily accessed and retrieved than low-frequency words. However, this finding was later challenged and attributed to the high correlation between age of acquisition (AoA), the age at which words are learned, and lexical frequency (Carroll & White, 1973). Studies controlling for the effects of AoA have failed to find an effect of word frequency on naming latencies, thus showing that the supposed word frequency effects found in many studies that did not control for AoA were in fact due to this variable (Bonin et al., 2002; Carroll & White, 1973; Dell'Acqua et al., 2000; Gilhooly & Gilhooly, 1980; Morrison et al., 1992). However, other studies have found that both frequency and AoA independently predicted latencies in different visual word-processing tasks (Cortese & Schock, 2013; Wilson, Cuetos, Davies, & Burani, 2013). Other studies that have used subjective ratings of word frequency (e.g., Lachman, 1973; Lachman et al., 1974), as well as more recent studies using larger picture sets and more accurate frequency measures (e.g., Alario et al., 2004; Barry et al., 1997; Ellis & Morrison, 1998; Snodgrass & Yuditsky, 1996), have shown effects of word frequency on naming latencies even when AoA was controlled.

A phonological locus for the word frequency effect has been proposed in several studies (Barry et al., 1997; Jescheniak & Levelt, 1994; La Heij, Puerta-Melguizo, van Oostrum, & Starreveld, 1999; Levelt et al., 1999; McCann & Besner, 1987; Wheeldon & Monsell, 1992). In other words, the frequency effect has its locus at the level of phonological encoding, because it affects the time taken for the activation and selection of the target lexeme. This effect has been explained by Jescheniak and Levelt

(1994) in terms of the lexeme's activation thresholds, which are low for high-frequency words and high for low-frequency words.

Word length (WL) refers to how long a word is, in numbers of letters, phonemes (phWL), and syllables (syllWL). These variables have been shown to influence reaction times in several visual word recognition tasks (see Barton, Hanif, Eklinder Björnström, & Hills, 2014, for a review). They also interact with frequency estimates, since highly frequent words tend to be shorter (Dell'Acqua et al., 2000). However, the effect of these variables on picture naming is not consistent across studies. Although some studies have shown shorter naming latencies for shorter words (e.g., Cuetos et al., 1999; Klapp, Anderson, & Berrian, 1973; Roelofs, 2002; Santiago, MacKay, Palma, & Rho, 2000), others have not (Alario et al., 2004; Bachoud-Lévi, Dupoux, Cohen, & Mehler, 1998; Damian, Bowers, Stadthagen-Gonzalez, & Spalek, 2010; Dell'Acqua et al., 2000; Snodgrass & Yuditsky, 1996; Valente et al., 2014).

It is generally assumed that these variables affect the phonological encoding stage of picture naming (Alario et al., 2004; Valente et al., 2014). In Levelt et al.'s (1999) model, this process consists in the retrieval and ordering of phonemes and their serial insertion into syllable frames. Thus, longer words take longer to be named than short ones.

The normative study

Method

Participants A total of 100 native speakers of TA participated in this study (mean education 16 years; mean age 24 years old, age range 18–35 years; 51 % females). They studied or were employed at the University of Carthage in Tunis, Tunisia. All of the participants had normal or corrected-to-normal vision and no history of language, learning, or attention difficulties. They had all acquired TA as their first language. All participants learned MSA at school around the age of 6 years old and spoke different TA dialects. They were also proficient in French, which they had learned as a second language in primary school from around the age of 8 and continued to receive formal instruction in until the end of high school. Our participants had TA as their dominant language (i.e., the language most frequently used), and used French in everyday communication in the form of frequently used lexical borrowings and idiomatic expressions blended in TA speech (i.e., code-switching and mixing).

Participants were randomly assigned to each one of the four tasks (*n* = 25 in each subgroup of the sample), so that each subgroup participated in only one of the tasks. The age means and standard deviations of the participants in each of the four subgroups were as follows: *M* = 25, *SD* = 3, in the NA group;

$M = 23$, $SD = 4$, in the FAM group; $M = 25$, $SD = 6$, in the IMA group; $M = 23$, $SD = 4$, in the FREQ group. The numbers of females in each of the four subgroups were as follows: 20 in the NA group, seven in the FAM group, 15 in the IMA group, and nine in the FREQ group. All groups were matched by age, $F(3, 96) = 1.68$, $p = .18$, and by the proportions of female and male participants in each group, $\chi^2(3) = 1.53$, $p = .68$. All participants had achieved the compulsory minimum education level in Tunisia (10 years).

Materials Four hundred black-and-white line drawings taken from Cycowicz et al. (1997) were used in the NA and FAM tasks. This set consisted of the 260 pictures in Snodgrass and Vanderwart (1980) and 140 additional line drawings constructed by Cycowicz et al. (1997).

For the FREQ and IMA tasks, ratings were collected for 348 names assigned to the pictures (listed under the Intended Names column in Appendix A of the supplemental materials). This list consisted of TA words, French loan words, and MSA words that are used in everyday oral communication in the Tunisian context.

Fifty-two pictures to which an intended name was not assigned because they have none in TA and/or are usually referred to with their French name by Tunisian speakers were excluded from the original set of 400 stimuli. For example, the modal name (i.e., the most common name given by participants) of *skirt* in TA is the French word *jupe*. The MSA names of those objects were not included because they are not used by Tunisian speakers in everyday oral communication.

These stringent exclusion criteria are supported by the data obtained in the NA task presented here (see the Results section for further details). Indeed, the modal names for the 52 finally excluded stimuli either were in French, did not correspond to the object represented by the picture, or were homonymous to the names of objects in the rated 348-word list. The final database consists of 348 pictures and their names (Appendix A). Appendix B (also in the supplement) contains the excluded 52 pictures and their related NA and FAM information, which researchers may find informative and use at their discretion.

Procedure We used a computerized procedure in each of the four tasks. Some normative studies, especially earlier ones, have used questionnaires to collect norms for the variables presented in this study (e.g., Alario & Ferrand, 1999; Snodgrass & Vanderwart, 1980). However such a method makes it difficult and more time-consuming to control the effects of order of presentation, among others. In several more recent studies, researchers have chosen to use stimulus presentation software to collect normative data for NA (e.g., Bates et al., 2003; Cortese & Fugett, 2004; Dell'Acqua et al., 2000; Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005), as well as for FREQ and IMA (e.g., Desrochers &

Thompson, 2009). This method allowed the homogenization of the data collection process (i.e., each stimulus was rated within the same time limit), as well as proper and effective randomization of the stimuli in each task in order to control for order-of-presentation and fatigue effects.

One picture-naming task (NA) and three rating tasks (FAM, FREQ, and IMA) were run on a PC using the DMDX software (Forster & Forster, 2003). Each subgroup of participants ($n = 25$) completed each task in one experimental session. The participants in one task did not take part in the others. The stimuli were divided among four blocks, and their order of administration was counterbalanced across participants. Within each block, items for the NA and FAM tasks ($n = 100$) and for the IMA and FREQ tasks ($n = 87$) were presented in a different random order for each participant.

Similar procedures were followed in all four tasks. Participants were tested individually in a quiet room and were seated in front of a PC monitor. At the beginning of each task, instructions in TA (adapted from Alario & Ferrand, 1999, for FAM and NA, and from Desrochers & Thompson, 2009, for FREQ and IMA) appeared on the screen and were read aloud by the experimenter. Six practice items were administered before the experimental trials. In the rating tasks, the scale was presented before the practice set and on top of each picture during the experiment. Participants used the numeric keys on the keyboard to enter their ratings. Each experimental trial ran as follows: A fixation point was presented at the center of the screen for 400 ms, followed immediately by the stimulus (either a word in TA or a picture) presented at the center of the screen. The stimulus remained on the screen for 6,000 ms in the rating tasks, and for 4,000 ms in the picture-naming task. Opportunities for breaks were provided at the end of each block.

In the NA task, participants were instructed to orally name each of the 400 drawings with the first name that came to their mind. They were told that a name could consist of more than one word. If they could not give the name of the picture, they were asked to give one of these justifications in TA: "I don't know the object" or "I don't know the name." Vocal responses were recorded with a microphone connected to the computer and with the DMDX software (Forster & Forster, 2003).

In the FAM task, participants were asked to rate the familiarity of 400 objects represented by the pictures using a 5-point scale adapted from Alario and Ferrand (1999), where 1 = *very unfamiliar* and 5 = *very familiar* pictures. Participants were told that familiar objects were those they often encounter in their daily life, whereas unfamiliar objects were unusual and rarely encountered.

In the FREQ task, participants were asked to rate the frequency of 348 names of the pictures (listed under the Intended Name column in Appendix A) using a 7-point scale (adopted from Balota et al., 2001), where 1 = words

they *never encountered* and 7 = words they *encountered several times a day*. Subjective frequency was defined as the degree to which participants saw or came across a word in their daily life.

In the IMA task, participants rated the imageability of 348 picture names—namely the ease with which a given word elicited a mental image—on a 7-point scale where 1 = a word *imaged with difficulty* and 7 = a word *easily and quickly imageable* (Desrochers & Thompson, 2009). Participants were told not to worry about how often they used a given number on the scale, as long as it faithfully represented their impression.

Additionally, we determined word length in number of spoken syllables; namely, sequences of phonemes served as the basis of syllabification of TA words. The ordering of these sequences into syllables was carried out following the admitted syllable structures in TA, namely consonant–vowel (CV), CVV, CVC, CCV, CCVC, CCVCC, CCVV, and CCVVC (Hamdi, Ghazali, & Barkat-Defradas, 2005).
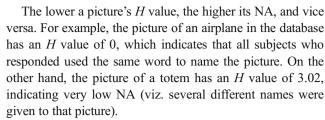
Results and discussion

A summary of the NA and rating data obtained from our sample of TA-speaking subjects is presented in Appendix A. The database contains the following information for each picture: (1) the number assigned to each picture, in decreasing order of NA% (first column); (2) the picture's name in English, as in Cycowicz et al.'s (1997) database (second column); (3) the picture's intended and modal names transcribed in Arabic script (third and fourth columns, respectively); (4) the modal name's English translation (fifth column); (5) two NA measures—the *H* statistic (Snodgrass & Vanderwart; 1980) and the percentage of participants giving the most common name in TA (sixth and seventh columns, respectively); (6) the means and standard deviations for FAM, FREQ, and IMA (subsequent columns); and (7) WL (in numbers of phonemes [phWL] and syllables [syllWL]) as counted by the researchers, since this information is not available for TA (the two final columns). The different alternative names given to each picture in the NA task are listed in Appendix C of the supplemental materials.

The information statistic, *H*, was computed using the following formula developed by Snodgrass and Vanderwart (1980):

$$H = \sum_{i=1}^{k} p_i \log_2\left(\frac{1}{p_i}\right),$$

where *k* refers to the number of names given to the picture and p_i indicates the proportion of participants who gave the name. Naming failures ("I don't know the name," "I don't know the object," and no responses) were taken into account when computing the NA percentages but were eliminated when computing the *H* statistic.

The lower a picture's *H* value, the higher its NA, and vice versa. For example, the picture of an airplane in the database has an *H* value of 0, which indicates that all subjects who responded used the same word to name the picture. On the other hand, the picture of a totem has an *H* value of 3.02, indicating very low NA (viz. several different names were given to that picture).

According to Snodgrass and Vanderwart (1980), the *H* statistic is a more reliable measure of the distribution of picture names than is the NA percentage. For example, a picture could have 92 % NA but an *H* value of 0 (i.e., perfect NA) if all of the subjects who gave a response used the same name. However, the percentage NA is also important as a complementary measure to the *H* statistic, since it gives us more detailed information about which items elicited a response from every single subject in the sample, and which ones caused naming failures.

Description and analysis of the normative data Table 1 presents the summary statistics for all of the variables in the database (NA, FAM, FREQ, IMA, and WL) and for the 348 stimuli. Both measures of NA (*H* and %) seem to indicate an overall low level of NA, with *M* = 1.13 and *SD* = 0.84 for the *H* statistic and *M* = 62.43 % and *SD* = 27.74 % for the percentage measure, which indicates great variability in the picture names given by participants. This may be partly accounted for in terms of regional dialect variations across participants. TA's regional varieties are mutually intelligible but present a few differences that include object names. Therefore, one object may have a different dominant name from one speaker's region to another (e.g., a faucet is named /sabɛlɑ/ in the capital city Tunis and /ʃiʃmɑ/ in other Tunisian regions). It is also noteworthy that some of the items showing an *H* value of 0 had a percentage slightly below 100 (e.g., *barrel* has an *H* value of 0 but 72 % NA). This is due to the fact that some pictures had naming failures (mostly no responses).

Three pictures had 0 % NA; namely, the participants' responses were all different, and no single most common name could be identified. One of these pictures (*fire hydrant*) failed to elicit a response from any of the participants, which can be explained by the fact that this particular object has no name in TA and is unfamiliar in a Tunisian context (*M* = 2.46, *SD* = 1.35). This item is included in the set of 52 excluded pictures. Seventeen pictures in the set were misidentified (e.g., the modal name for the picture of a thimble was صطل [*bucket*]), due to the unfamiliarity of these objects in a Tunisian context (*M* = 2.63, *SD* = 1.17). Nine of these 17 pictures were in the list of 52 pictures excluded from the FREQ and IMA tasks (and the final 348-item database). Additionally, 42 pictures were given French names by the participants (e.g., the modal name for the picture of a screwdriver was its French equivalent, *tournevis*). Eighteen of these were in the excluded 52-picture set, and the rest have existing, albeit less frequent,

Table 1 Summary statistics for all TA variables

	Mean (M)	Standard Deviation (SD)	Median	Asymmetry	Kurtosis	Range	Min	Max	25th Percentile	75th Percentile	Interquartile Range
NA H	1.13	0.84	1.12	0.32	–0.89	3.32	0.00	3.32	0.40	1.77	1.37
NA%	62.43	27.74	64.00	–0.25	–1.18	100.00	0.00	100.00	40.00	88.00	48.00
FAM	3.59	0.70	3.60	–0.20	–0.74	3.25	1.67	4.92	3.08	4.12	1.04
IMA	5.73	0.84	5.98	–1.93	4.36	4.95	1.80	6.75	5.47	6.27	0.81
FREQ	3.97	1.17	3.94	0.06	–0.67	5.40	1.44	6.84	3.06	4.84	1.78
phWL	5.83	2.05	6.00	1.34	2.16	11.00	3.00	14.00	4.00	7.00	3.00
syllWL	2.20	0.88	2.00	0.67	0.45	4.00	1.00	5.00	2.00	3.00	1.00

NA H, name agreement information statistic; NA%, name agreement as a percentage; FAM, familiarity; IMA, imageability; FREQ, subjective frequency; phWL, word length in number of phonemes; syllWL, word length in number of syllables

names in TA ($M = 3.79$, $SD = 1.60$). For example, the modal name of the picture of a hat was the French word *chapeau*, whereas the intended TA name for this object was طَرْبُوشَة. This reflects the marked interaction of French with TA in Tunisia (Daoud, 2001).

The fact that only 17 pictures were misidentified in the set, versus 42 pictures that were correctly identified but given French names, indicates that the low NA in this database was not due to unfamiliarity with the objects among the Tunisian sample of participants, but rather was a direct result of the variability inherent to dialects, a variability heightened by the language contact between TA and French. However, it is noteworthy that despite the overall low NA, the database lists 162 items with an average to high NA ($H = 1$ and lower), and 52 with a perfect NA ($H = 0$).

The results of the NA task support two methodological choices: (1) the exclusion of the 52 items (listed in Appendix B) from the word-rating tasks, and (2) collecting the FREQ and IMA ratings for the intended rather than for the modal names. As we explained above, 4.3 % of the pictures' modal names reflected misidentifications of the objects represented by the pictures, and 10.5 % were in French. Therefore, in order to obtain ratings for as many TA words corresponding to the pictures as possible, we chose simply to translate the English names in Cycowicz et al. (1997) into their equivalent TA names. The ratings of the FAM and FREQ tasks indicated that the pictures were highly familiar to TA subjects ($M = 3.59$, $SD = 0.72$) and that their names were partially familiar ($M = 3.98$, $SD = 1.17$). The IMA task data, on the other hand, show that most names easily evoked a mental image to participants ($M = 5.73$, $SD = 0.84$), which is not surprising, seeing that all names in the set represent concrete objects.

Correlations among TA variables Correlational analyses were conducted among all of the TA variables (NA% and H, FAM, IMA, and FREQ). Three items were removed from the percentage NA data (the ones that had

0 % NA) and one from the NA/H data (*fire hydrant*, which elicited no names).

The correlation matrix is presented in Table 2. Significant correlations were found among all of the abovementioned variables (all $ps < .01$). As expected, and as had been found in previous studies (Alario & Ferrand, 1999; Manoiloff et al., 2010), we observed a strong negative correlation ($r = –.91$) between the two measures of NA, NA/H and NA%. A strong positive correlation was also found between FAM and FREQ ($r = .73$). The weakest of the correlations among our five main variables (NA% and H, FAM, IMA, and FREQ) was between FREQ and NA/H ($r = –.35$). Additionally, moderate correlations were found among the rest of the variables.

The strong relationship found between familiarity of the pictures and their names in TA seems to indicate that the names of the most familiar objects are also the most frequently used and heard in daily communication.

Correlations were also performed between all four TA variables and WL (both phWL and syllWL). Most correlations were significant at $p < .01$ (phWL and FAM were significant at $p < .05$), except for the correlation between FAM and syllWL, which

Table 2 Correlations between all TA variables (used as predictors in the reading aloud task)

	NA H	NA%	FAM	IMA	FREQ	phWL	syllWL	RTs
NA H	1							
NA%	–.91**	1						
FAM	–.39**	.52**	1					
IMA	–.40**	.54**	.53**	1				
FREQ	–.35**	.49**	.73**	.69**	1			
phWL	.21**	–.24**	–.13*	–.22**	–.33**	1		
syllWL	.21**	–.22**	–.100	–.15*	–.25**	.88**	1	
RTs	.39**	–.43**	–.25**	–.49**	–.47**	.48**	.46**	1

NA H, name agreement information statistic; NA%, name agreement as a percentage; FAM, familiarity; IMA, imageability; FREQ, subjective frequency; phWL, word length in number of phonemes; syllWL, word length in number of syllables; RTs, reaction times. * $p < .05$. ** $p < .01$.

was marginally significant ($p = .06$). The strongest correlation was found between phWL and syllWL ($r = .88$) and the weakest between IMA and syllWL ($r = -.15$). All other correlations were weak, and usually negative, except for the correlations between both measures of WL and NA/H, which were positive.

The significant and negative correlations found between WL (both phWL and syllWL) and both IMA and FREQ, albeit weak, suggest that most frequent words are also shorter and evoke a mental image more quickly. The significant and positive correlations between NA/H and WL (phWL and syllWL) indicate that longer words are more likely to have other possible names. The significant and negative correlation between the NA% and WL (phWL and syllWL) variables suggests that the longer the word, the more difficult it is to name.
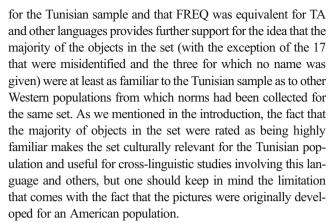
TA versus English, French, and Spanish norms Table 3 presents descriptive data for NA, FAM, IMA, and FREQ in TA, French, English, and Spanish. Comparisons and correlations between TA and both the French and Spanish norms were carried out for NA and FAM (taken from Alario & Ferrand, 1999, and Manoiloff et al., 2010, respectively) for the 348 pictures shared between the three studies. Additionally, we carried comparisons and correlations between the present NA and FAM norms and the English ones for the 239 pictures in common. Seeing that FREQ and IMA ratings were not available for the whole set, we extracted the stimuli for which norms were available in French, Spanish, and English (see Table 3 for details).

From a descriptive point of view, the most important differences were between the two measures of NA in TA and the three other languages: The NA/H value was much higher and NA% much lower in TA than in English, French, and Spanish. With respect to FAM, the TA ratings were higher than the French ones. However, there were no remarkable differences between the TA and English FAM ratings. Overall, pictures were rated as being more familiar to the Tunisian sample. There were no differences of note between the TA ratings and those in other languages for FREQ and IMA. The fact that FAM was higher

for the Tunisian sample and that FREQ was equivalent for TA and other languages provides further support for the idea that the majority of the objects in the set (with the exception of the 17 that were misidentified and the three for which no name was given) were at least as familiar to the Tunisian sample as to other Western populations from which norms had been collected for the same set. As we mentioned in the introduction, the fact that the majority of objects in the set were rated as being highly familiar makes the set culturally relevant for the Tunisian population and useful for cross-linguistic studies involving this language and others, but one should keep in mind the limitation that comes with the fact that the pictures were originally developed for an American population.

The correlation matrix between the ratings collected for TA and English, French, and Spanish norms is presented in Table 4. Significant (at the .01 and .05 levels) and positive correlations were found between the norms in TA and other languages, except for IMA in Spanish ($r = .09$). The strongest correlations were found between TA and both French and English norms of FAM ($rs = .73$ and .81, respectively). All other correlations were weak to moderate.

The weak correlations found between TA and the French, English, and Spanish measures of NA, as well as the comparison between the descriptive data for this variable in all four languages (TA, French, English, and Spanish), suggest that it was much more difficult to generate a single most common name for TA speakers than for English, French, or Spanish ones. However, it is noteworthy that other normative studies using the Snodgrass and Vanderwart (1980) picture set have had low NA ratings comparable to those found in TA (Dimitropoulou et al., 2009; Nishimoto et al., 2005; Tsaparina et al., 2011; Weekes et al., 2007). For example, Weekes et al.'s normative database for Chinese has an H mean of 0.97 and an

Table 3 Means (M) and standard deviations (SD) for all variables in TA, French, English, and Spanish

	TA		French		English		Spanish	
	M	SD	M	SD	M	SD	M	SD
NA H	1.13	0.84	0.35	0.43	0.56	0.53	0.71	0.62
NA%	62.43	27.74	85	21	86	14	81	21
FAM	3.59	0.70	2.70	1.21	3.29	0.96	2.84	1.06
IMA	5.76	0.80	6.32	0.87	5.95	0.33	6.08	0.51
FREQ	4.05	1.17	3.90	1.27	5.38	0.60	5.77	0.90

NA H, name agreement information statistic; NA%, name agreement as a percentage; FAM, familiarity; IMA, imageability; FREQ, subjective frequency

Table 4 Correlations between TA and French, English, and Spanish norms for NA, FAM, IMA, and FREQ

	French	English	Spanish
NA H	.30**	.41**	.15**
NA%	.32**	.40**	.15**
FAM	.73**	.81**	.27**
IMA	.12*	.18**	.09
FREQ	.21**	.66**	.48**

NA H, name agreement information statistic; NA%, name agreement as a percentage; FAM, familiarity; IMA, imageability; FREQ, subjective frequency. For NA and FAM, the comparisons between TA and both Spanish (from Manoiloff et al., 2010) and French (from Alario & Ferrand, 1999) norms are for 348 pictures, and those between TA and English norms are for 239 pictures in common with Snodgrass and Vanderwart's (1980) set. For IMA and FREQ, comparisons were carried out on 320 words for French (Desrochers & Thompson, 2009), 189 (Izura, Hernández-Muñoz, & Ellis, 2005) and 193 words (Desrochers et al., 2010) for Spanish, and 199 (Stadthagen-Gonzales & Davis, 2006) and 203 (Balota et al., 2001) words for English. $^{*}p < .05$. $^{**}p < .01$.

NA percentage of 65 %, and it lists only 15 pictures with perfect NA (vs. 52 in TA). Similarly, Tsaparina et al.'s Russian database has an H mean of 0.82 and an NA percentage of 80 %, and it lists only 35 pictures with a perfect NA. Thus, the relatively low NA values found in our study are by no means atypical.

Furthermore, we believe that it is useful to count on the entire database, rather than restricting it to high-NA items only (which would mean that we would only keep items that had 75 % NA and higher). Similar databases in other languages have not eliminated items with very low NA. For example, NA% ranged from 8 % to 100 % in the Spanish database of Sanfeliu and Fernandez (1996), and in their Chinese database, Weekes et al. (2007) listed items with an NA of 4 %. Our main objective was to provide a psycholinguistic database for TA that was as large as possible. We leave it to researchers to choose whether to control NA by matching their stimuli on this variable or by simply selecting only high-NA items. We also believe that researchers will benefit from a large database in which the ranges of values for all variables are wide (seeing the correlation of NA% with the other variables in our study, keeping only high-NA items would also limit the range of the other variables to high values). For example, some researchers might want to explore the effect of high versus low frequency in word production or recognition studies in TA, or to use these psycholinguistic variables for tasks such as object decision, word reading, or lexical decision. For the sake of practicality, we ordered the items in the database from higher to lower NA in order to make the use of the database easier and to facilitate the stimulus selection process for researchers.

The association between TA and other languages for FAM and FREQ seems to indicate that pictures and their names are equally familiar for Tunisian speakers and speakers of other languages. IMA and NA seem to be the variables most influenced by cultural context and language in our TA database, since they both present the weakest correlations with the norms in the other three languages. In other words, it seems that the ability to generate names for the objects represented by the pictures (i.e., NA) or mental images for the names of the objects (i.e., IMA) highly depends on language. This is in line with similar comparisons performed in previous normative studies, in which NA has been shown to be most affected by cultural differences (Alario & Ferrand, 1999; Dell'Acqua et al., 2000; Manoiloff et al., 2010; Sanfeliu & Fernandez, 1996).

The influence of the new variables in reading aloud in Tunisian Arabic

We obtained values for seven key psycholinguistic variables for a large set of object names in TA. Seeing that this is the first psycholinguistic database in TA, we wished to provide researchers with a glimpse of the possible effects of these variables on language processing in TA. Picture-naming studies usually involve a familiarization phase preceding the experiment itself, in which participants are shown the word with which to name each picture (Alario et al., 2004). This is done in order to limit the variability in the names given to the pictures across participants (which might arise for pictures with imperfect name agreement). Such a time- and resource-consuming picture-naming procedure for the whole 348-item set was beyond the scope of the present normative study. That is why we chose a simple task that does not require a familiarization phase—that is, reading aloud. Thus, the aim of this experiment was to explore the influence of these key psycholinguistic variables by means of reading aloud in TA.

The seven variables presented in the previous section can be divided into word form variables, such as word length in phonemes and syllables (Barca et al., 2002; Davies, Barbón, & Cuetos, 2013); lexical phonological or lexeme-level variables, as the H statistic for name agreement, name agreement as a percentage, and subjective frequency (Barry et al., 1997; Cuetos et al., 1999); and semantic variables, as familiarity and imageability (Cortese & Schock, 2013; Yap & Balota, 2009). Current computational models of reading aloud predict the effects of word length and lexical phonological variables (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2010), as well as the influence of semantic variables on reading aloud (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989).

Consistently, many studies using a regression design have shown the independent contributions of these variables to reading-aloud latencies. For instance, word length effects in terms of numbers of phonemes, letters, and syllables have been shown to be among the best predictors of word-reading latencies (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Barca et al., 2002; Burani, Arduino, & Barca, 2007; Cortese & Khanna, 2007; Davies, Wilson, Cuetos, & Burani, 2014; Yap & Balota, 2009). The contribution of lexical variables such as subjective frequency (Balota et al., 2004; Cortese & Khanna, 2007) has also been shown in reading aloud. In the same way, semantic effects such as familiarity (Cuetos & Barbón, 2006) and imageability (Balota et al., 2004; Cortese & Schock, 2013; Cuetos & Barbón, 2006) have been shown to affect reading latencies. However, to the best of our knowledge, no study has addressed reading aloud with a regression design in Arabic.

Method

Participants Twenty-five adult native TA speakers participated in this study (mean age 27 years, education 19 years; 52 % females). They were recruited at Université Laval in Québec, Canada. All of the participants had normal or corrected-to-normal vision and no history of language, learning, or attention difficulties. They also had a language profile
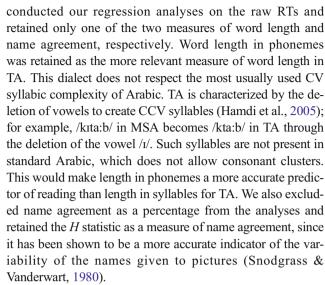
similar to that of the participants in the normative study and had not spent more than 3 years in Canada.

Task A word-reading task was used in which participants were asked to name the 348 TA words listed in our database. All of the words were vowelized; that is, we placed the diacritic marks indicating vowels above each word's letters, to ensure that participants read the words correctly.

Procedure Participants were tested individually in a sound-proof room and were instructed to name the words in TA as quickly and accurately as possible. They completed the task in one experimental session. The stimuli were divided into four blocks ($n = 87$ words each), and their order of administration was counterbalanced across participants. Within each block, words were presented in a different random order for each participant.

The DMDX software (Forster & Forster, 2003) was used to present the stimuli and record the response onset by means of a headset with a microphone. Naming latencies were measured from word onset until the vocal response. Participants were seated in front of a computer monitor. At the beginning of the task, instructions in TA appeared on the screen and were read aloud by the experimenter. Six practice items were administered before the experimental trials. Each trial ran as follows: A fixation point was presented at the center of the screen for 400 ms, followed immediately by the word presented at the center of the screen. The stimulus remained on the screen for 1,500 ms, followed by a blank screen that lasted for 1,000 ms. Opportunities for breaks were provided at the end of each block. All RTs were extracted from the recorded responses using the CheckVocal program (Protopapas, 2007).

Data analysis The bottom row of Table 2 shows the correlations between the key variables and RTs. Analysis of the correlation matrix for the raw values of the variables revealed the presence of two coefficients greater than .75, for the correlations between word length in phonemes and word length in syllables and between the H statistic for name agreement and name agreement as a percentage. These coefficients indicated a high level of multicollinearity (Cohen, Cohen, West, & Aiken, 2003). In order to deal with collinearity, we transformed to $\log_{10}(x + 1)$ all of the variables and RTs. The log transformation reduced the coefficient between the H statistic for name agreement and name agreement as a percentage, $r(348) = .680$, $p < .001$; however, the coefficient between word length in phonemes and word length in syllables continued to be greater than .75, $r(348) = .872$, $p < .001$. We attempted further to deal with this collinearity by centering the variables (i.e., subtracting the mean of each variable from its values), following Davies et al. (2014), but this procedure also failed to reduce the collinearity between word length in phonemes and word length in syllables. Consequently, we

conducted our regression analyses on the raw RTs and retained only one of the two measures of word length and name agreement, respectively. Word length in phonemes was retained as the more relevant measure of word length in TA. This dialect does not respect the most usually used CV syllabic complexity of Arabic. TA is characterized by the deletion of vowels to create CCV syllables (Hamdi et al., 2005); for example, /kɪta:b/ in MSA becomes /kta:b/ in TA through the deletion of the vowel /ɪ/. Such syllables are not present in standard Arabic, which does not allow consonant clusters. This would make length in phonemes a more accurate predictor of reading than length in syllables for TA. We also excluded name agreement as a percentage from the analyses and retained the H statistic as a measure of name agreement, since it has been shown to be a more accurate indicator of the variability of the names given to pictures (Snodgrass & Vanderwart, 1980).

Following Cortese and Schock (2013) and Yap and Balota (2009), the predictor variables were grouped. They were then entered in the regression models in three different steps. Step 1 included 16 dichotomous variables coding for the features of the initial phonemes in TA: voiced, alveolar, postalveolar, pharyngeal, palatal, dental, trill, glottal, uvular, pharyngealized, labiodental, approximant, plosive, velar, bilabial, and nasal. This was done to avoid the possible bias related to voice-key triggering and articulatory differences among the different initial phonemes (Cortese & Schock, 2013). Step 2 included all of the newly obtained variables for TA except one (i.e., four out of five): word length in phonemes, subjective frequency, the H statistic for name agreement, familiarity, and imageability. Step 3 included each one of the five key variables separately. We did this in order to study the specific contribution of each key variable above and beyond that of the other variables. Thus, we ran five regression models, one for each key variable included in Step 3.

Results

The means and standard deviations of the variables are shown in Table 1. Mispronunciation errors (1.43 %) and no responses (5.80 %) were removed from the analysis of RTs. This resulted in the exclusion of 7.23 % of the total data. The mean RT for the whole set of words was 748.61 ms ($SD = 117.39$). Five hierarchical regressions with three steps each were conducted with raw RTs as the dependent variable. Table 5 shows the results of these analyses. After controlling for the effects of phonological onset and the other four new variables, word length in phonemes, $\beta = .342$, $p < .001$; the H statistic for name agreement, $\beta = .163$, $p < .001$; subjective frequency, $\beta = -.257$, $p < .001$; familiarity, $\beta = .173$, $p < .01$; and imageability, $\beta = -.274$, $p < .001$, were all significant predictors of reading aloud.

Table 5 Standardized βs, R^2s, and ΔR^2s for the regression analyses of the word-reading task

Step		phWL	H	FREQ	FAM	IMA
Step 1	Phonological onset variables					
	R^2		.095			
	ΔR^2		.095[**]			
Step 2	All variables but one					
	phWL	n/a	.364[***]	.390[***]	.366[***]	.333[***]
	H	.214[***]	n/a	.161[***]	.141[**]	.220[***]
	FREQ	−.432[***]	−.254[***]	n/a	−.132[*]	−.427[***]
	FAM	.262[***]	.137[*]	.045	n/a	.171[**]
	IMA	−.254[***]	−.327[***]	−.373[***]	−.274[***]	n/a
	R^2	.435	.509	.509	.515	.493
	ΔR^2	.340[***]	.414[***]	.414[***]	.420[***]	.398[***]
Step 3	Single key variable					
	Variable	.342[***]	.163[***]	−.257[***]	.173[**]	−.274[***]
	R^2	.528	.528	.528	.528	.528
	ΔR^2	.093[***]	.019[***]	.019[***]	.013[**]	.035[***]

phWL, word length in phonemes; H, H statistic for name agreement; FREQ, subjective frequency; FAM, familiarity; IMA, imageability. Columns refer to the different regression models, and their titles show the variables entered in the last step of the models. *Single key variable* refers to the variable entered in the last step of the regression models. ΔR^2 is the incremental increase in the model R^2 that results from the addition of a predictor or set of predictors in a new step of the model. [*] $p < .05$. [**] $p < .01$. [***] $p < .001$.

Discussion

In the present study, we aimed at investigating the effect of the newly obtained variables on word reading in TA. As was predicted by theoretical models (Coltheart et al., 2001; Perry et al., 2010; Plaut et al., 1996), several studies on reading aloud using regression designs have shown that word length, lexical, and semantic variables all affected reading-aloud latencies. Consistent with previous similar studies in other languages (for a review, see Balota, Yap, & Cortese, 2006), we found that the largest predictors of reading latencies in Arabic were word length (in phonemes) and (subjective) frequency. This was followed by the semantic effects of imageability and familiarity.

Word length effects have been mostly described in terms of length in letters and syllables (Barca et al., 2002; Burani et al., 2007; Cortese & Khanna, 2007; Cuetos & Barbón, 2006; Davies et al., 2014; Yap & Balota, 2009) as predictors of reading latencies. In the present study, we have used length in phonemes. Unlike other studies in Western alphabetic scripts such as English, we chose to use length in phonemes rather than length in letters because of the characteristics of the Arabic script. Indeed, only the consonants are represented in Arabic script, and the vowels are represented by diacritic marks placed on top or under the word to indicate pronunciation (Ibrahim, 2013). Thus, we claim that when comparing Western alphabetic and Arabic scripts, the length in phonemes in Arabic would be the equivalent of length in letters in Western scripts. In this study, length in phonemes, like length in letters in other languages, affected reading latencies in TA.

The effect of lexical variables, such as subjective frequency, has been found in English (Balota et al., 2004; Cortese & Khanna, 2007). We also found that this variable strongly affected reading latencies in TA. Additionally, name agreement (measured by the H index) also affected naming latencies. Concretely, stimuli with fewer candidate names (i.e., high agreement) elicited shorter latencies than words with a wider range of possible candidates (i.e., low agreement). As we discussed in the introduction, there are two possible loci of name agreement in the literature: the level of structural encoding (i.e., the object recognition stage) or the level of lemma or lexical selection (Barry et al., 1997; Cuetos et al., 1999). Seeing that in the task we used only words, and no pictures were presented, the fact that name agreement affected latencies in a word-reading task (i.e., in the absence of picture recognition processes) further supports the idea that name agreement exerts its influence at the level of lemma or lexical selection (Barry et al., 1997; Cuetos et al., 1999; Vitkovitch & Tyrrell, 1995). Further evidence in support of the lemma selection and subsequent phonological encoding loci for name agreement has come from studies using event-related brain potentials (ERPs). Valente et al. (2014) used a regression design for picture naming. As in our study, they found that name agreement was a significant predictor of naming latencies and that this variable had a late effect on word production (around 400 ms after picture onset and until 100 ms before response articulation). Similarly, Shao, Roelofs, Acheson, and Meyer (2014) conducted a picture-naming study with a group of high- and another of low-name-agreement (measured by the

H index) stimuli using ERPs. They found that NA affected naming latencies. Interestingly, they also found that the amplitude of the N2, a negative ERP component associated with response inhibition, was larger for the low-NA group in the time window of 170–330 ms post-picture-onset. This suggests that the effect of NA corresponds to lexical selection and that having several candidate names for a stimulus requires inhibition of the potential competitors in order to produce a response. Thus, it seems that the lemmas of words referring to objects with several candidate names coactivate the lemmas of all the possible names for that object, inducing larger latencies due to the need to inhibit all candidates but one.

Semantic variables like familiarity (Cuetos & Barbón, 2006) and imageability (Balota et al., 2004; Cortese & Schock, 2013; Cuetos & Barbón, 2006) also predict reading latencies. In our study, we found that both semantic variables affected reading. However, why would semantic variables affect reading in Arabic in the presence of diacritic marks for vowel pronunciation, as in the present study? The presence of such marks would make orthography-to-phonology mappings transparent in Arabic. In such a case, the lack of semantic effects in reading would be expected in Arabic, as compared to other transparent scripts such as Italian (Barca et al., 2002; Burani et al., 2007). Semantic effects as AoA and familiarity in reading are not new for transparent languages. They have also been found in Spanish (Cuetos & Barbón, 2006; Wilson et al., 2013) and Turkish (Raman, 2006) word reading. In Arabic, the diacritic marks for vowels that make mappings transparent are not compulsory. Thus, readers are used to finding only the consonant strings (i.e., unvowelized words) in most written contexts. This makes semantic reliance the only way to correctly pronounce a word when the vowel diacritic marks are not present (e.g., the consonant string مصلحة written without any diacritic marks may be read in two ways: /msˤalħa/ [broom] or /masˤlħa/ [a benefit]). This could make Arabic readers rely more on semantic strategies even in the presence of vowel marks, since the simple orthography-to-phonology mappings could lead to incorrect pronunciations. In order to validate this hypothesis, more empirical evidence will be needed in future studies.

Another point of interest in the results of our regression analyses is the fact that the beta coefficient of familiarity was positive, which seems to indicate that the more familiar the object represented by the word was, the slower the word was read. However, as can be seen from the correlation matrix in Table 2, RTs and familiarity ratings were, as expected, significantly and negatively correlated. This means that, as had been found in the previous literature, the more familiar an item was, the faster it was processed. Thus, it seems that when we entered this variable in the regression models, a change in the sign of the beta coefficient took place. This is possibly a case of positive net suppression (Krus & Wilkinson, 1986). It would seem that this phenomenon occurred because

familiarity is more correlated with other variables in the model—that is, subjective frequency and imageability—than it is with naming latencies. Additionally, subjective frequency and imageability were both more strongly correlated with RTs than familiarity was. According to Krus and Wilkinson, in such cases a variable like familiarity would suppress the error variance of variables like subjective frequency and imageability more than it would predict the variance of the dependent variable (i.e., naming latencies). That is not to say that familiarity does not influence naming latencies at all, but simply that it does so to a lesser extent than subjective frequency and imageability.

In sum, the results of reading in TA showed similarities to those found in other studies using alphabetic Western scripts. To the best of our knowledge, this is the first report on the effect of psycholinguistic key variables such as familiarity and imageability in reading using a large set of words in Arabic. Further studies will need to be conducted in TA to explore the effects of these variables in different tasks (e.g., picture naming, lexical decision, etc.).

Conclusion

The first aim of the present study was to create an extensive normative database of pictures and their names for TA. The database contains norms for seven important psycholinguistic variables: NA (*H* and %), FAM, IMA, FREQ, and WL (phWL and syllWL). The second aim was to investigate the effects of these newly obtained variables on word reading in TA.

Evidence has shown that each of these variables influences different stages of language processing in different experimental tasks and in different languages. NA, the degree to which the speakers of a language agree on the names of objects, has consistently been shown to be the most robust determinant of naming latencies in picture-naming tasks (e.g., Alario et al., 2004). The effect of FAM in this task is somehow mitigated, but some studies have shown a significant influence of this variable. For example, Hirsh and Funnell (1995) identified FAM as a strong predictor of picture-naming latencies in patients suffering from the semantic variant of primary progressive aphasia. The influence of this variable has somehow been equated to that of FREQ, with each variable affecting different stages of processing. The locus of the FAM effect is at the level of semantic activation, whereas FREQ has been known to significantly affect the level of phonological encoding in picture-naming, reading, and lexical-decision tasks (e.g., Davies, Rodríguez-Ferreiro, Suárez, & Cuetos, 2013). WL has also been found to affect word reading. For example, Davies, Rodríguez-Ferreiro, et al. (2013) found that the reading performance of healthy and dyslexic Spanish children was affected by WL, with longer words taking more time to be read.

The influence of the abovementioned variables on processing in Arabic has been the subject of little or no inquiry. That is why we conducted a reading-aloud task in TA with the stimuli of the database and by means of multiple regression analyses. As in studies in other languages, we found that the best predictors of reading aloud in Arabic were word length and frequency (e.g., Balota et al., 2004; Yap & Balota, 2009). Also, we used name agreement (*H*) as a predictor. Normally, this type of variable is not considered in reading-aloud studies. We found that it did predict word-reading latencies. This comes as support for the idea that name agreement has, at least partially, its locus at the lemma selection level (Barry et al., 1997; Cuetos et al., 1999).

One interesting finding was the influence of semantic variables in word reading in the presence of diacritic marks for vowel pronunciation in TA. The Arabic script is highly transparent only when it includes such marks; otherwise, pronunciation of the consonant string is highly inconsistent, and various pronunciations are possible for the same consonant string (Ibrahim, 2013). These alternative pronunciations are only disentangled by the context provided in a sentence. So, in the presence of diacritics, as in other transparent scripts, the influence of semantic variables was not predicted for reading aloud (Barca et al., 2002; Burani et al., 2007). Notwithstanding, the semantic variables familiarity and imageability also affected reading aloud in TA. We believe that Arabic readers rely on semantics even in the presence of vowel marks, because this is the only nonambiguous way to read Arabic in all possible contexts.

In conclusion, the present database offers the opportunity to investigate the effects of each of these psycholinguistic variables in a spoken variety of Arabic. To the best of our knowledge, this study is the first to offer such a sizeable normative database for Arabic and to study the effects of the abovementioned variables in reading aloud. It will thus be of great use in research involving this language, since it provides the means to properly control experimental studies involving Arabic-speaking subjects, both healthy and impaired, and will allow for comparisons with other intra- and cross-linguistic studies.

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,* 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,* 140–155. doi:10.3758/BF03195559

Bachoud-Lévi, A. C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study. *Journal of Memory and Language, 346,* 331–346. doi:10.1006/jmla.1998.2572

Bakhtiar, M., Nilipour, R., & Weekes, B. (2012). Predictors of timed picture naming in Persian. *Behavior Research Methods, 17,* 18–19. doi:10.3758/BF03193165

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316. doi:10.1037/0096-3445.133.2.283

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29,* 639–647. doi:10.3758/BF03200465

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 285–375). Amsterdam, The Netherlands: Academic Press.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers, 34,* 424–434. doi:10.3758/BF03195471

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology, 50A,* 560–585. doi:10.1080/783663595

Barton, J. J. S., Hanif, H. M., Eklinder Björnström, L., & Hills, C. (2014). The word-length effect in reading: A review. *Cognitive Neuropsychology, 31,* 378–412. doi:10.1080/02643294.2014.895314

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,* 344–380. doi:10.3758/BF03196494

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,* 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,* 89–114. doi:10.1348/000712602162463

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,* 158–167. doi:10.3758/BF03195507

Boudelaa, S., & Marslen-Wilson, W. D. (2010). Aralex: A lexical database for Modern Standard Arabic. *Behavior Research Methods, 42,* 481–487. doi:10.3758/BRM.42.2.481

Boudelaa, S., & Marslen-Wilson, W. D. (2013). Morphological structure in the Arabic mental lexicon: Parallels between standard and dialectal Arabic. *Language and Cognitive Processes, 28,* 1453–1473. doi:10.1080/01690965.2012.719629

Burani, C., Arduino, L. S., & Barca, L. (2007). Frequency, not age of acquisition, affects Italian word naming. *European Journal of Cognitive Psychology, 19,* 828–866. doi:10.1080/09541440600847946

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25,* 85–95. doi:10.1080/14640747308400325

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York, NY: Routledge.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256. doi:10.1037/0033-295X.108.1.204

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1084–1096. doi:10.1037/0278-7393.16.6.1084

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36,* 384–387. doi:10.3758/BF03195585

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology, 60,* 1072–1082. doi:10.1080/17470210701315467

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology, 66,* 946–972. doi:10.1080/17470218.2012.722660

Cuetos, F., & Barbón, A. (2006). Word naming in Spanish. *European Journal of Cognitive Psychology, 18,* 415–436. doi:10.1080/13594320500165896

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers, 31,* 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,* 171–237. doi:10.1006/jecp.1996.2356

Damian, M. F., Bowers, J. S., Stadthagen-Gonzalez, H., & Spalek, K. (2010). Does word length affect speech onset latencies when producing single words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 892–905. doi:10.1037/a0019446

Daoud, M. (2001). The language situation in Tunisia. *Current Issues in Language Planning, 2,* 1–52. doi:10.1080/14664200108668018

Davies, R., Barbón, A., & Cuetos, F. (2013). Lexical and semantic age-of-acquisition effects on word naming in Spanish. *Memory & Cognition, 41,* 297–311. doi:10.3758/s13421-012-0263-8

Davies, R., Rodríguez-Ferreiro, J., Suárez, P., & Cuetos, F. (2013). Lexical and sub-lexical effects on accuracy, reaction time and response duration: Impaired and typical word and pseudoword reading in a transparent orthography. *Reading and Writing, 26,* 721–738. doi:10.1007/s11145-012-9388-1

Davies, R., Wilson, M., Cuetos, F., & Burani, C. (2014). Reading in Spanish and Italian: Effects of age of acquisition in transparent orthographies? *Quarterly Journal of Experimental Psychology, 67,* 1808–1825. doi:10.1080/17470218.2013.872155

De Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 824–845. doi:10.1037/0278-7393.15.5.824

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,* 588–615. doi:10.3758/BF03200832

Desrochers, A., Liceras, J. M., Fernández-Fuertes, R., & Thompson, G. L. (2010). Subjective frequency norms for 330 Spanish simple and compound words. *Behavior Research Methods, 42,* 109–117. doi:10.3758/BRM.42.1.109

Desrochers, A., & Thompson, G. L. (2009). Subjective frequency and imageability ratings for 3,600 French nouns. *Behavior Research Methods, 41,* 546–557. doi:10.3758/BRM.41.2.546

Dimitropoulou, M., Duñabeitia, J. A., Blitsas, P., & Carreiras, M. (2009). A standardized set of 260 pictures for Modern Greek: Norms for name agreement, age of acquisition, and visual complexity. *Behavior Research Methods, 41,* 584–589. doi:10.3758/BRM.41.2.584

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 515–523. doi:10.1037/0278-7393.24.2.515

Ferguson, C. A. (1959). Diglossia. *Word, 15,* 325–340.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116–124. doi:10.3758/BF03195503

Gilhooly, K. J., & Gilhooly, M. L. M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology, 71,* 105–110. doi:10.1111/j.2044-8295.1980.tb02736.x

Hamdi, R., Ghazali, S., & Barkat-Defradas, M. (2005). Syllable Structure in Spoken Arabic: A comparative investigation. *Interspeech,* 2245–2248.

Hirsh, K. W., & Funnell, E. (1995). Those old, familiar things: Age of acquisition, familiarity and lexical access in progressive aphasia. *Journal of Neurolinguistics, 9,* 23–32. doi:10.1016/0911-6044(95)00003-8

Ibrahim, R. (2013). Reading in Arabic: New evidence for the role of vowel signs. *Creative Education, 04,* 248–253. doi:10.4236/ce.2013.44036

Izura, C., Hernández-Muñoz, N., & Ellis, A. W. (2005). Category norms for 500 Spanish words in five semantic categories. *Behavior Research Methods, 37,* 385–397. doi:10.3758/BF03192708

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 824–843. doi:10.1037/0278-7393.20.4.824

Khwaileh, T., Body, R., & Herbert, R. (2014). A normative database and determinants of lexical retrieval for 186 Arabic nouns: Effects of psycholinguistic and morpho-syntactic variables on naming latency. *Journal of Psycholinguistic Research, 43,* 749–769. doi:10.1007/s10936-013-9277-z

Klapp, S. T., Anderson, W. G., & Berrian, R. W. (1973). Implicit speech in reading reconsidered. *Journal of Experimental Psychology, 100,* 368–374.

Kremin, H., Perrier, D., De Wilde, M., Dordain, M., Le Bayon, A., Gatignol, P., . . . Arabia, C. (2001). Factors predicting success in picture naming in Alzheimer's disease and primary progressive aphasia. *Brain and Cognition, 46,* 180–183. doi:10.1006/brcg.2000.1270

Krus, D. J., & Wilkinson, S. M. (1986). Demonstration of properties of a suppressor variable. *Behavior Research Methods, Instruments, & Computers, 18,* 21–24. doi:10.3758/BF03200988

La Heij, W., Puerta-Melguizo, M. C., van Oostrum, M., & Starreveld, P. A. (1999). Picture naming: Identical priming and word frequency interact. *Acta Psychologica, 102,* 77–95. doi:10.1016/S0001-6918(99)00020-7

Lachman, R. O. Y. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology, 99,* 199–208.

Lachman, R., Shaffer, J. P., & Hennrikus, D. (1974). Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time. *Journal of Verbal Learning and Verbal Behavior, 13,* 613–625. doi:10.1016/S0022-5371(74)80049-6

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38. **disc. 38–75**.

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,* 452–460. doi:10.3758/BRM.42.2.452

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental*

Psychology: Human Perception and Performance, 13, 14–24. doi:10.1037/0096-1523.13.1.14

Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition, 20,* 705–714. doi:10.3758/BF03202720

Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. *Behavior Research Methods, 37,* 398–416. doi:10.3758/BF03192709

Nisi, M., Longoni, A. M., & Snodgrass, J. G. (2000). Misure italiane per l'accordo sul nome, familiarità ed età di acquisizione, per le 260 figure di Snodgrass e Vanderwart (1980) [Italian norms for name agreement, familiarity, and age of acquisition for the S&V (1980) set of pictures]. *Giornale Italiano Di Psicologia, 27,* 205–218. **Retrieved from http://psycnet.apa.org/psycinfo/2001-17858-004**.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17,* 273–281. doi:10.1080/17470216508416445

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology, 61,* 106–151. doi:10.1016/j.cogpsych.2010.04.001

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115. doi:10.1037/0033-295X.103.1.56

Plaut, D., & Shallice, T. (1993). Deep dyslexia: A case-study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377–500. doi:10.1080/02643299308253469

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39,* 859–862. doi:10.3758/BF03192979

Raman, I. (2006). Are acquired dyslexias and dysgraphias language: Specific or universal? *Visual Cognition, 13,* 1044–1053. doi:10.1080/13506280500153200

Raman, I., Raman, E., & Mertan, B. (2014). A standardized set of 260 pictures for Turkish: Norms of name and image agreement, age of acquisition, visual complexity, and conceptual familiarity. *Behavior Research Methods, 46,* 588–595. doi:10.3758/s13428-013-0376-4

Roelofs, A. (2002). Syllable structure effects turn out to be word length effects: Comment on Santiago et al. (2000). *Language and Cognitive Processes, 17,* 1–13. doi:10.1080/01690960042000139

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass–Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers, 28,* 537–555. doi:10.3758/BF03200541

Santiago, J., MacKay, D. G., Palma, A., & Rho, C. (2000). Sequential activation processes in producing words and syllables: Evidence

from picture naming. *Language and Cognitive Processes, 15,* 1–44. doi:10.1080/016909600386101

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568. doi:10.1037/0033-295X.96.4.523

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119,* 159–187. doi:10.1016/j.actpsy.2005.01.002

Shao, Z., Roelofs, A., Acheson, D. J., & Meyer, A. S. (2014). Electrophysiological evidence that inhibition supports lexical selection in picture naming. *Brain Research, 1586,* 130–142. doi:10.1016/j.brainres.2014.07.009

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,* 174–215. doi:10.1037/0278-7393.6.2.174

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers, 28,* 516–536. doi:10.3758/BF03200540

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38,* 598–605. doi:10.3758/BF03193891

Tsaparina, D., Bonin, P., & Méot, A. (2011). Russian norms for name agreement, image agreement for the colorized version of the Snodgrass and Vanderwart pictures and age of acquisition, conceptual familiarity, and imageability scores for modal object names. *Behavior Research Methods, 43,* 1085–1099. doi:10.3758/s13428-011-0121-9

Valente, A., Burki, A., & Laganaro, M. (2014). ERP correlates of word production predictors in picture naming: A trial by trial multiple regression analysis from stimulus onset to response. *Frontiers in Neuroscience, 8,* 1–13. doi:10.3389/fnins.2014.00390

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *Quarterly Journal of Experimental Psychology, 48,* 822–848. doi:10.1080/14640749508401419

Weekes, B., Shu, H., Hao, M., Liu, Y., & Tan, L. H. (2007). Predictors of timed picture naming in Chinese. *Behavior Research Methods, 17,* 335–342. doi:10.3758/BF03193165

Wheeldon, L. R., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *Quarterly Journal of Experimental Psychology, 44A,* 723–761. doi:10.1080/14640749208401307

Wilson, M. A., Cuetos, F., Davies, R., & Burani, C. (2013). Revisiting age-of-acquisition effects in Spanish visual word recognition: The role of item imageability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1842–1859. doi:10.1037/a0033090

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60,* 502–529. doi:10.1016/j.jml.2009.02.001