First learned words are not forgotten: Age-of-acquisition effects in the tip-of-the-tongue experience

Eduardo Navarrete[1] · Massimiliano Pastore[1] · Rosa Valentini[2] · Francesca Peressotti[1]

Abstract A large body of evidence indicates that the age at which a word is acquired predicts the time required to retrieve that word during speech production. Here we explored whether age of acquisition also predicts the experience of being unable to produce a known word at a particular moment. Italian speakers named a sequence of pictures in Experiment 1 or retrieved a word as a response to a definition in Experiment 2. In both experiments, the participants were instructed to indicate when they were in a tip-of-the-tongue (TOT) state. Generalized mixed-effects models performed on the TOT and correct responses revealed that word frequency and age of acquisition predicted the TOT states. Specifically, low-frequency words elicited more TOTs than did high-frequency words, replicating previous findings. In addition, late-acquired words elicited more TOTs than did early-acquired words. Further analyses revealed that the age of acquisition was a better predictor of TOTs than was word frequency. The effects of age of acquisition were similar with subjective and objective measures of age of acquisition, and persisted when several psycholinguistic variables were taken into consideration as predictors in the generalized mixed-effects models. We explained these results in terms of weaker semantic-to-phonological connections in the speech production system for late-acquired words.

✉ Eduardo Navarrete
  eduardo.navarrete@unipd.it

[1] Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, Padua, Italy

[2] Dipartimento di Studi Linguistici e Letterari, Università di Padova, Padua, Italy

Language production is usually rather efficient. When presented with a picture of a "dog" never seen before, speakers encounter no difficulty in naming the picture correctly as a "dog." Nonetheless, on some occasions, language production can be inefficient and speakers may be unable to retrieve the appropriate word. Failures during word retrieval occur even when the person knows there is a word to transmit her or his communicative intention. When temporarily unable to retrieve a known word, a speaker is experiencing a phenomenon called *tip-of-the-tongue* (hereafter, TOT; e.g., A.S. Brown, 1991, 2012; R. Brown & McNeill, 1966).

The frustrating sensation in which a known word is momentarily inaccessible is not an "all-or-nothing" feeling. Speakers experiencing a TOT can often accurately retrieve some information associated with the intended word, such as partial phonological information or syntactic information, the initial phoneme, or its grammatical gender (e.g., A.S. Brown, 1991; Miozzo & Caramazza, 1997; Vigliocco, Antonini, & Garrett, 1997). Thus, the TOT phenomenon constitutes perhaps the most intuitive evidence for the notion that the translation of our communicative intentions into a specific set of sounds is completed in distinct stages of processing. Consistent with this intuition, theories of speech production agree on the notion that word production involves the retrieval of different types of information (e.g., Caramazza, 1997; Dell, 1986; Levelt, Roelofs, & Meyer, 1999). For instance, in the course of naming a picture, speakers need to retrieve the conceptual representation corresponding to the picture. Later on, speakers must retrieve the lexical and phonological information associated with the intended word in order to be able to initiate the articulatory–motor program and utter the word.

⚫ Springer

The TOT phenomenon is thought to reflect successful completion of access to the concept but failure during the retrieval of linguistic information. The most influential interpretation of the origin of TOTs assumes that the retrieval deficit arises in the transmission of activation between the conceptual and phonological systems (Burke, MacKay, Worthley, & Wade, 1991; Burke & Shafto, 2004). Specifically, during a TOT the connections up to the word's phonological content are weakened, and therefore phonological retrieval is not fully achieved. Burke and colleagues (1991) argued that one of the critical causes that determines the strengthening of the connections is frequency of use: The less frequently a word is produced, the lower the opportunities to strengthen the semantic-to-phonological links, and consequently, the probabilities of experiencing a TOT (with that word) increase. At least two empirical observations are congruent with this hypothesis: One is that TOT states are more susceptible to be experienced with low-frequency than with high-frequency words (e.g., Burke et al., 1991; Harley & Bown, 1998); the second is the finding that bilingual speakers are more prone to TOTs than monolingual speakers (Gollan & Acenas, 2004; Gollan, Bonanni, & Montoya, 2005; Gollan & Silverberg, 2001; Pyers, Gollan, & Emmorey, 2009). Presumably the semantic-to-phonological mappings are weaker in bilinguals because they speak each language less time than monolinguals. That is, because bilinguals divide their frequency of use between two languages, the strengths of the connections in their speech production system are weaker than among monolinguals. In sum, the TOT phenomenon has relevant implications for the understanding of how linguistic information (i.e., lexical and phonological knowledge) is retrieved during language production, and word frequency seems to be a crucial variable determining the success of this retrieval process.

Nonetheless, word retrieval during language production depends on several other variables. It is a well-established phenomenon that the speed and accuracy with which speakers retrieve a word is influenced by the age in life at which that word was first learned. Specifically, early-acquired words tend to be named more quickly and accurately than late-acquired words, an observation known as the *age-of-acquisition effect* (hereafter, *AoA effect*; for reviews, see Johnston & Barry, 2006; Juhasz, 2005). Thus, it could be the case that failures during word retrieval, as in TOT states, may depend on the age at which the words have been acquired. The main aim of the present research was to test this hypothesis.

Although AoA is one of the most reliable factors determining word production, there is still some disagreement about the representation of AoA in the speech production system (e.g., Catling & Johnston, 2009). One of the first interpretations of the effect suggested that early-acquired words are stored as "complete" forms in the phonological lexicon, whereas late-acquired words are stored in a more

"fragmented" fashion (G.D.A. Brown & Watson, 1987; see also Gerhand & Barry, 1998). Such a difference in the formats of the stored phonological representation would allow early-acquired words to access their phonological representations more rapidly, resulting in shorter naming latencies. Although this hypothesis has been challenged by compelling evidence and lacks convincing evidence supporting it (e.g., Monaghan & Ellis, 2002), the notion that AoA may arise, at least in part, during phonological encoding has been recently supported by other studies (e.g., Hernandez, & Fiebach, 2006; Kittredge, Dell, Verkuilen, & Schwartz, 2008; Navarrete, Scaltritti, Mulatti, & Peressotti, 2013). A second line of research, first advanced by Bates, Burani, D'Amico, and Barca (2001), hypothesizes that the AoA effect is mainly localized at the level of lexical retrieval. For instance, Belke, Brysbaert, Meyer, and Ghyselinck (2005) reported an interaction between AoA and semantic interference in a picture-naming task (i.e., late-acquired target words showed greater interference than early-acquired words). Under the assumption that semantic interference in picture-naming tasks arises during the selection of the lexical node, Belke and colleagues concluded that such an interaction indicates that (part of) the AoA effect emerges at the lexical retrieval level. However, AoA effects have been observed not only in language production, but in many other tasks, such as word reading, lexical decision, and semantic categorization, as well as in aphasic linguistic performance (for a review, see Johnston & Barry, 2006), suggesting that this variable could also affect the input to the linguistic system and/or a stage that is common to both production and comprehension tasks, such as the semantic system.

It therefore seems reasonable to hypothesize that AoA does not selectively affect a single stage of processing during speech production, but that AoA effects are distributed across several stages (Brysbaert & Ghyselinck, 2006; Catling & Johnston, 2009; Ellis & Lambon Ralph, 2000; Holmes & Ellis, 2006; Navarrete et al., 2013). Accordingly, a large empirical and computational tradition has focused on the factors that determine the phenomenon rather than on the layer of processing within the speech production system at which the phenomenon is localized. With regard to this, theories agree that two factors are critical in AoA effects: (a) the weights of the mappings between conceptual and linguistic representations, and (b) the "nature" of these mappings. In their influential article, Ellis and Lambon Ralph presented a connectionist network in which early-acquired words have greater influence in modulating the network's weights between the input and output units of the network. The difference in the modulations would thus be explained in terms of reductions in the plasticity of the system for late-acquired words. That is, as the network learns new words, it becomes stable and decreases in its capacity to assimilate new weight patterns between input and output units. Lambon Ralph and Ehsan (2006) further argued that this account predicts that AoA effects should

mainly emerge in those circumstances in which the input-to-output mappings are arbitrary, as for instance in the semantic-to-phonology mappings involved in picture-naming tasks. By contrast, when the mappings are not arbitrary, it would be possible to generalize from early-acquired (trained) words to later acquired (untrained) words, leading to a cancellation or reduction of AoA effects (see also Mermillod, Bonin, Méot, Ferrand, & Paindavoine, 2012; Zevin & Seidenberg, 2002). Word reading, mainly in languages with transparent orthography-to-phonology conversion systems, is an example of nonarbitrary input-to-output mappings. In this case, reading late-acquired words could be performed through the grapheme-to-phonology conversion rules that were acquired when the network was exposed to the first words (i.e., early-acquired words). Congruent with this hypothesis, AoA effects are larger in picture-naming tasks than in word-reading tasks (for a review, see Johnston & Barry, 2006). In sum, there is theoretical consistency in localizing both the TOT phenomenon and (part of) the AoA effect in the mappings from semantic up to phonological representations. Nevertheless, to the extent of our knowledge, no attempt at studying the role of AoA in TOT states has previously been made. Our main aim here was to explore this role. Specifically, we investigated whether the probability of experiencing a TOT is predicted by AoA.

The present research

Although the relationship between frequency and TOTs has been studied extensively, there is only indirect evidence of the relationship between AoA and TOTs. For instance, in a normalized naming study of a set of photographs of celebrities, Bonin, Perret, Méot, Ferrand, and Mermillod (2008) observed that AoA predicts the percentages of TOTs. Another study showing some indirect, not conclusive, evidence for the relation between AoA and TOTs is that of Kittredge and colleagues (2008). These authors analyzed the semantic and phonological picture-naming errors of 50 aphasic patients by controlling several variables, such as frequency and AoA. Two main observations in relation to the AoA variable were reported. The first was that AoA predicted phonological errors (e.g., "mat" for "cat"), but not semantic errors (e.g., "dog" for "cat"), suggesting that AoA affects the retrieval of the phonological content of the word (see the General Discussion below for further discussion of this point). The second observation, more relevant to our purposes here, was that AoA predicted the numbers of omissions that patients made. Similar to the TOT phenomenon, omissions reflect the inability to retrieve a target word. However, whereas TOT states reflect a failure during linguistic access, it is impossible to distinguish the level at which an omission occurs.

Indeed, omissions may occur at the conceptual level or during lexical and/or phonological retrieval.

In order to explicitly test the influence of AoA in TOT states with common nouns and in a not-brain-damaged population, two experiments were conducted. In Experiment 1, a set of 154 Italian nouns were selected for a picture-naming task. In Experiment 2, a set of 230 definitions were presented for a word-naming-to-definition task. Before the tasks, participants were informed that a TOT state described the feeling "when you are sure you know the name to denominate the picture/definition but can't remember it," and they were instructed to report when they were experiencing a TOT. Our main goal was to explore whether AoA predicts TOTs. AoA and word frequency are highly correlated variables, since early-acquired words tend to be higher in frequency (e.g., Chalard, Bonin, Méot, Boyer, & Fayol, 2003). Thus, in order to check for AoA effects we needed to control for word frequency effects. We did this by adopting a mixed-effect approach (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013). The statistical approach that we adopted here allowed us to specify which of the two variables (i.e., AoA or frequency) was a better predictor of TOTs (see below for details).

In summary, we expected to replicate previous studies and report that low-frequency words elicited more TOTs that high-frequency words (e.g., Burke et al., 1991; Harley & Bown, 1998). In relation to the critical AoA variable, if access to linguistic representations during speech production is AoA-sensitive, we should observe that AoA predicts TOTs over and above the effect of word frequency. Specifically, under the assumption that AoA determines the strengthening of the links between semantic and phonological units (Ellis & Lambon Ralph, 2000), we expected to observe more TOTs for late-acquired than for early-acquired words. This would be independent of the task at hand (i.e., picture naming or naming to definition).

Experiment 1: picture naming

In Experiment 1, we explored whether AoA predicts the probability of experiencing a TOT in a picture-naming task. In order to estimate the AoA values, we followed the procedure generally used for constructing the normative databases of AoA, by asked adult participants who did not take part in the naming experiment when they thought they had first acquired each of the 154 experimental words. The word written frequencies of the selected items were counted on the basis of the Bertinetto et al. (1995) Italian corpus.

Besides the correlation with frequency, AoA tends to correlate in picture-naming tasks with other important psycholinguistic variables, such as name agreement measures (e.g., name agreement and the H statistic). Name agreement

describes the proportion of participants who give the most common name to one specific item, whereas the *H* statistic is a logarithmic function describing the different names that an item receives and the proportion of participants giving each name (Snodgrass & Vanderwart, 1980). It has been shown that the *H* statistic captures more information about the variability of names across participants than does the simple percentage-of-agreement measure (Snodgrass & Vanderwart, 1980; see also Alario & Ferrand, 1999; Snodgrass & Yuditsky, 1996). Therefore, in this research we decided to control for the influence of the *H* statistic as a measure of name agreement. An *H* value of 0 indicates that all participants provided the same name in response to a picture, and the *H* value increases with the number of different given names. It has also been shown that word length correlates with AoA measures in picture-naming tasks. In a study conducted by Pérez (2007), naming latencies showed a significant correlation between AoA and two word length measures, namely the number of phonemes and the number of syllables (e.g., Morrison, Chappell, & Ellis, 1997). Here we controlled for word length effects by including as variables the numbers of phonemes and syllables.

Method

Participants Thirty native Italian speakers took part in the experiment (mean age= 33, *SD*= 17, range: 19–64).

Materials A total of 154 Italian common nouns were selected (see Appendices A and B). The target stimuli consisted of 58 pictures selected from different databases (Alario & Ferrand, 1999; Bonin, Peereman, Malardier, Méot, & Chalard, 2003; Dell'Acqua, Lotto, & Job, 2000) and 96 photographs selected from Internet. AoA values were estimated from a norming study with 24 new participants who did not take part in the main experiment (mean age= 20, *SD*= 1, range: 19–23). In this study, participants were instructed to rate the age at which they thought they had first learned each of the selected words, on a 1–9 Likert scale (1= *learned at 0–1 year*, 9= *learned at age 15+*, with 2-year age bands in between).

Design Ten different randomization lists containing the experimental items were created. Care was taken that two items of the same semantic category were never presented in consecutive trials, in order to avoid semantic priming (e.g., Navarrete, Del Prato, Peressotti, & Mahon, 2014). Each experimental list was used a total of three times (across the 30 participants). Four filler pictures depicting very common objects (i.e., high-frequency and early-acquired objects) were selected as training items and presented twice before the experimental items. None of the participants reported TOTs with the filler pictures.

Procedure An experimental trial involved the following events. A fixation cross was shown in the center of the screen for 500 ms and was followed by the target picture for 10 s or until the participant's response. Participants were instructed to press the space bar to start the next trial. Stimulus presentation, logging of response times, and response recording were controlled by the DMDX program (Forster & Forster, 2003).

Participants were instructed to report TOTs whenever they occurred. When unable to name the pictures within the 10 s of the stimulus's presentation, participants were asked to decide whether or not they knew the target name. They were instructed to respond "don't know" if they did not recognize the picture or if they thought that they had never known the appropriate name for that concept. If participants responded that they knew what the picture represented and that there was a word to name it, the experimenter asked whether or not the participant was in a TOT state. The experimenter waited about 5 s before providing the target name. After that, the experimenter asked the participant whether the intended target was the name that the participant was trying to retrieve (in the case that she or he was in a TOT state) or if she or he knew the word (in the other cases).

Analysis Responses were classified as GOT (as in "got it") when the participant correctly produced the intended target name within the 10 s of the stimulus's presentation. Failures to name the target within the 10 s were classified into one of four major types: (a) positive TOT (TOT+), when the participant reported a TOT state that was later confirmed to be for the intended target word, or that was spontaneously self-resolved by the participant within the 5-s delay before the experimenter provided the target word; (b) negative TOT (TOT–), when the participant reported a TOT state but the intended target word did not correspond to the word that the participant was thinking of; (c) DK ("don't know"), when the participant reported not knowing the target picture or not knowing the target word provided by the experimenter; and (d) notGOT, when the participant did not respond (without experiencing a TOT state) or produced an erroneous picture name. We explored whether AoA and frequency predicted the probability of experiencing a TOT state using generalized linear mixed-effects models (GLMM; see the Results, below). Given the dichotomous nature of the dependent variable, we used a logistic model. Analyses were performed using the lme4 package (Bates, Maechler, & Bolker, 2011) with the R program (R Development Core Team, 2011).

Response scoring Due to a mistake in the randomization of the lists, the item *flight of steps* was not presented to two participants. The items *dagger*, *locomotive*, *stapler*, *wharf*, *wine shop*, and *coat hanger* were named more than 40% of the time with alternative names from those designed by the experimenter, and were excluded from the analysis. In addition, a

total of 33 trials featured problems with the voice key and were discarded from the analysis. The analysis was based on a total of 4,405 trials. Participants successfully named 3,537 (80.3%) of the stimuli with the names designated by the experimenter (i.e., GOT responses). This also included responses containing the lemma form of the target word, diminutive forms, or responses with extra detailed information. In addition, participants successfully named 176 (4%) of the stimuli with names that could be scored as plausible responses but that were different from those expected. These responses included dialectal names or alternative names (e.g., *rimmel* instead of *mascara*, both Italian words to name mascara). The remaining responses consisted of 126 (2.9%) TOTs (range = 0–16). The TOTs were further divided into positive and negative TOTs, with 112 TOTs+ and 14 TOTs–. There were also a total of 49 (1.1%) DK responses and 517 (11.7%) notGOT responses (see Table 1).

Results

GLMMs were performed to investigate the roles of AoA and the other variables in GOT and TOT+ responses. Specifically, the following models were tested and compared. The null model (i.e., M0) contained intercepts only and no predictors. We first explored the influence of the order in which items were presented by adding the predictor item position in M1. Next, we explored the influence of the H statistic by adding this variable as a predictor in M2. Afterward, we explored word length variables by taking into consideration the variables number of phonemes (NumPhon) and number of syllables (NumSyll) as predictors in the M3a and M3b models, respectively. Subsequently, we explored the influences of frequency and AoA in M4 and M5, respectively. In each model we set the same random effects: subjects and items. For comparing the aforementioned models, we performed the likelihood ratio test and took into consideration the Bayesian information criterion (BIC; Schwarz, 1978). In Table 2, we report the results for the model comparison. ΔBIC indicates the differences between the null model (M0) and the other models; a positive ΔBIC value implies that a given model is better than the null model. In the last column we have included

Bayes factor (BF) approximations, using the formula exp(ΔBIC/2) (Raftery, 1995); by using the BF, we were able to compare the relative evidence for different models. For example, a BF value of 3 indicates that one model is three times more likely than the null model. In general, the higher the ΔBIC and BF, the more likely the model is in comparison to the null model.

A direct comparison between M0 and M1 showed that the inclusion of the predictor item position did not improve the model fit, as indicated by the negative ΔBIC value of –4.3. The comparison between M0 and M2, on the other hand, showed that M2 explained the data 247 times better than M0. The upper left panel in Fig. 1 shows that words that elicited TOTs tended to have higher H values than did words named correctly (i.e., GOT responses). We therefore kept the H statistic as a critical predictor of TOTs and explored the influences of the other variables. The inclusion of the word length variables produced lower ΔBIC values (i.e., 4.12 and 5.44 for M3a and M3b, respectively) than the model containing the H statistic as the only predictor (i.e., 11.02 in M2), suggesting that the inclusion of these variables did not improve the model fit. Consequently, the length variables were not included in further comparisons.

The influence of frequency was explored in M4. This model showed a better fit in relation to the model with H as the unique predictor. This can be seen in the bigger ΔBIC parameter for M4 (13.11) than for M2 (11.02). The upper right panel in Fig. 1 shows that words that elicited positive TOTs tended to be less frequent than words that were correctly named (i.e., GOT responses). Subsequently, in M5 we explored the critical variable of AoA by adding this variable as a fixed factor to M4. Critically, M5 showed a better fit than M4: The ΔBIC parameter was bigger in M5 (21.16) than in M4 (13.11). This result suggests that AoA is a critical predictor of TOTs once the influences of the H statistic and frequency effects have been taken into consideration. The results show that TOTs are more likely to be observed with late-acquired words (see Fig. 1, bottom).

Some researchers have proposed that rather than AoA estimates, it is more appropriate to use objective AoA measures (Ellis & Morrison, 1998; but see Zevin & Seidenberg, 2002).

Table 1 Percentages and total numbers of responses (in parentheses) of each scoring response type, separated by experiments (see the main text for details)

Response Type	Experiment 1: Picture Naming		Experiment 2: Naming to Definition
	Overall Set	Lotto et al.'s Subset	Overall Set
GOT	80.3 (3,537)	84.9 (2,147)	69.5 (9,030)
TOT+	2.5 (112)	2 (51)	3.7 (483)
Alternatives	4 (176)	3 (76)	1.3 (171)
TOT–	0.3 (14)	0.2 (4)	0.4 (54)
DK	1.1 (49)	0.3 (8)	2.3 (296)
notGOT	11.7 (517)	9.6 (243)	24.1 (3,129)

Table. 2 Fit indices from the analysis in Experiment 1

	Fixed Effects	Model df	Chisq (df)	p	BIC	ΔBIC	Approx. BF
M0		3			917		
M1	Item position	4	3.89 (1)	= .049	921	− 4.3	0.11
M2	H	4	19.22 (1)	< .001	906	11.02	247
M3a	H + NumPhon	5	20.52 (2)	< .001	913	4.12	7
M3b	H + NumSyll	5	21.85 (2)	< .001	911	5.44	15
M4	H + LogFreq	5	29.51 (2)	< .001	904	13.11	702
M5	H + LogFreq. + AoA	6	45.76 (3)	< .001	895	21.16	> 10,000

df, degree of freedom; Chisq (*df*), chi-squared and degree of freedom; *p*, probability value; BIC, Bayesian information criterion; ΔBIC, differences between the null model (M0) and other models; Approx. BF, Bayes factor approximation, exp(ΔBIC/2).

Objective measures of AoA are derived from children's correct responses on picture-naming tasks. In order to explore whether the influence of AoA in TOTs that we have observed in this experiment was due to the use of subjective measures, we conducted the same analysis with objective measures of AoA. In a recent study, Lotto, Surian, and Job (2010) provided an Italian database of objective AoA measures from children between 2 and 11 years of age. We restricted the further analysis to the subset of items for which Lotto and colleagues' database provides objective AoA measures. Specifically, the analyses were performed of a subset of 85 items. The results mirrored those obtained with subjective AoA measures (see Table 3).

Discussion

The results of this experiment suggest that the probability of experiencing a TOT state during a picture-naming task depends on the age at which the word is acquired in life. However, before drawing further conclusions, we will report Experiment 2 in order to extend the main effect of AoA on TOT states and to address possible concerns that could be raised about Experiment 1. For instance, it is worth noting that the results of this experiment were performed on a small number of TOTs (i.e., 125). Indeed, the rate of TOTs that we obtained was lower than the rates obtained in other TOT studies that have also used a picture-naming task (e.g., Gollan & Acenas, 2004). One could argue that our findings were due to the small set of words. It was therefore crucial to test the reliability of the main phenomenon with a larger number of TOTs. In order to increase the number of TOT observations, in Experiment 2 we used a more "standard" task in TOT research. Participants were exposed to the definitions of common Italian nouns and instructed to retrieve the corresponding noun. In addition, we increased the number of materials by selecting a total of 230 definitions and tested almost double the number of participants in Experiment 1.

Another possible concern regarding the main findings of Experiment 1 refers to the fact that the group of participants who estimated the AoA measures was younger than the group of participants that took part in the naming experiment. Some studies have revealed that AoA ratings are better predictors of word processing when the AoA raters and participants are of the same age (e.g., Cuetos, Samartino, & Ellis, 2012). Although the fact that the same pattern was observed with subjective and objective AoA seems to reduce this possibility, we aimed to replicate our main findings with both rater and experimental groups of similar ages. Finally, some studies have reported that AoA and familiarity are highly correlated variables, in as much as early-acquired words tend to be more familiar than late-acquired words (e.g., Barca, Burani, & Arduino, 2002; Bates et al., 2001). Furthermore, Barca and colleagues (2002) reported a significant correlation between the orthographic neighborhood and AoA variables in a word-reading study in Italian. In Experiment 2, we explored whether the effect of AoA on determining TOT states would remain when familiarity and phonological neighborhood predictors were added in a generalized mixed analysis (see also Pérez, 2007, for an effect of phonological neighborhood on picture-naming latencies).

Experiment 2: naming to definition

Method

Participants Fifty-eight native Italian speakers, students at the University of Padua, took part in the experiment. Their mean age was 20 years (SD = 2.2, range: 18–29). Then a group of 130 new participants from the same population took part in two norming studies. Of these, 80 participated in the AoA study (mean age = 20 years, SD = 1.2, range: 18–25) and 50 in the familiarity study (mean age = 20 years, SD = 4, range: 18–36).

Materials A total of 230 Italian common nouns were selected (113 came from the pool of items used in Exp. 1; see
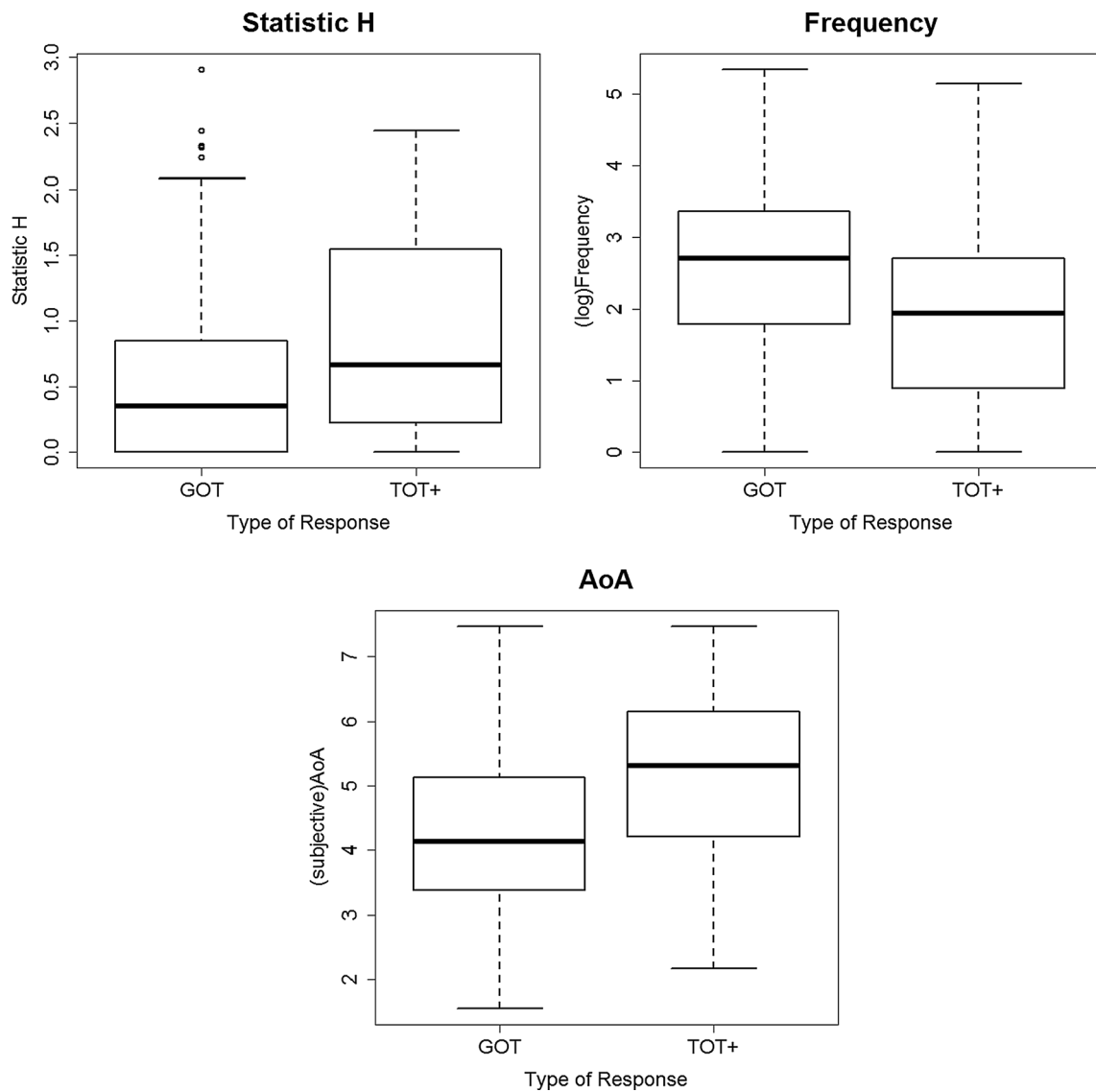
Fig. 1 Distributions of the *H* statistic (upper left), (log) frequency (upper right), and AoA (bottom) variables for GOT responses and TOT+ responses in Experiment 1

Appendices A and B). A definition was created for each of the selected words in order to elicit the verbal response in the naming-to-definition task. Definitions were created by two Italian native speakers (the authors R.V. and F.P.). AoA values were estimated as in Experiment 1, and familiarity values were estimated by means of a 1–5 Likert scale with which participants

Table 3 Fit indices from the analysis of the subset of items included in Lotto and colleagues' (2010) database in Experiment 1

	Fixed Effects	Model *df*	Chisq (*df*)	*p*	BIC	ΔBIC	Approx. BF
M0_L		3			651		
M1_L	Item position	4	2.48 (1)	= .114	656	− 5.34	0.06
M2_L	*H*	4	22.06 (1)	< .001	636	14.76	1,607
M3a_L	*H* + NumPhon	5	22.61 (2)	< .001	644	6.93	32
M3b_L	*H* + NumSyll	5	22.89 (2)	< .001	644	7.15	35
M4_L	*H* + LogFreq	5	27.76 (2)	< .001	639	12.09	422
M5_L	*H* + LogFreq+ AoA	6	47.91 (3)	< .001	626	24.40	> 10,000

For the fit statistics, see the note to Table 2.

were instructed to rate how often they thought they were exposed to each of the selected words (1, *very unfamiliar*; 5, *very familiar*). The participants in both norming studies were exposed randomly to half of the selected words (i.e., 115). For each word, a total of 40 observations were collected in the AoA and 25 in the familiarity normative studies.

Design and procedure Stimuli were presented in a random order, with the only constraint being that two items from the same semantic category were never presented in consecutive trials. Each definition was read aloud by the experimenter, and participants were required to name the noun within approximately 10 s. Participants were tested individually, and response times and response recordings were not taken. Response scoring was the same as was described in the previous experiment.

Analysis and response scoring The same analyses were carried out as in Experiment 1. The items *dawn*, *lamp*, *arsenal*, *nostril*, *fairy tale*, and *gust* were excluded from the analysis because 35% of the times they were named with names alternative to those designated by the experimenter. The analysis was based on 12,992 trials, in which participants named 9,030 of the stimuli correctly (i.e., GOT responses, 69.5%). This included 171 responses with names that could be scored as plausible responses but that were different from what was expected (1.3%). The remaining responses consisted of 537 TOTs (range= 0–17), with 483 (3.7%) TOTs+ and 54 (0.4%) TOTs–. In addition, participants made 296 (2.3%) DK responses and 3,129 (24.1%) notGOT responses (see Table 1).

The same GLMM analyses were conducted as in the previous experiment. In this experiment, before testing the influences of frequency and AoA, we explored the influences of six control variables: H statistic, number of phonemes, number of syllables, phonological neighborhood (Phon_N), frequency of the phonological neighborhood (Phon_N_MFreq), and familiarity. The Phon_N and Phon_N_MFreq values were taken from the PhonItalia database (Goslin, Galluzzi, & Romani, 2014). The first four models were identical to models M1–M3b in Experiment 1. M1 contained the predictor item position, M2 the predictor H statistic, and M3a and M3b the word length predictors (i.e., NumPhon and NumSyll, respectively). The influences of the phonological neighborhood variables were explored in models M4a and M4b. In M5, we added the predictor familiarity. Finally, we explored the influence of frequency in M6 and the influence of AoA in M7. Following the same procedure as in Experiment 1, when a predictor improved the fit, that predictor was kept in the model for further comparisons. This procedure allowed us to estimate the relevance of the critical variables AoA and frequency, once all of the control predictors had been tested.

Results

TOTs The results are reported in Table 4. We observed no item position effect (i.e., M1). In M2 the inclusion of the H statistic as a predictor increased the fit of the model in relation to M0. The predictor number of phonemes did not improve the model fit. By contrast, the inclusion of the predictor number of syllables produces a bigger ΔBIC parameter in M3b (i.e., 66.81) than in M2 (i.e., 64.01). We thus kept the predictor NumSyll and estimated the roles of the two variables related to phonological neighborhood. Neither Phon_N nor Phon_N_MFreq improved the model fit of M3b. We then added to M3b the predictor familiarity, which did improve the model fit, so that M5 had a greater ΔBIC parameter than M3b.

In sum, the preliminary analysis on the control variables suggested influences of the predictors H statistic, number of syllables, and familiarity. Specifically, increasing H and the number of syllables increased the probability of experiencing a TOT, whereas increasing familiarity had the opposite effect (see Fig. 2). We then estimated the relevance of frequency and AoA effects. By adding the predictor frequency to M4b, the model fit did not improve. In contrast, the inclusion of the AoA variable increased the model fit, suggesting a relevant role of AoA. As in Experiment 1, the probability of experiencing a TOT increased with late-acquired words (see Fig. 2). It is worth noting here that the analyses showed no effects of frequency once familiarity was considered. Similarly, no effect of phonological neighborhood size emerged once length in syllables was added to the model (but see below, in the Further Tests of the AoA Effect section). However, given the higher levels of collinearity between familiarity and frequency and between string length and number of neighbors (e.g., Sadat, Martin, Costa, & Alario, 2014), further research will be needed to establish the precise relation and role of such variables in the TOT phenomenon (for further discussion, see Harley & Bown, 1998; Vitevitch & Sommers, 2003).

Further tests of the AoA effect In a final analysis, we further tested the reliability of the AoA effect in TOT states. We compared the null model (i.e., M0) against seven different models, each of which contained one of the control variables as the only predictor (i.e., all of the variables except AoA). The results showed that the models that improved the fit in comparison to M0 contained the following variables as predictors: H statistic, number of syllables, phonological neighborhood, familiarity, and frequency (see Table 5). These results reveal effects of each of these variables in the TOT pattern, indicating that, when considered individually, all of these factors affect the likelihood of TOT states. In particular, high naming agreement, shorter lengths, high-density phonological neighborhood, high levels of familiarity, and high frequency values seem all to have protective effects on TOT occurrence. We therefore compared a model simultaneously containing all these five variables as

Table 4 Fit indices from the analysis in Experiment 2

	Fixed Effects	Model *df*	Chisq (*df*)	*p*	BIC	ΔBIC	Approx. BF
M0		3			3,202		
M1	Item position	4	0.073 (1)	= .786	3,211	− 9.06	0.01
M2	*H*	4	73.14 (1)	< .001	3,138	64.01	> 10,000
M3a	*H* + NumPhon	5	78.28 (2)	< .001	3,142	60.01	> 10,000
M3b	*H* + NumSyll	5	85.08 (2)	< .001	3,135	66.81	> 10,000
M4a	*H* + NumSyll + Phon_N	6	87.32 (3)	< .001	3,142	59.91	> 10,000
M4b	*H* + NumSyll + Phon_N_MFreq	6	86.16 (3)	< .001	3,143	58.74	> 10,000
M5	*H* + NumSyll + Familiarity	6	186.08 (3)	< .001	3,043	158.65	> 10,000
M6	*H* + NumSyll + Familiarity+ LogFreq	7	186.45 (4)	< .001	3,052	149.89	> 10,000
M7	*H* + NumSyll + Familiarity + AoA	7	205.25 (4)	< .001	3,033	168.69	> 10,000

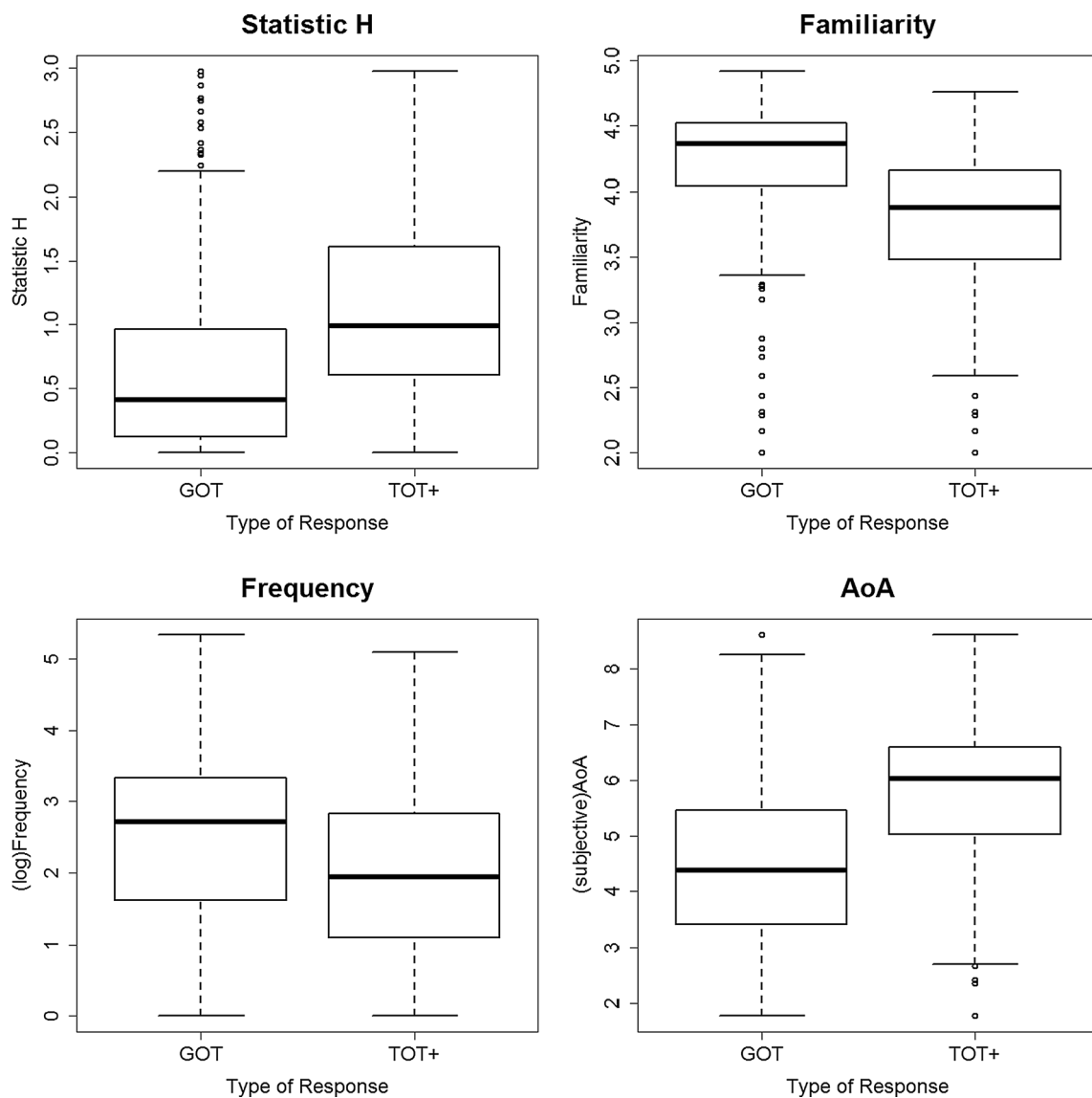For the fit statistics, see the note to Table 2.

Fig. 2 Distributions of the *H* statistic (upper left), familiarity (upper right), (log) frequency (lower left), and AoA (lower right) variables for GOT responses and TOT+ responses in Experiment 2

Table 5 Fit indices reporting the control variables that, as the only predictors, improved the fit in comparison to M0, along with the model containing all of these variables (i.e., MAll) and the model containing all of these variables plus the critical variable AoA (i.e., MAll_AoA) in Experiment 2 (see the main text for details)

	Fixed Effects	Model df	Chisq (df)	p	BIC	ΔBIC	Approx. BF
M0		3			3,202		
H	H	4	73.14 (1)	< .001	3,138	64.01	> 10,000
NumSyll	NumSyll	4	18.47 (1)	< .001	3,192	9.31	104
Phon_N	Phon_N	4	16.08 (1)	< .001	3,195	6.94	33
Familiarity	Familiarity	4	122.23 (1)	< .001	3,089	113.1	> 10,000
LogFreq	LogFreq	4	30.43 (1)	< .001	3,180	21.29	> 10,000
MAll	H + NumSyll + Phon_N + Familiarity + LogFreq	8	186.47 (5)	< .001	3,061	140.77	> 10,000
MAll_AoA	H + NumSyll + Phon_N + Familiarity + LogFreq + AoA	9	206.14 (6)	< .001	3,050	151.31	> 10,000

For the fit statistics, see the note to Table 2.

fixed effects (i.e., MAll) and a second model containing the five variables plus the critical variable AoA (i.e., MAll_AoA). The comparison between these two models showed that the model containing AoA improved the fit (see Table 5). This result confirms the AoA effect even when all relevant variables (i.e., predictors) are taken into consideration.

General discussion

In the present study, we assessed whether the probability of experiencing a TOT state depends on the age at which the word is acquired in life. Our findings suggest that this is the case. Specifically, we observed that participants suffered more TOT states with late-acquired than with early-acquired words. This effect persists even when the frequency variable is taken into account. Indeed, the statistical analysis showed that AoA was a better predictor of TOT states than frequency. This pattern was observed in two different tasks, picture naming (Exp. 1) and naming to definition (Exp. 2), and when the influences of other word property variables were taken in consideration. In addition, the results with subjective AoA measures (i.e., estimates provided by a group of adult participants) were identical to those observed with objective AoA measures (i.e., when children can name objects) in Experiment 1, replicating the high correlations observed between these two measures (e.g., Lotto et al., 2010; Morrison et al., 1997).

Furthermore, the two experiments reported here replicated previous studies identifying other relevant variables on the TOT phenomenon. For instance, pictures and definitions with less consistency in the words they elicit (i.e., higher H statistic scores) are more prone to TOT states, replicating previous observations with common nouns (Mitchell, 1989) and proper names (Bonin et al., 2008). We also observed that TOT probability increases with long words in the naming-to-definition task, consistent with Hanley and Chapman's (2008) study on proper names (for similar evidence with common nouns in a population of dyslexic children, see Hanly & Vandenberg, 2010). In addition, less familiar words tend to evoke more TOTs (for similar findings, see Burke et al., 1991). In spite of these observations, the most relevant observation of our experiments refers to the fact that AoA is the critical variable predicting the TOT phenomenon. AoA effects remained significant even when the effect of all other relevant variables were considered (e.g., name agreement, word length, neighborhood variables, familiarity, and frequency).

These results have relevant implications for psycholinguistic accounts of the TOT phenomenon. For instance, according to Burke and colleagues (1991), transmission deficits up to the phonological units are determined by three factors: the frequency of use of the word, the recent use of the word, and the age of the speaker. In relation to the first of these factors, TOT states would be more likely to be observed with low-frequency words because of their weaker connections to phonological units, in comparison to high-frequency words (see also Gollan, Montoya, Fennema-Notestine, & Morris, 2005). Our results are congruent with this interpretation. Nevertheless, our results show that AoA is a more reliable predictor than frequency. Under the assumption that early-acquired words have stronger and more consolidated connections within the speech production system (e.g., Ellis & Lambon Ralph, 2000), it might be more likely to observe failures during the transmission of information with late-acquired than with early-acquired words. In other words, the "advantage" for early-acquired words (i.e., they suffer fewer TOTs) would be explained by their strengthened semantic-to-phonology connections, in comparisons to words acquired later in life. We conclude here that in order to fully characterize the TOT phenomenon, theories will need to deal with the AoA variable.

The age in life at which the semantic-to-phonology connections are established for a word will determine the probability of retrieving that word in later years (e.g., in

adulthood). This raises the question of how vocabulary is acquired during childhood. Hills and collaborators have characterized vocabulary acquisition by exploiting the structure of growing networks (Hills, J. Maouene, Riordan, & Smith, 2010; Hills, M. Maouene, J. Maouene, Sheya, & Smith, 2009). The authors argued that vocabulary acquisition is mainly determined by the principle of preferential acquisition, which states that the order of acquisition of words is driven by the connectivity of the words in the learning environment. Children would thus learn first the most well-connected words in the learning environment to which they were exposed—that is, the words used more frequently in their social context. As a consequence of this learning process, these words would maintain an advantage over new words learned in the course of life. According to this approach, semantic access and word retrieval processes are presumably highly interconnected, and therefore this poses the problem of understanding to what extents AoA affects semantic processes, linguistic processes, or both. The aim of the present research was to test whether AoA affects linguistic retrieval by using a TOT paradigm that clearly tapped into such a process (see Catling & Johnston, 2009; Navarrete et al., 2013, for different methodological approaches exploring a similar issue). Given that the TOT phenomenon reflects failures during linguistic retrieval but not during semantic access, the AoA effect that we reported here should be (mainly) localized at cognitive processes involved in the retrieval of linguistic information (i.e., lexical and phonological units) during speech production. A prediction that follows from this reasoning is that highly proficient bilingual speakers who have acquired their second language in adulthood should show AoA effects in TOT states when speaking in their first language, but not when they speak in their second language. This is so because none of the words that they have learned in their second language could have been acquired early.

In the present research, we referred to a general architecture of the speech production system. Further research will be needed to localize at which level of the production system the transmission deficit in the TOT states occurs (i.e., early-TOT or late-TOT states, in the terms used by Gollan et al., 2014). However, some interesting hints emerge from the pattern of effects obtained in the present research. The fact that variables related to the phonological forms of words, such as phonological neighborhood and number of syllables, significantly predict TOT states suggests that TOTs can occur at the level of phonological encoding. In addition, the fact that variables related to the retrieval of lexical units, such as frequency and familiarity, also predict TOT states suggests that such states can also occur during lexical processing. This is consistent with the notion that there may be more than one type of TOT (see Gollan et al., 2014). Further

studies will be needed to explore whether AoA mainly affects one or several levels of processing, as was suggested by Catling and Johnston (2009; see also Navarrete et al., 2013). Evidence suggesting that it is more likely to localize AoA effects in the lexical-to-phonological mappings than in the semantic-to-lexical mappings derives from the study by Kittredge and colleagues (2008). In their analysis of the picture-naming errors of 50 aphasic patients, they showed that AoA predicted phonological errors but not semantic errors. Although the first type of error occurs during the retrieval of the phonological content of the words, semantic errors reflected failures during lexical retrieval. Under the assumption that lexical-to-phonological mappings are more arbitrary than semantic-to-lexical mappings (Dell & Kim, 2005), this pattern would be congruent with the notion that AoA effects arise mainly in tasks involving less arbitrary mappings (Lambon Ralph & Ehsan, 2006). Our study shows that AoA determines the transmission deficit during TOT states. Further studies will be needed to specify whether AoA has more influence on the semantic-to-lexical mappings or the lexical-to-phonological mappings.

Appendix A

Table 6 Properties of the experimental items used in Experiments 1 and 2: Mean values, standard deviations (SD), and range values

	Experiment 1			Experiment 2		
	Mean	SD	Range	Mean	SD	Range
AoA	4.4	1.3	1.5–7.5	4.8	1.5	1.8–8.6
Frequency	1.1	0.5	0–2.3	1.1	0.5	0–2.3
Familiarity	–	–	–	4.1	0.5	2–5
H	0.7	0.7	0–2.9	0.9	0.8	0–3
NumPhon	7.1	1.8	3–14	7	1.9	3–13
NumSyll	3	1	2–5	3	0.9	1–6
Phon_N	–	–	–	2.7	2.8	0–16
Phon_N_MFreq	–	–	–	1.6	1.2	0–4.6

AoA, age of acquisition; frequency, log Colfis total frequency; H, statistic H; NumPhon, number of phonemes; NumSyll, number of syllables; Phon_N, phonological neighborhood size; Phon_N_MFreq, summed neighborhood frequency. Frequency values were retrieved from the COLFIS database (which comprises 3,798,275 lexical occurrences; Bertinetto et al., 1995). Phonological neighborhood values (i.e., Phon_N, Phon_N_MFreq) were retrieved from the PhonItalia database (Goslin et al., 2014).

Appendix B

Table 7 List of the experimental items used in Experiments 1 and 2, with the numbers of TOTs that each item elicited

Italian Name	English Name	Used in Experiment	Number of TOT+ in Experiment 1	Number of TOT+ in Experiment 2
apparecchio	dental appliance	1	0	–
arco	arch	1	2	–
asparagi	asparagus	1	0	–
autobus	bus	1	0	–
balestra	crossbow	1	6	–
barca	boat	1	1	–
batteria	drum	1	0	–
broccolo	broccoli	1	1	–
cannella	cinnamon	1	2	–
carciofo	artichoke	1	1	–
castagna	chestnut	1	0	–
cervo	deer	1	0	–
fagiano	pheasant	1	2	–
falco	hawk	1	0	–
fenicottero	flamingo	1	0	–
fico	fig	1	0	–
fungo	mushroom	1	0	–
garofano	carnation	1	0	–
kiwi	kiwi	1	1	–
lampone	raspberry	1	1	–
libreria	bookcase	1	0	–
mango	mango	1	0	–
melograno	pomegranate	1	2	–
molletta	clothespin	1	0	–
mora	blackberry	1	0	–
orecchini	earrings	1	0	–
paletta	scoop	1	4	–
pavone	peacock	1	3	–
pellicano	pelican	1	4	–
picchio	woodpecker	1	2	–
pisello	pea	1	0	–
rapa	turnip	1	1	–
rondine	swallow	1	0	–
roulotte	caravan	1	3	–
tenaglia	pincer	1	3	–
altalena	swing	1, 2	0	0
amaca	hammock	1, 2	1	3
antenna	antenna	1, 2	1	1
arpa	harp	1, 2	1	0
ascia	ax	1, 2	1	3
avvoltoio	vulture	1, 2	3	7
bastone	cane	1, 2	0	0
biberon	feeding bottle	1, 2	0	0
bilancia	scales	1, 2	0	0
binocolo	binoculars	1, 2	0	3
birillo	skittle	1, 2	0	1

Table 7 (continued)

Italian Name	English Name	Used in Experiment	Number of TOT+ in Experiment 1	Number of TOT+ in Experiment 2
bomba	bomb	1, 2	0	0
botte	barrel	1, 2	0	3
bussola	compass	1, 2	2	0
cacciavite	screwdriver	1, 2	1	1
camino	fireplace	1, 2	0	0
campanile	bell tower	1, 2	1	1
camper	camper	1, 2	0	1
cannone	cannon	1, 2	0	0
canottiera	vest	1, 2	0	0
caramella	candy	1, 2	0	0
carriola	wheelbarrow	1, 2	2	1
catapulta	catapult	1, 2	7	8
cavalletto	easel	1, 2	2	8
caviale	caviar	1, 2	1	4
cerniera	zipper	1, 2	0	0
chiodo	(metal) nail	1, 2	1	1
cicogna	stork	1, 2	2	2
ciuccio	pacifier	1, 2	0	1
civetta	little owl	1, 2	0	0
clessidra	hourglass	1, 2	0	1
colomba	dove	1, 2	0	1
coperchio	cover	1, 2	0	0
cornamusa	bagpipe	1, 2	2	10
culla	cradle	1, 2	0	0
cupola	cupola	1, 2	1	1
damigiana	demijohn	1, 2	0	9
deserto	desert	1, 2	0	0
diga	dam	1, 2	0	1
dirigibile	airship	1, 2	5	8
elicottero	helicopter	1, 2	0	0
falce	sickle	1, 2	4	0
faro	lighthouse	1, 2	1	0
fiammifero	match	1, 2	0	0
fionda	slingshot	1, 2	2	2
fisarmonica	accordion	1, 2	1	2
flauto	flute	1, 2	0	0
forBICe	scissor	1, 2	0	0
freccia	arrow	1, 2	0	0
frusta	whip	1, 2	1	1
gabbia	cage	1, 2	0	0
gabbiano	seagull	1, 2	3	2
ghianda	acorn	1, 2	0	3
gilet	vest	1, 2	0	2
gondola	gondola	1, 2	0	2
grattugia	grater	1, 2	0	2
gufo	owl	1, 2	0	0
imbuto	funnel	1, 2	0	1
incenso	incense	1, 2	1	3
lavagna	blackboard	1, 2	0	0

Table 7 (continued)

Italian Name	English Name	Used in Experiment	Number of TOT+ in Experiment 1	Number of TOT+ in Experiment 2
magazzino	warehouse	1, 2	1	1
mandolino	mandolin	1, 2	2	8
manette	handcuffs	1, 2	0	0
mantello	cloak	1, 2	0	0
mappa	map	1, 2	1	0
mascara	mascara	1, 2	0	0
mattarello	rolling pin	1, 2	3	1
mazza	bat	1, 2	1	3
mestolo	ladle	1, 2	1	0
microscopio	microscope	1, 2	0	3
mitragliatrice	machine gun	1, 2	0	2
mongolfiera	hot-air balloon	1, 2	0	1
nacchere	castanets	1, 2	3	11
oca	goose	1, 2	0	3
organo	organ	1, 2	1	3
pannocchia	corncob	1, 2	0	1
passeggino	stroller	1, 2	0	1
pennello	brush	1, 2	0	0
pentola	cooking pot	1, 2	0	0
pergamena	parchment	1, 2	1	1
pettine	comb	1, 2	0	0
piramide	pyramid	1, 2	0	1
portico	portico	1, 2	0	2
pozzo	well	1, 2	0	1
puzzola	skunk	1, 2	1	1
rastrello	rake	1, 2	0	3
remo	oar	1, 2	0	2
rubinetto	tap	1, 2	0	4
sandali	sandals	1, 2	1	0
sassofono	saxophone	1, 2	0	2
scalinata	flight of steps	1, 2	0	0
scivolo	slide	1, 2	0	0
secchio	bucket	1, 2	0	2
sega	saw	1, 2	0	1
sgabello	stool	1, 2	0	0
sigaro	cigar	1, 2	0	0
slitta	sledge	1, 2	0	0
sottomarino	submarine	1, 2	1	3
struzzo	ostrich	1, 2	2	1
tamburo	drum	1, 2	0	0
tanica	jerry can	1, 2	1	6
tenda	tent	1, 2	0	0
timone	rudder	1, 2	2	6
torre	tower	1, 2	0	0
trapano	drill	1, 2	0	4
trattore	tractor	1, 2	1	0
trave	beam	1, 2	1	1
triciclo	tricycle	1, 2	0	0
tromba	trumpet	1, 2	0	0

Table 7 (continued)

Italian Name	English Name	Used in Experiment	Number of TOT+ in Experiment 1	Number of TOT+ in Experiment 2
tuorlo	yolk	1, 2	0	0
ventaglio	fan	1, 2	2	0
zattera	raft	1, 2	0	0
zoccoli	clogs	1, 2	0	2
acquario	aquarium	2	_	1
agrume	citrus	2	_	4
ancora	anchor	2	_	1
anemometro	anemometer	2	_	4
aratro	plow	2	_	3
arcobaleno	rainbow	2	_	1
bambola	doll	2	_	0
batuffolo	wad cotton	2	_	1
bavaglino	bib	2	_	1
biga	(roman) chariot	2	_	6
biglia	ball	2	_	0
boia	executioner	2	_	3
burattino	puppet	2	_	8
burrasca	storm	2	_	5
calcagno	heel	2	_	4
caleidoscopio	kaleidoscope	2	_	11
calice	chalice	2	_	4
capitello	capital	2	_	3
cappuccio	hood	2	_	0
carcassa	carcass	2	_	3
carillon	carillon	2	_	3
catrame	tar	2	_	9
cavaliere	knight	2	_	0
cavatappi	corkscrew	2	_	5
cenere	ash	2	_	1
centauro	centaur	2	_	17
cerotto	band aid	2	_	0
cinghiale	boar	2	_	0
cipria	powder	2	_	0
circo	circus	2	_	0
cisterna	cistern	2	_	0
cocchio	chariot	2	_	2
cofano	car bonnet	2	_	6
corallo	coral	2	_	1
corno	horn	2	_	2
cornucopia	cornucopia	2	_	3
corona	crown	2	_	0
crisma	chrism	2	_	4
cucciolo	puppy	2	_	0
curaro	curare	2	_	1
dinamite	dynamite	2	_	4
dinosauro	dinosaur	2	_	0
drago	dragon	2	_	0
elica	propeller	2	_	2
elmo	helmet	2	_	1

Table 7 (continued)

Italian Name	English Name	Used in Experiment	Number of TOT+ in Experiment 1	Number of TOT+ in Experiment 2
ematoma	hematoma	2	–	2
fata	fairy	2	–	2
fibbia	buckle	2	–	3
flanella	flannel	2	–	0
forfora	dandruff	2	–	0
frullatore	mixer	2	–	4
genero	son-in-law	2	–	5
genio	genie	2	–	0
geroglifico	hieroglyph	2	–	11
giostra	carousel	2	–	0
giradischi	record player	2	–	8
gobba	hunchback	2	–	0
graffiti	graffiti	2	–	4
incantesimo	spell	2	–	3
lapide	tombstone	2	–	0
leggio	lectern	2	–	8
locomotiva	locomotive	2	–	6
mancia	tip	2	–	0
marionetta	marionette	2	–	2
maschera	mask	2	–	0
mausoleo	mausoleum	2	–	7
metronomo	metronome	2	–	17
mitra	miter	2	–	6
molecola	molecule	2	–	2
muffa	mold	2	–	1
nano	dwarf	2	–	0
nuora	daughter-in-law	2	–	2
origami	origami	2	–	4
orma	footprint	2	–	3
orto	vegetable garden	2	–	0
papiro	papyrus	2	–	4
paracadute	parachute	2	–	1
pedone	pawn	2	–	2
petardo	firecracker	2	–	0
plastilina	plasticine	2	–	1
podio	podium	2	–	1
pompiere	fireman	2	–	1
pulpito	pulpit	2	–	10
racchetta	racket	2	–	0
ragnatela	cobweb	2	–	4
renna	reindeer	2	–	0
residuo	residue	2	–	1
robot	robot	2	–	0
rotaia	rail	2	–	4
rotolo	roll	2	–	2
rublo	ruble	2	–	2
ruggine	rust	2	–	1
salvadanaio	money box	2	–	0
satellite	satellite	2	–	3

Table 7 (continued)

Italian Name	English Name	Used in Experiment	Number of TOT+ in Experiment 1	Number of TOT+ in Experiment 2
sciabola	sabre	2	–	2
serra	greenhouse	2	–	1
sestante	sextant	2	–	3
sfinge	sphinx	2	–	2
silo	silo	2	–	3
sirena	mermaid	2	–	0
spartito	sheet music	2	–	5
strega	witch	2	–	0
sumo	sumo	2	–	6
tarocco	tarot	2	–	2
tessera	puzzle piece	2	–	2
tic	tic	2	–	0
tombino	manhole	2	–	0
trotto	trot	2	–	3
urna	urn	2	–	7
velcro	Velcro	2	–	5
zolla	clod	2	–	0

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

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

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,* 390–412. doi:10.1016/j.jml.2007.12.005

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

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,* 255–278. doi:10.1016/j.jml.2012.11.001

Bates, E., Burani, C., D'Amico, S., & Barca, L. (2001). Word reading and picture naming in Italian. *Memory & Cognition, 29,* 986–999. doi:10.3758/BF03195761

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes [Software] (R package version 0.999375-39). Retrieved from http://CRAN.R-project.org/package=lme4

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical-semantic competition hypothesis. *Cognition, 96,* B45–B54.

Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., & Thornton, A.M. (1995). CoLFIS (Corpus e lessico di frequenza dell'Italiano scritto) [Corpus and frequency lexicon of written Italian]. Retrieved from the Institute of Cognitive Sciences and Technologies, www.istc.cnr.it/grouppage/colfisEng

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

Bonin, P., Perret, C., Méot, A., Ferrand, L., & Mermillod, M. (2008). Psycholinguistic norms and face naming times for photographs of celebrities in French. *Behavior Research Methods, 40,* 137–146. doi:10.3758/BRM.40.1.137

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109,* 204–223. doi:10.1037/0033-2909.109.2.204

Brown, A. S. (2012). *The tip of the tongue state.* Hove, UK: Psychology Press.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325–337. doi:10.1016/S0022-5371(66)80040-3

Brown, G. D. A., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition, 15,* 208–216. doi:10.3758/BF03197718

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition, 13,* 992–1011. doi:10.1080/13506280544000165

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30,* 542–579.

Burke, D. M., & Shafto, M. A. (2004). Aging and language production. *Current Directions in Psychological Science, 13,* 21–24.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14,* 177–208.

Catling, J. C., & Johnston, R. A. (2009). The varying effects of age of acquisition. *Quarterly Journal of Experimental Psychology, 62,* 50–62. doi:10.1080/17470210701814352

Chalard, M., Bonin, P., Méot, A., Boyer, B., & Fayol, M. (2003). Objective age-of-acquisition (AoA) norms for a set of 230 object names in French: Relationships with psycholinguistic variables, the English data from Morrison, et al. 1997 and naming latencies. *European Journal of Cognitive Psychology, 15,* 209–245.

Cuetos, F., Samartino, T., & Ellis, A. W. (2012). Age acquisition norms from elderly Spanish people: Characteristics and the prediction of word recognition performance in Alzheimer's disease. *Psicológica, 33,* 59–76. Retrieved from www.uv.es/psicologica/articulos1.12/4CUETOS.pdf

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S., & Kim, A. E. (2005). Speech errors and word form encoding. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 17–41). Hove, UK: Psychology Press.

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

Development Core Team, R. (2011). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1103–1123. doi:10.1037/0278-7393.26.5.1103

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

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

Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 267–283. doi:10.1037/0278-7393.24.2.267

Gollan, T. H., & Acenas, L.-A. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish–English and Tagalog–English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 246–269. doi:10.1037/0278-7393.30.1.246

Gollan, T. H., Bonanni, M. P., & Montoya, R. I. (2005a). Proper names get stuck on bilingual and monolingual speakers' tip-of-the-tongue equally often. *Neuropsychology, 19,* 278–287.

Gollan, T. H., Ferreira, V. S., Cera, C., & Flett, S. (2014). Translation-priming effects on tip-of-the-tongue states. *Language, Cognition and Neuroscience, 29,* 278–288. doi:10.1080/01690965.2012.762457

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005b). Bilingualism affects picture naming but not picture classification. *Memory & Cognition, 33,* 1220–1234. doi:10.3758/BF03193224

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition, 4,* 63–83. doi:10.1017/S136672890100013X

Goslin, J., Galluzzi, C., & Romani, C. (2014). PhonItalia: A phonological lexicon for Italian. *Behavior Research Methods, 46,* 872–886. doi:10.3758/s13428-013-0400-8

Hanley, J. R., & Chapman, E. (2008). Partial knowledge in a tip-of-the-tongue state about two- and three-word proper names. *Psychonomic Bulletin & Review, 15,* 156–160. doi:10.3758/PBR.15.1.156

Hanly, S., & Vandenberg, B. (2010). Tip-of-the-tongue and word retrieval deficits in dyslexia. *Journal of Learning Disabilities, 43,* 15–23.

Harley, T. A., & Bown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology, 89,* 151–174.

Hernandez, A. E., & Fiebach, C. J. (2006). The brain bases of reading late learned words: Evidence from functional MRI. *Visual Cognition, 13,* 1027–1043.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science, 20,* 729–739. doi:10.1111/j.1467-9280.2009.02365.x

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language, 63,* 259–273.

Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition, 13,* 884–910.

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition, 13,* 789–845. doi:10.1080/13506280544000066

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin, 131,* 684–712. doi:10.1037/0033-2909.131.5.684

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). In which is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology, 25,* 463–492. doi:10.1080/02643290701674851

Lambon Ralph, M. A., & Ehsan, S. (2006). Age of acquisition effects depend on the mapping between representations and the frequency of occurrence: Empirical and computational evidence. *Visual Cognition, 13,* 928–948.

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.

Lotto, L., Surian, L., & Job, R. (2010). Objective age of acquisition for 223 Italian words: Norms and effects on picture naming speed. *Behavior Research Methods, 42,* 126–133. doi:10.3758/BRM.42.1.126

Mermillod, M., Bonin, P., Méot, A., Ferrand, L., & Paindavoine, M. (2012). Computational evidence that frequency trajectory theory does not oppose but emerges from age-of-acquisition theory. *Cognitive Science, 36,* 1499–1531.

Miozzo, M., & Caramazza, A. (1997). Retrieval of lexical–syntactic features in tip-of-the tongue states. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1410–1423. doi:10.1037/0278-7393.23.6.1410

Mitchell, D. B. (1989). How many memory systems? Evidence from aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 31. doi:10.1037/0278-7393.15.1.31

Monaghan, J., & Ellis, A. W. (2002). Age of acquisition and the completeness of phonological representations. *Reading and Writing, 15,* 759–788.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology, 50A,* 528–559. doi:10.1080/027249897392017

Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B. Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language, 76,* 253–272.

Navarrete, E., Scaltritti, M., Mulatti, C., & Peressotti, F. (2013). Age-of-acquisition effects in delayed picture-naming tasks. *Psychonomic Bulletin & Review, 20,* 148–153. doi:10.3758/s13423-012-0310-2

Pérez, M. A. (2007). Age of acquisition persists as the main factor in picture naming when cumulative word frequency and frequency trajectory are controlled. *Quarterly Journal of Experimental Psychology, 60,* 32–42. doi:10.1080/17470210600577423

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

Pyers, J. E., Gollan, T. H., & Emmorey, K. (2009). Bimodal bilinguals reveal the source of tip-of-the-tongue states. *Cognition, 112,* 323–329.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology 1995* (pp. 111–163). Cambridge, MA: Blackwell.

Sadat, J., Martin, C. D., Costa, A., & Alario, F. X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology, 68,* 33–58.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464. doi:10.1214/aos/1176344136

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

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science, 8,* 314–317.

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition, 31,* 491–504. doi:10.3758/BF03196091

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47,* 1–29. doi:10.1006/jmla.2001.2834