The Quarterly Journal of Experimental Psychology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje20

Reading in Spanish and Italian: Effects of age of acquisition in transparent orthographies?

Rob Davies\textsuperscript{a}, Maximiliano Wilson\textsuperscript{bc}, Fernando Cuetos\textsuperscript{d} & Cristina Burani\textsuperscript{c}

\textsuperscript{a} Department of Psychology, Social Work and Public Health, Oxford Brookes University, Oxford, UK
\textsuperscript{b} Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec and Département de Réadaptation, Université Laval, Quebec City, Quebec, Canada
\textsuperscript{c} Institute for Cognitive Sciences and Technologies, National Research Council, Rome, Italy
\textsuperscript{d} Faculty of Psychology, University of Oviedo, Oviedo, Spain

Accepted author version posted online: 24 Jan 2014. Published online: 14 Mar 2014.

To cite this article: Rob Davies, Maximiliano Wilson, Fernando Cuetos & Cristina Burani (2014) Reading in Spanish and Italian: Effects of age of acquisition in transparent orthographies?, The Quarterly Journal of Experimental Psychology, 67:9, 1808-1825, DOI: 10.1080/17470218.2013.872155

To link to this article: http://dx.doi.org/10.1080/17470218.2013.872155

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or
Reading in Spanish and Italian: Effects of age of acquisition in transparent orthographies?

Rob Davies¹, Maximiliano Wilson²,³, Fernando Cuetos⁴, and Cristina Burani³

¹Department of Psychology, Social Work and Public Health, Oxford Brookes University, Oxford, UK
²Centre de Recherche de l’Institut Universitaire en Santé Mentale de Québec and Département de Réadaptation, Université Laval, Quebec City, Quebec, Canada
³Institute for Cognitive Sciences and Technologies, National Research Council, Rome, Italy
⁴Faculty of Psychology, University of Oviedo, Oviedo, Spain

Despite the similar transparency of their orthographies, reading in Italian has been found to be affected by frequency but not age of acquisition (AoA) [Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. Behaviour Research Methods, Instruments and Computers, 34, 424–434] while reading in Spanish is affected by AoA but not frequency [Cuetos, F., & Barbón, A. (2006). Word naming in Spanish. European Journal of Cognitive Psychology, 18, 415–436]. We examined this cross-linguistic difference, firstly, through a reanalysis of the Italian and Spanish reading latencies. After eliminating several between-experiment differences, we replicated the AoA effect in Spanish but not in Italian and the frequency effect in Italian but not in Spanish. The cross-linguistic comparison could not equate stimulus imageability; therefore, secondly, we compared the Italian reading latencies with new Spanish reading latencies for imageability-matched words. We found frequency effects but neither an AoA effect nor a language by AoA interaction. We argue that the previously reported cross-linguistic difference in the AoA effect resulted from a between-study difference in stimulus imageability. More imageable words induced more semantic involvement in reading, yielding an AoA effect in Spanish.

Keywords: Orthography; Age of acquisition; Frequency; Transparent; Reading.

Words encountered later in life are normally processed more slowly and less accurately than those encountered earlier (e.g., Cortese & Khanna, 2007; Gilhooly & Logie, 1981; Morrison & Ellis, 1995; see reviews by Johnston & Barry, 2006; Juhasz, 2005). This age of acquisition (AoA) effect has been shown to be larger for words with inconsistent spelling–sound mappings and smaller or absent for words with consistent mappings in reading aloud in English (J. Monaghan & Ellis, 2002) and Italian (Wilson, Ellis, & Burani, 2012). The modulation of the AoA effect by the relative spelling–sound consistency of different words within a language leads to the expectation of a similar conditioning of AoA by the relative consistency of words in between-language comparisons. Recent observations in Spanish and Italian do not match that expectation, however.

Correspondence should be addressed to Rob Davies, Department of Psychology, Fylde College, Lancaster University, Bailrigg, Lancaster, LA1 4YF, UK. E-mail: r.davies1@lancaster.ac.uk

We thank Padraic Monaghan and two anonymous reviewers for their very helpful comments.

This work was supported by an EU Sixth Framework Marie Curie Research Training Network (MRTN-CT-2004-512141): Language and Brain (RTN: LAB, http://www.hull.ac.uk/RTN-LAB/).
These languages have orthographies with similarly high spelling–sound consistency across words, yet the AoA effect has been found in word naming in Spanish (Cuetos & Barbón, 2006) but not in Italian when words respected the stress distribution of the language (Barca, Burani, & Arduino, 2002). This cross-linguistic difference should have important implications for our understanding of when AoA effects can be observed, but the between-language comparison is confounded by between-study differences. In the present study, we attempted to uncover what aspects of the mappings involved in reading actually condition the observation of AoA effects in Spanish and Italian.

A substantial body of empirical research supports the assumption of AoA or age-limited learning effects in reading. Earlier acquired words typically elicit faster, more accurate responses than late-acquired words in reading tasks like word naming (e.g., in English, Brown & Watson, 1987; Coltheart, Laxon, & Keating, 1988; Cortese & Khanna, 2007; Ellis & Monaghan, 2002; Gerhand & Barry, 1998; Gilhooly & Logie, 1981; J. Monaghan & Ellis, 2002; Morrison & Ellis, 1995, 2000; Stadthagen-Gonzalez, Bowers, & Damian, 2004; in Spanish, Cuetos & Barbón, 2006; in Turkish, Raman, 2006) and lexical decision (e.g., in Dutch, Brysbaert, Lange, & Van Wijnendaele, 2000; in English, Cortese & Khanna, 2007; Gerhand & Barry, 1999; Morrison & Ellis, 1995, 2000; Stadthagen-Gonzalez et al., 2004; in French, Bonin, Chalard, Méot, & Fayol, 2002; in Italian, Burani, Arduino, & Barca, 2007; Colombó & Burani, 2002). While AoA effects are thus well established experimentally, there has been considerable debate over their interpretation.

Two kinds of concern have been raised. The first is that rated AoA correlates highly with or is predicted by a number of other key psycholinguistic variables (Álvarez & Cuetos, 2007; Barbarotto, Laiacona, & Capitani, 2005; Chalard, Bonin, Méot, Boyer, & Fayol, 2003; Lotto, Surian, & Job, 2010; Morrison, Chappell, & Ellis, 1997; Pérez & Navalón, 2005; Pind, Jónsdóttir, Gissurardóttir, & Jónsson, 2000; see, also, Baayen, Feldman, & Schreuder, 2006; Bonin, Barry, Méot, & Chalard, 2004; Cortese & Khanna, 2007; Zevin & Seidenberg, 2002, 2004), including word imageability, frequency, and length. The implication of such correlations is that rated AoA is a composite variable whose predictive power in analyses might be owed to its overlap with those other variables.

The second concern is that AoA values reflect an outcome (Zevin & Seidenberg, 2002): the age at which a word is learnt, on average. In most of the studies cited in the foregoing, researchers have used rated AoA in analyses of reading performance. Such ratings have been characterized as subjective but have been found to correlate highly with objective estimates of AoA that can be derived by examining the age at which children are capable of correctly naming pictured objects (Morrison et al., 1997). The objective AoA recorded in Morrison et al. (1997) and similar studies is, explicitly, a measure of children’s picture naming performance. The problem introduced by an outcome interpretation of AoA measures is that their effect can be taken to reflect the influence of those variables that determine the relative ease with which words can be learnt (Zevin & Seidenberg, 2002, 2004; see, also, Bonin et al., 2004; Bonin, Méot, Mermillod, Ferrand, & Barry, 2009; Mermillod, Bonin, Méot, Ferrand, & Paindavoine, 2012).

The first concern has been effectively addressed, we think, in recent observations reported by Cortese and Khanna (2007). These authors observed an independent effect of rated AoA on word naming and lexical decision latencies in regression analyses that also took into account the contribution of other factors, including word frequency, imageability, and length. Cortese and Khanna (2007) argued that if the effect of AoA is found to explain variance in reading behaviour above and beyond the contribution of other correlated factors then that AoA effect should be considered to contribute unique information to our account of reading. The second concern, that AoA is an outcome variable, will be considered in relation to findings from a series of computational simulations of increasing refinement.

Ellis and Lambon Ralph (2000) showed that neural networks trained to associate input with
output patterns will evidence the effect of the order in which patterns are introduced in training. If patterns were introduced early in training, and if they continued to be presented even as other patterns were introduced later, the network produced an output closer to the target for early- than for late-introduced patterns although the total number of presentations was equated. The persistent early advantage was found to stem from the decreasing plasticity of the neural network. As the accumulation of experience drove the adaptation of connection weights to suit the input–output mappings for early items, and as space for change in connection weights approached their limits, the network became committed to the input–output patterns it had experienced, and it became harder for later items to be learnt as well.

The Ellis and Lambon Ralph (2000) simulations deployed input–output mappings where units varied at random. This arbitrariness and the unpredictability of arbitrary mappings were argued by Zevin and Seidenberg (2002; see also J. Monaghan & Ellis, 2002) to be critical to the observation of AoA effects. Zevin and Seidenberg (2002) manipulated the cumulative frequency of items and, to implement AoA, the distribution of frequency over time (the frequency trajectory). Critically, they did so for representations of mappings designed to resemble the quasiregular orthography-to-phonology mappings of English. They found that words encountered frequently in early training were learned early, supporting the view that AoA can be understood as an outcome of frequency of encounter. Where early and late orthography-to-phonology mappings were similar (as in sets of consistently pronounced words), the early advantage diminished over time because what was learned about early items carried over to the processing of late items. Where the overlap between mappings was removed, however, the advantage associated with early high frequency endured because what was learned about early items did not carry over to late items. The latter had to be learned by rote, and that rote learning requirement, given decreasing network plasticity, caused an early item advantage.

P. Monaghan and Ellis (2010) argued that Zevin and Seidenberg (2002) had not observed a frequency trajectory effect because late items in their simulations had received sufficient but unrealistic early exposure to shape connection weights, while early items, encountered rarely later in training, had been subject to forgetting. In their model of reading development, P. Monaghan and Ellis (2010) entered items into training in a staged fashion according to their frequency of occurrence in texts ordered by school grade. After introduction, an item’s frequency in training over time was then made to follow the variation in its grade-by-grade frequency. The tailoring of the training regime to a realistic estimation of experience over time proved critical. P. Monaghan and Ellis (2010) showed that an AoA effect could be observed with spelling–sound mappings resembling those in English, given a realistic training regime. This AoA effect was found to be independent both of cumulative frequency and of frequency trajectory, the factors for which word learning might be an outcome.

In the P. Monaghan and Ellis (2010; see also Mermillod et al., 2012) simulations, the AoA effect was larger for items with inconsistent mappings than for items with consistent mappings. The modulation of the AoA effect by mapping consistency is broadly in line with the results of previous simulations (Ellis & Lambon Ralph, 2000; J. Monaghan & Ellis, 2002; Smith, Cottrell, & Anderson, 2001; Zevin & Seidenberg, 2002) where the AoA effect was larger if late items could not be predicted from information about early items. It is theoretically significant, then, that Bonin and colleagues observed an effect of frequency trajectory in lexical decision (Bonin et al., 2004) and in picture naming (Bonin et al., 2009), while Cortese and Khanna (2007) found a stronger AoA effect in lexical decision than in word naming. This is because performance is argued to involve the processing of mappings to semantics in lexical decision (e.g., Cortese & Khanna, 2007; Plaut, 1997) or from semantics in picture naming (e.g., Levelt, Roelofs, & Meyer, 1999), and lexical mappings to or from semantics are largely arbitrary (Zevin & Seidenberg, 2002). While the
P. Monaghan and Ellis (2010) results reflect AoA effects in a model of reading equipped only with orthography-to-phonology (O–P) mappings, simulation results reported by Lambon Ralph and Ehsan (2006) and by Mermillod et al. (2012) provide evidence that AoA or frequency trajectory effects are more prominent in tasks involving arbitrary mappings akin to mappings to or from semantics.

A semantic location for the AoA effect might stem not just from the arbitrariness of mappings involving semantics. Steyvers and Tenenbaum (2005) observed that AoA appears to shape the structure of semantic networks. Semantic knowledge is assumed to develop in a gradual fashion, with later acquired concepts added on to earlier acquired concepts. Given the potential role of AoA in structuring semantic knowledge, the evidence reviewed suggests that the involvement of semantics will entail stronger AoA effects in lexical tasks.

As the effect of AoA in reading seems to be modulated by spelling–sound consistency, one can predict that the size of the AoA effect will be different across languages, according to the relative transparency of their scripts: (a) The AoA effect in reading will be smaller in transparent scripts with almost fully consistent O–P mappings than in those with less consistent orthographies; (b) the size of the AoA effect will be similar in similarly consistent orthographies. Spanish and Italian, the languages investigated in the present study, have very transparent O–P mappings. It is thus predicted that the effects of AoA in reading aloud should be small or absent in both languages. This is not what has been found. In Italian, with the exception of one study in which AoA effects were found only when using irregularly stressed words (Wilson et al., 2012), a frequency effect but no AoA effect has been repeatedly observed in reading aloud (Barca et al., 2002, BBA02; Bates, Burani, D’Amico, & Barca, 2001; Burani et al., 2007). Conversely, an AoA effect but no frequency effect has been observed in the similarly consistent orthography, Spanish (Cuetos & Barbón, 2006, CB06). This Italian (BBA02)–Spanish (CB06) contrast is puzzling because it arguably calls into question the assumption that AoA effects are conditioned by the predictability of the mappings involved in lexical tasks. It is therefore of theoretical interest to explore why in two similarly consistent orthographies the effects of frequency and AoA were not also similar.

THE PRESENT INVESTIGATION

We report two studies. In the first, we reanalysed the reading latencies recorded in CB06 and BBA02. In Experiment 2, we compared the BBA02 reading data with latencies for a new set of Spanish words matched on several psycholinguistic variables.

Both the BBA02 and the CB06 studies employed regression analyses to examine the influence of several variables on the word naming of university students. However, besides the difference in language, there were differences in the kinds of words presented and in the choices made at analysis. First, while CB06 entered, as regression predictors, the raw variables corresponding to the attributes of their stimuli (e.g., the rated AoA or estimated frequency of words), BBA02 entered latent factors derived by principal components analysis (PCA) of the raw variables. Secondly, CB06 took into account the phonetic characteristics of word initials (an important source of variance, Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004) by entering dummy variables to code for phonetic features, while BBA02 controlled for the phonetic characteristics of initials by excluding data pertaining to a subset of their items (110 out of 626 words).

These differences, in combination, might have been responsible for the between-language difference in the AoA effect. This possibility was examined in Experiment 1 by removing the differences through an analysis that would take the same approach to both languages. We evaluated whether the Spanish–Italian difference in the size of the AoA effect would be replicated in a joint reanalysis of data from both languages using the same predictors and the same analytic strategy.
The results of Experiment 1 directed attention to another between-study difference: stimulus choice. A salient difference between the stimulus sets is that the Spanish words were highly imageable object names while the Italian words included nouns sampled from across the range of rated imageability. Highly imageable words are argued to have richer semantic representations or to activate more semantic features than abstract words (e.g., Jones, 1985). Greater semantic activation has been associated with faster naming (Balota et al., 2004; Pexman, Lupker, & Hino, 2002; Wilson, Cueto, Davies, & Burani, 2013), and this association has been captured, in some analyses, by the effect of imageability on word naming (Balota et al., 2004). These observations predict faster reading latencies for the Spanish words with higher imageability values than for the Italian words. However, imageability could have shaped not just overall reading speed but also the AoA effect.

That imageability can modulate the AoA effect follows from the assumption that the frequency-independent AoA effect is linked to mappings involving semantics (Bates et al., 2001; Brysbaert & Ghyselinck, 2006; Mermillod et al., 2012; Steyvers & Tenenbaum, 2005). If it is assumed that semantic information is activated early in a cascaded reading process and that it contributes to phonological coding in word naming (Balota et al., 2004) then this contribution is likely to be especially significant for highly imageable words, rich in semantic content (Pexman et al., 2002). However, Cortese and Khanna (2007) have shown that if imageability and AoA are entered simultaneously with other variables in a regression analysis of word naming latencies, the semantic influence on reading will be seen as an AoA effect observed with a null imageability effect. Thus a reading process which relies more on semantics could be expected to reveal stronger AoA effects, as has been shown in Spanish word naming by Wilson et al. (2013).

Experiment 2 investigated whether the kind of words presented by CB06 and BBA02 were critically different and therefore processed differently. We hypothesized that if that were the case then controlling for the imageability of items would eliminate the interaction between language and AoA. Therefore in Experiment 2 we analysed the BBA02 data together with data from a new study of word naming using Spanish words matched to the BBA02 words on imageability.

EXPERIMENT 1

Method

The BBA02 and CB06 papers report methodologically similar investigations. Participants were native language speakers, students attending local universities. BBA02 tested 30 participants aged 20–30 years (15 males, 15 females). CB06 tested 53 participants aged 18–23 years (12 males, 41 females). Word naming latencies were recorded using voice keys linked to PCs. BBA02 participants read 626 words whereas CB06 participants read 240 words.

Stimuli

Our analyses employed word attribute variables that had been used as predictors in both previous studies: adult written frequency (in Spanish, the Spanish Lexicon (LEXESP) database, Sebastián, Martí, Carreiras, & Cueto, 2000; in Italian, the Corpus and Frequency Lexicon of Written Italian (CoLFIS) database, Bertinetto et al., 2011); child written frequency (in Spanish, Martínez & García, 2004; in Italian, Marconi, Ott, Pesenti, Ratti, & Tavella, 1993); word length in letters (henceforth, length); orthographic neighbourhood size (N-size, see CB06 and BBA02 for details); and rated AoA and imageability, both obtained using 7-point scales and equivalent instructions (see CB06 and BBA02 for details). We report a summary of item characteristics in Table 1.

To minimize skew, we transformed to log10(x + 1) the child frequency, adult frequency, and N-size variables. The CB06 Spanish words are more frequent, shorter, more imageable, and earlier acquired and possess more orthographic neighbours than the BBA02 Italian words (Table 1).

Our aim was to estimate the extent to which the effects of critical psycholinguistic variables differed
Table 1. Summary of item characteristics in each language for Experiment 1 and Experiment 2 with between-languages comparisons of mean values for each variable

<table>
<thead>
<tr>
<th>Variable</th>
<th>Experiment 1</th>
<th></th>
<th></th>
<th></th>
<th>Experiment 2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Both languages</td>
<td>Spanish</td>
<td>Italian</td>
<td></td>
<td>Both languages</td>
<td>Spanish</td>
<td>Italian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>866 items</td>
<td>240 items</td>
<td>626 items</td>
<td></td>
<td>1252 items</td>
<td>626 items</td>
<td>626 items</td>
<td></td>
</tr>
<tr>
<td>Log₁₀(adult frequency per million + 1)</td>
<td>1.14 0.57</td>
<td>1.22 0.48</td>
<td>*</td>
<td>1.13 0.52</td>
<td>1.15 0.42</td>
<td>1.12 0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log₁₀(child frequency per million + 1)</td>
<td>1.41 0.78</td>
<td>1.59 0.49</td>
<td>***</td>
<td>1.33 0.85</td>
<td>1.27 0.72</td>
<td>1.21 0.55</td>
<td>1.33 0.85</td>
<td></td>
</tr>
<tr>
<td>Word length in letters</td>
<td>6.20 1.58</td>
<td>5.95 1.63</td>
<td>**</td>
<td>6.30 1.55</td>
<td>6.24 1.53</td>
<td>6.18 1.50</td>
<td>6.30 1.55</td>
<td></td>
</tr>
<tr>
<td>Rated imageability</td>
<td>5.32 1.12</td>
<td>6.14 0.47</td>
<td>***</td>
<td>5.01 1.13</td>
<td>5.03 1.14</td>
<td>5.04 1.14</td>
<td>5.01 1.13</td>
<td></td>
</tr>
<tr>
<td>Rated AoA</td>
<td>3.37 1.20</td>
<td>2.53 0.84</td>
<td>***</td>
<td>3.69 1.17</td>
<td>3.70 1.16</td>
<td>3.70 1.16</td>
<td>3.69 1.17</td>
<td></td>
</tr>
<tr>
<td>Log₁₀(N-size + 1)</td>
<td>0.26 0.32</td>
<td>0.32 0.40</td>
<td>***</td>
<td>0.23 0.28</td>
<td>0.37 0.38</td>
<td>0.51 0.42</td>
<td>0.23 0.28</td>
<td></td>
</tr>
<tr>
<td>Centred adult frequency</td>
<td>0.00 0.57</td>
<td>0.07 0.48</td>
<td>*</td>
<td>-0.03 0.60</td>
<td>0.00 0.52</td>
<td>0.02 0.42</td>
<td>-0.02 0.60</td>
<td></td>
</tr>
<tr>
<td>Centred child frequency</td>
<td>0.00 0.78</td>
<td>0.19 0.49</td>
<td>***</td>
<td>-0.07 0.85</td>
<td>0.00 0.72</td>
<td>-0.06 0.55</td>
<td>0.06 0.85</td>
<td></td>
</tr>
<tr>
<td>Centred letters</td>
<td>0.00 1.58</td>
<td>-0.25 1.63</td>
<td>**</td>
<td>0.10 1.55</td>
<td>0.00 1.53</td>
<td>-0.06 1.50</td>
<td>0.06 1.55</td>
<td></td>
</tr>
<tr>
<td>Centred imageability</td>
<td>0.00 1.12</td>
<td>0.82 0.47</td>
<td>***</td>
<td>-0.31 1.13</td>
<td>0.00 1.14</td>
<td>0.02 1.14</td>
<td>-0.02 1.13</td>
<td></td>
</tr>
<tr>
<td>Centred AoA</td>
<td>0.00 1.20</td>
<td>-0.83 0.84</td>
<td>***</td>
<td>0.32 1.17</td>
<td>0.00 1.16</td>
<td>0.01 1.16</td>
<td>-0.01 1.17</td>
<td></td>
</tr>
<tr>
<td>Centred N-size</td>
<td>0.00 0.32</td>
<td>0.07 0.40</td>
<td>***</td>
<td>-0.03 0.28</td>
<td>0.00 0.38</td>
<td>0.14 0.42</td>
<td>-0.14 0.28</td>
<td></td>
</tr>
</tbody>
</table>

Note: Centred variables are centred on the means for all items (both languages) of the corresponding raw or log transformed variables. AoA = Age of Acquisition; N-size = orthographic neighbourhood size.

Comparisons for each variable, two-tailed independent-samples t tests comparing Spanish and Italian: *p < .05, **p < .01, ***p < .001.
between languages. We created between-language interaction terms by multiplying the psycholinguistic variables with a language coding variable (Spanish = 0, Italian = 1) because information about interactions is carried by the multiplicative product of their components (Cohen, Cohen, West, & Aiken, 2003). However, in creating interaction terms we raised the problem of multicollinearity since interaction variables and their components are highly correlated. Several coefficients greater than .75 can be seen in Table 2 where we report correlations between reading latencies, a language coding variable, the critical psycholinguistic variables, and interaction terms generated by multiplying the latter two together. These large correlations indicate a high level of multicollinearity, which would be associated with regression results that would be misleading and unstable across samples because correlated predictors are associated with overlapping portions of variance in the outcome measure.

To deal with this multicollinearity, following Cohen et al. (2003), we centred the critical psycholinguistic predictor variables, subtracting the mean of each variable from its values, using the mean computed over items in both languages. Interaction terms were then created as the multiplicative product of a centred language coding variable and the psycholinguistic predictors—for example, the language by adult frequency interaction was the product of multiplying the centred language and adult frequency variables. Centring variables substantially reduces the correlation between variables corresponding to main effects (e.g., adult frequency) and variables corresponding to interactions (e.g., language by adult frequency). This can be seen if one compares the upper (uncentred) and lower (centred) correlation matrices in Table 2. This reduction in pairwise correlations is consistent with a fall in the level of multicollinearity and thus in the risks associated with high levels of multicollinearity. To diagnose the multicollinearity among our predictors, we calculated the condition number (Baayen, 2008) over the critical psycholinguistic variables plus the computed interaction terms. For Experiment 1 data, the condition number is 71 when predictors are uncentred, indicating high levels of multicollinearity, but is 6 when the predictors are centred, indicating negligible levels of multicollinearity.

**Results and discussion**

To compare word naming in Spanish and Italian, we focused on latencies because participants produced few errors. We used linear mixed-effects (LME) modelling (Baayen, 2008), allowing us to account for random effects due to uncontrolled differences between items or participants. Such random effects potentially confound between-language comparisons. The data for LME modelling consisted of the “raw” (not averaged) latencies of all correct responses. Prior to analysis, we excluded 940 observations (≈3% of the total) corresponding to errors, voice key misfires, or outliers (reaction times, RTs ≤ 200 ms). These exclusions left 30,560 RTs, which were log10 transformed to reduce skew in the latency distribution.

The critical psycholinguistic variables and computed interaction terms were entered as predictors in our analyses. In addition, we entered variables coding for the phonetic characteristics of word initial phonemes, employing a phonetic feature coding scheme based on that used by Balota et al. (2004). The use of the phonetic coding variables eliminated a confound between the original studies because CB06 did but BBA02 did not code for the phonetic characteristics of word initials. (A further difference between the CB06 and BBA02 studies is that the former but not the latter presented items including some morphologically complex words, derivations, or inflections. In analyses that are not reported, we found no significant effect of morphological complexity.)

We stepped through a series of models. First, assuming the same random effects of subjects and items, we compared models differing in fixed effects:

1. A model with just initial phoneme factors.
2. A model with initial phoneme factors plus the language coding variable.
3. A model with initial phoneme factors, the language coding variable, and the centred critical psycholinguistic variables (adult frequency,
Table 2: Experiment 1: Pairwise correlations between predictor variables and log10(RT)

Correlations for raw variables

| Predictor variables | Log10(RT) | Language | Adultfreq | Childfreq | Letters | IMG | AoA | N-size | Lang × Adultfreq | Lang × Childfreq | Lang × Letters | Lang × IMG | Lang × AoA | Lang × N-size |
|---------------------|-----------|----------|-----------|-----------|---------|-----|-----|--------|-----------------|----------------|---------------|------------|------------|-------------|---------------|
| Language            | 0.23***   |          | -0.11***  | -0.13***  |         |     |     |        |                 |                |               |            |            |              |               |
| Adultfreq           |           | -0.11*** | -0.13***  | -0.19***  | -0.26***|     |     |        |                 |                |               |            |            |              |               |
| Childfreq           |           |          | -0.13***  | -0.17***  | -0.69***|     |     |        |                 |                |               |            |            |              |               |
| Letters             |           |          |           |          |         |     |     |        |                 |                |               |            |            |              |               |
| IMG                 |           |          |           |          |         |     |     |        |                 |                |               |            |            |              |               |
| AoA                 |           |          |           |          |         |     |     |        |                 |                |               |            |            |              |               |
| N-size              |           |          |           |          |         |     |     |        |                 |                |               |            |            |              |               |

Correlations for centred variables

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C(lang)</td>
<td>0.23***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C(adultfreq)</td>
<td></td>
<td>-0.11***</td>
<td>-0.13***</td>
<td>-0.17***</td>
<td>-0.26***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C(childfreq)</td>
<td></td>
<td></td>
<td>-0.13***</td>
<td>-0.17***</td>
<td>-0.69***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C(letters)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C(IMG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C(AoA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C(N-size)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Condition numbers for each predictor set are shown. RT = reaction time; c = centred; lang = language; adultfreq = log10(adult frequency per million + 1); childfreq = log10(child frequency per million + 1); IMG = rated imageability; AoA = rated age of acquisition; N-size = log10(N-size + 1) For all correlations. ***p < 0.001, **p < 0.01, *p < 0.05.
child frequency, AoA, imageability, length, and N-size).

4. A model with the same predictors as Model 3 plus terms corresponding to the interactions between the centred language coding variable and each of the centred critical psycholinguistic variables.

We contrasted the utility of the more complex models by comparing them with the simpler models using the Likelihood Ratio Test (LRT; Baayen, 2008). We found that each increment in complexity was justified by improved fit to the data: comparing Models 1 and 2, \( \chi^2(1) = 11, p = .001 \); comparing Models 2 and 3, \( \chi^2(6) = 268, p = 2 \times 10^{-16} \); and comparing Models 3 and 4, \( \chi^2(6) = 21, p = .002 \).

Secondly, we evaluated whether inclusion of random effects was necessary in the final model (Model 4), using LRT comparisons between models with the same fixed effects but differing random effects. We found that both random effects were justified, in comparisons of models that included: (a) both random effects of subjects and items, as specified for Model 4; (b) just the random effect of subjects; (c) just the random effect of items. The difference between Models (a) and (b) was significant, \( \chi^2(1) = 1,592, p = 2 \times 10^{-16} \). The difference between Models (a) and (c) was also significant, \( \chi^2(1) = 11,662, p = 2 \times 10^{-16} \).

Finally, we checked whether our findings were influenced by outliers, fitting Model 4 with both random effects to a trimmed dataset omitting 745 outlier observations. These outliers were distinguished by being associated with standardized Model 4 residuals > 2.5 (Baayen, 2008). The original and trimmed dataset models showed the same pattern of effects, indicating that our results are robust to the presence of outliers. We report a summary of the final model, on the trimmed dataset, in Table 3, noting significant effects of language (Italian latencies were slower), adult frequency (more frequent words elicited shorter latencies), and length (longer words elicited longer latencies), as well as significant interactions between language and length, and language and AoA.

We explored the interactions by examining the impact of the critical psycholinguistic variables on reading in each language considered alone. We report summaries of the resulting models in Table 3. For both languages, reported models included random effects of participants and items and were fitted to trimmed datasets. It can be seen that the adult and child frequency effects were significant only in Italian, though the interactions between the effects of language and adult or child frequency were not indicated to be significant. The effect of word length was larger in Spanish than in Italian but significant in both languages. The AoA effect (later acquired words elicited longer RTs) was significant for Spanish only. Likewise, the effect of N-size was significant only in Spanish (words with more neighbours elicited longer RTs).

A joint analysis of word naming data for both Spanish and Italian indicated significant effects due to language, word length, and adult word frequency as well as significant interactions between the effects of language and word length, and language and AoA. Critically, simple effects analyses by language indicated that the AoA effect was significant in Spanish but not in Italian. Thus, our reanalysis of the data recorded by CB06 and BBA02 replicated the original finding.

We hypothesized that the between-language difference in the AoA effect remained because, while we controlled for other methodological differences between the studies, a salient difference remained: the greater imageability of the Spanish words. The average imageability of the Spanish items (average rated imageability of 6.1) was significantly greater than that of the Italian items (average rated imageability of 5.0). It should be noted that the distribution of imageability values in the two language samples was substantially different. All the Spanish items were object names like “tigre” (tiger), with Spanish rated imageability ranging between 5 and 7. In contrast, many of the Italian items included abstract words like “litigio” (quarrel) with a relatively even number of items occurring at every point on the
Table 3. Summary of linear mixed-effects (LME) models of log_{10}(RT) reading in Spanish and Italian in each experiment

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>p(MCMC)</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>p(MCMC)</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>p(MCMC)</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>p(MCMC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.7026</td>
<td>0.0062</td>
<td>437.11</td>
<td>***</td>
<td>2.7060</td>
<td>0.0068</td>
<td>397.20</td>
<td>***</td>
<td>2.6823</td>
<td>0.0078</td>
<td>344.41</td>
<td>***</td>
<td>2.7042</td>
<td>0.0055</td>
<td>489.19</td>
<td>***</td>
</tr>
<tr>
<td>Affricative</td>
<td>0.0161</td>
<td>0.0057</td>
<td>2.81</td>
<td></td>
<td>0.0271</td>
<td>0.0058</td>
<td>4.71</td>
<td>***</td>
<td>0.0046</td>
<td>0.0145</td>
<td>2.94</td>
<td></td>
<td>0.0099</td>
<td>0.0044</td>
<td>2.33</td>
<td></td>
</tr>
<tr>
<td>Abstrait</td>
<td>0.0165</td>
<td>0.0031</td>
<td>5.30</td>
<td>***</td>
<td>0.0115</td>
<td>0.0038</td>
<td>3.07</td>
<td>***</td>
<td>0.0017</td>
<td>0.0051</td>
<td>0.34</td>
<td>***</td>
<td>0.0182</td>
<td>0.0023</td>
<td>8.07</td>
<td>***</td>
</tr>
<tr>
<td>Biliteral</td>
<td>0.0006</td>
<td>0.0025</td>
<td>0.24</td>
<td>***</td>
<td>0.0073</td>
<td>0.0026</td>
<td>2.77</td>
<td>***</td>
<td>-0.0094</td>
<td>0.0041</td>
<td>-2.30</td>
<td>***</td>
<td>0.0037</td>
<td>0.0018</td>
<td>2.02</td>
<td>***</td>
</tr>
<tr>
<td>Dental</td>
<td>0.0083</td>
<td>0.0029</td>
<td>2.87</td>
<td>***</td>
<td>0.0155</td>
<td>0.0029</td>
<td>5.31</td>
<td>***</td>
<td>0.0081</td>
<td>0.0053</td>
<td>1.51</td>
<td>***</td>
<td>0.0079</td>
<td>0.0019</td>
<td>4.08</td>
<td>***</td>
</tr>
<tr>
<td>Fracitive</td>
<td>0.0202</td>
<td>0.0034</td>
<td>5.94</td>
<td>***</td>
<td>0.0379</td>
<td>0.0043</td>
<td>8.86</td>
<td>***</td>
<td>0.0066</td>
<td>0.0050</td>
<td>1.33</td>
<td>***</td>
<td>0.0084</td>
<td>0.0024</td>
<td>3.58</td>
<td>***</td>
</tr>
<tr>
<td>Labiodental</td>
<td>-0.0120</td>
<td>0.0039</td>
<td>-3.06</td>
<td>***</td>
<td>-0.0214</td>
<td>0.0047</td>
<td>-4.56</td>
<td>***</td>
<td>-0.0235</td>
<td>0.0073</td>
<td>-3.21</td>
<td>***</td>
<td>-0.0022</td>
<td>0.0029</td>
<td>-0.78</td>
<td>***</td>
</tr>
<tr>
<td>Liquid</td>
<td>-0.0166</td>
<td>0.0041</td>
<td>-4.11</td>
<td>***</td>
<td>-0.0120</td>
<td>0.0044</td>
<td>-2.74</td>
<td>***</td>
<td>0.0017</td>
<td>0.0073</td>
<td>0.24</td>
<td>***</td>
<td>-0.0178</td>
<td>0.0030</td>
<td>-5.84</td>
<td>***</td>
</tr>
<tr>
<td>Nasal</td>
<td>-0.0019</td>
<td>0.0027</td>
<td>-0.72</td>
<td></td>
<td>-0.0067</td>
<td>0.0027</td>
<td>-2.49</td>
<td>***</td>
<td>0.0073</td>
<td>0.0049</td>
<td>1.48</td>
<td>***</td>
<td>-0.0042</td>
<td>0.0020</td>
<td>-2.07</td>
<td>***</td>
</tr>
<tr>
<td>Palatal</td>
<td>0.0319</td>
<td>0.0061</td>
<td>1.95</td>
<td>***</td>
<td>0.0401</td>
<td>0.0061</td>
<td>6.67</td>
<td>***</td>
<td>0.0017</td>
<td>0.0200</td>
<td>0.53</td>
<td>***</td>
<td>0.0168</td>
<td>0.0044</td>
<td>3.79</td>
<td>***</td>
</tr>
<tr>
<td>Velar</td>
<td>0.0318</td>
<td>0.0027</td>
<td>5.19</td>
<td>***</td>
<td>0.0166</td>
<td>0.0028</td>
<td>6.04</td>
<td>***</td>
<td>0.0170</td>
<td>0.0046</td>
<td>3.72</td>
<td>***</td>
<td>0.0114</td>
<td>0.0020</td>
<td>5.58</td>
<td>***</td>
</tr>
<tr>
<td>Voiced</td>
<td>0.0004</td>
<td>0.0018</td>
<td>0.20</td>
<td></td>
<td>-0.0042</td>
<td>0.0018</td>
<td>-2.38</td>
<td>***</td>
<td>0.0172</td>
<td>0.0038</td>
<td>5.47</td>
<td>***</td>
<td>-0.0060</td>
<td>0.0014</td>
<td>-0.39</td>
<td>***</td>
</tr>
<tr>
<td>C(log_{10}adultfrequency)</td>
<td>-0.0050</td>
<td>0.0018</td>
<td>-2.75</td>
<td>***</td>
<td>-0.0050</td>
<td>0.0016</td>
<td>-3.15</td>
<td>***</td>
<td>-0.0020</td>
<td>0.0043</td>
<td>-0.47</td>
<td>***</td>
<td>-0.0077</td>
<td>0.0015</td>
<td>-5.14</td>
<td>***</td>
</tr>
<tr>
<td>C(letter)</td>
<td>-0.0017</td>
<td>0.0018</td>
<td>-0.92</td>
<td></td>
<td>-0.0032</td>
<td>0.0014</td>
<td>-2.29</td>
<td>***</td>
<td>-0.0038</td>
<td>0.0049</td>
<td>-0.77</td>
<td></td>
<td>-0.0032</td>
<td>0.0014</td>
<td>-2.29</td>
<td>***</td>
</tr>
<tr>
<td>C(imageability)</td>
<td>0.0064</td>
<td>0.0006</td>
<td>11.52</td>
<td>***</td>
<td>0.0049</td>
<td>0.0005</td>
<td>9.89</td>
<td>***</td>
<td>0.0095</td>
<td>0.0010</td>
<td>9.29</td>
<td>***</td>
<td>0.0039</td>
<td>0.0005</td>
<td>8.48</td>
<td>***</td>
</tr>
<tr>
<td>C(letter) × C(imageability)</td>
<td>-0.0001</td>
<td>0.0011</td>
<td>-0.12</td>
<td></td>
<td>0.0004</td>
<td>0.0009</td>
<td>0.43</td>
<td></td>
<td>-0.0029</td>
<td>0.0027</td>
<td>-1.08</td>
<td></td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>C(letter) × C(letter)</td>
<td>0.0019</td>
<td>0.0011</td>
<td>1.82</td>
<td></td>
<td>0.0007</td>
<td>0.0011</td>
<td>0.64</td>
<td></td>
<td>0.0045</td>
<td>0.0019</td>
<td>2.36</td>
<td></td>
<td>0.0012</td>
<td>0.0008</td>
<td>1.51</td>
<td></td>
</tr>
<tr>
<td>C(AoA)</td>
<td>0.0036</td>
<td>0.0028</td>
<td>1.27</td>
<td></td>
<td>0.0005</td>
<td>0.0030</td>
<td>0.18</td>
<td></td>
<td>0.0096</td>
<td>0.0040</td>
<td>2.36</td>
<td></td>
<td>-0.0026</td>
<td>0.0020</td>
<td>-1.29</td>
<td></td>
</tr>
<tr>
<td>C(N-size)</td>
<td>0.0257</td>
<td>0.0089</td>
<td>2.63</td>
<td>***</td>
<td>0.0162</td>
<td>0.0105</td>
<td>1.54</td>
<td>***</td>
<td>0.0052</td>
<td>0.0030</td>
<td>1.75</td>
<td>***</td>
<td>-0.0016</td>
<td>0.0029</td>
<td>-0.54</td>
<td>***</td>
</tr>
<tr>
<td>C(letter) × C(letter)</td>
<td>-0.0029</td>
<td>0.0002</td>
<td>-2.46</td>
<td>**</td>
<td>0.0034</td>
<td>0.0009</td>
<td>3.74</td>
<td>***</td>
<td>-0.0007</td>
<td>0.0003</td>
<td>-0.50</td>
<td>***</td>
<td>-0.0021</td>
<td>0.0006</td>
<td>-1.31</td>
<td>***</td>
</tr>
<tr>
<td>C(letter) × C(letter)</td>
<td>0.0016</td>
<td>0.0035</td>
<td>0.57</td>
<td></td>
<td>0.0344</td>
<td>0.0041</td>
<td>1.47</td>
<td></td>
<td>0.0060</td>
<td>0.0041</td>
<td>1.47</td>
<td></td>
<td>0.0037</td>
<td>0.0013</td>
<td>2.86</td>
<td>***</td>
</tr>
</tbody>
</table>

**Note:** C() variables are centred on mean values of all items for both languages. Summary of linear mixed-effects (LME) models. RT = reaction time; MCMC = Markov chain Monte Carlo. $p$(MCMC) is reported: $\sim p < .10$. $^* p < .05$. $^{**} p < .01$. $^{***} p < .001$. 

1817
imageability range from 1 to 7. This difference between the samples might have induced a greater semantic contribution in reading, reflected in an AoA effect for Spanish (see Wilson et al., 2013, for a similar interpretation). To test this hypothesis, we compared the BBA02 reading latencies with latencies for a new set of Spanish words matched on imageability.

EXPERIMENT 2

Method

We collected Spanish word naming data in a new study (a detailed description is reported by Davies, Barbón, & Cuetos, 2013). Twenty-five Spanish-speakers (average age = 21 years; 24 females, 1 male) read 2,764 words sampling a broad range of AoA and frequency values. Words were randomly assigned to four blocks (each read on a separate day). Presentation order was randomized both within and between blocks. Our analysis made a cross-linguistic comparison of reading latencies for the 626 Italian BBA02 words and a matched subset of 626 words from the new sample of Spanish words.

Stimuli

We report a summary of item characteristics in Table 1. The Italian and Spanish words were matched on every variable (p > .10) except log child frequency and log N-size (p < .05). All words were morphologically simple. Note that for Experiment 2 data, the condition number is 49 when predictors are uncentred, indicating high multicollinearity, but is 5 when the predictors are centred, indicating negligible multicollinearity (Table 4).

Results and discussion

Our analysis of reading in Spanish and Italian again focused on latencies because there were too few errors for analysis. The criteria for incorrect responses were the same as those in Experiment 1, excluding 1,169 latencies (3.4% of the total), leaving 33,261 RTs for analysis.

First, assuming the same random effects of subject and item, we compared models differing in fixed effects. Successive LRTs showed that each increment in complexity was justified by improved fit to the data: comparing Models 1 (just phonetic variables) and 2 (adding the centred language coding variable), χ²(1) = 5, p = .02; comparing Models 2 and 3 (adding centred critical psycholinguistic variables), χ²(6) = 276, p = 2 × 10⁻¹⁶; and comparing Models 3 and 4 (adding interactions between language and the centred critical psycholinguistic variables), χ²(6) = 22, p = .001.

Secondly, LRT comparisons between models with the same fixed effects structure (Model 4) but differing random effects showed that random effects of both participants and items were justified. The difference between Model (a), a model with both random effects, and Model (b), a model with just the random effect of participant, was significant, χ²(1) = 839, p = 2 × 10⁻¹⁶. The difference between Model (a) and Model (c), a model with just the random effect of items, was significant, χ²(1) = 11,143, p = 2 × 10⁻¹⁶.

Finally, we checked whether our findings were influenced by outliers, fitting Model 4 with both random effects to a trimmed dataset omitting 937 outlier observations. The original and trimmed dataset models showed the same pattern of significant effects. We report a summary of the final model, on the trimmed dataset, in Table 3. There was a significant effect of adult and a near-significant effect of child frequency (more frequent words elicited shorter latencies), as well as significant effects of word length (longer words elicited longer latencies) and language (Spanish reading latencies were shorter), and a significant interaction between language and length.

We explored the interaction of language with length by examining the reading data for each language separately. We report summaries of the resulting models in Table 3, reporting models including random effects of participants and items, fitted to trimmed datasets. It can be seen that the length effect coefficient was larger in Italian than in Spanish, though the effect was...
<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>Log$_{10}$(RT)</th>
<th>Language</th>
<th>Adultfreq</th>
<th>Childfreq</th>
<th>Letters</th>
<th>IMG</th>
<th>AoA</th>
<th>N-size</th>
<th>Lang × Adultfreq</th>
<th>Lang × Childfreq</th>
<th>Lang × Letters</th>
<th>Lang × IMG</th>
<th>Lang × AoA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lang</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adultfreq</td>
<td>-.09***</td>
<td>.03***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.16***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Childfreq</td>
<td>-.06***</td>
<td>.08***</td>
<td>.64***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.06***</td>
<td>.08***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letters</td>
<td>.12***</td>
<td>.04***</td>
<td>-.06***</td>
<td>.16***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMG</td>
<td>-.01</td>
<td>-.02**</td>
<td>-.10***</td>
<td>.33***</td>
<td>.04***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AoA</td>
<td>.05***</td>
<td>-.01</td>
<td>-.24***</td>
<td>-.67***</td>
<td>.15***</td>
<td>.69***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-size</td>
<td>-.14***</td>
<td>-.37***</td>
<td>.03***</td>
<td>.07***</td>
<td>-.62***</td>
<td>.11***</td>
<td>-.17***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lang × Adultfreq</td>
<td>.10***</td>
<td>.78***</td>
<td>.50***</td>
<td>.43***</td>
<td>-.04***</td>
<td>-.06***</td>
<td>-.12***</td>
<td>-.25***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lang × Childfreq</td>
<td>.08***</td>
<td>.73***</td>
<td>.36***</td>
<td>.65***</td>
<td>-.09***</td>
<td>.14***</td>
<td>-.35***</td>
<td>-.20***</td>
<td>.85***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lang × Letters</td>
<td>.20***</td>
<td>.94***</td>
<td>-.07***</td>
<td>-.01</td>
<td>.30***</td>
<td>-.03***</td>
<td>.04***</td>
<td>-.47***</td>
<td>.70***</td>
<td>.63***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lang × IMG</td>
<td>.17***</td>
<td>.95***</td>
<td>-.06***</td>
<td>.16***</td>
<td>.03***</td>
<td>.22***</td>
<td>-.17***</td>
<td>-.33***</td>
<td>.72***</td>
<td>.75***</td>
<td>.88***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lang × AoA</td>
<td>.17***</td>
<td>.90***</td>
<td>-.12***</td>
<td>-.17***</td>
<td>.10***</td>
<td>-.24***</td>
<td>.31***</td>
<td>-.35***</td>
<td>.64***</td>
<td>.46***</td>
<td>.88***</td>
<td>.76***</td>
<td></td>
</tr>
<tr>
<td>Lang × N-size</td>
<td>.02***</td>
<td>.48***</td>
<td>.07***</td>
<td>.17***</td>
<td>-.40***</td>
<td>.06***</td>
<td>-.14***</td>
<td>.31***</td>
<td>.44***</td>
<td>.45***</td>
<td>.26***</td>
<td>.48***</td>
<td>.36***</td>
</tr>
</tbody>
</table>

**Table 4. Experiment 2: Pairwise correlations between predictor variables and log$_{10}$(RT)**

Correlations for raw variables

Correlations for centred variables

Note: Condition numbers for each predictor set are shown. RT = reaction time; c = centred; lang = language; adultfreq = log$_{10}$(adult frequency per million + 1); childfreq = log$_{10}$(child frequency per million + 1); IMG = rated imageability; AoA = rated age of acquisition; N-size = log$_{10}$(N-size + 1). For all correlations, ***p < 0.001, **p < 0.01, *p < 0.05.
significant in both languages. Critically, while there were effects of adult frequency in both languages, there were no effects of AoA and no language by AoA interaction.

The interaction between language and AoA was no longer significant when Italian and Spanish words were matched on imageability. There was no AoA effect when considering the Spanish and Italian reading data, either together or apart. Therefore our results provide decisive support for the view that the CB06 AoA effect was observed because the Spanish words were highly imageable.

**GENERAL DISCUSSION**

Previous research indicated a surprising contrast in the factors that influence word naming in two similarly transparent orthographies. Cuetos and Barbón (2006, CB06) reported that reading in Spanish was affected by AoA but not frequency, while Barca et al. (2002, BBA02) reported that reading in Italian was affected by frequency but not AoA. We hypothesized that this between-language difference arose either because of methodological differences between the two studies that included the number and nature of the variables introduced and the type of data analyses, or because the high imageability of the Spanish words amplified the AoA effect. Our results support the second view, opening new insights into the contribution of semantics in reading.

In Experiment 1, we reanalysed the CB06 Spanish and BBA02 Italian data in one cross-linguistic comparison. Analysing both data sets together eliminated a number of between-study differences that otherwise confounded the cross-language comparison. We analysed reading latencies using centred psycholinguistic predictor variables and variables representing their interactions with language. We found significant effects of language, frequency, and length. A number of interactions, including the critical interaction between AoA and language, also reached significance. The simple effects analyses showed that the effect of AoA was significant only for Spanish. In Experiment 1, we thus replicated the pattern of effects found in the original studies.

In Experiment 2, we compared the Italian naming latencies from the BBA02 study with new reading data for Spanish words matched to the BBA02 items on imageability and other critical variables. Results showed effects of adult frequency, word length, and language and one significant interaction between language and length with a larger length effect for Italian. Matching the stimulus sets eliminated the between-language difference in the effect of AoA, consistent with our explanation that the greater imageability of the Spanish CB06 items amplified the AoA effect in that experiment.

Connectionist models of learning show that the age at which words are learnt should affect reading and that that effect will be more prominent for words involving less predictable mappings (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; Mermillod et al., 2012; P. Monaghan & Ellis, 2010; Zevin & Seidenberg, 2002). The P. Monaghan and Ellis (2010) simulations persuasively demonstrate that if the presentation of training stimuli is carefully phased to match the frequency distribution of words over time then AoA effects will emerge. We think that these observations evidence the impact of the frequency trajectory of words, construing that term broadly as variation over age in the frequency of occurrence of items. The question arises, however, whether analyses like those we report would differ in outcome if researchers were to use frequency trajectory measures like those calculated in previous analyses of reading behaviour (Bonin et al., 2004; Zevin & Seidenberg, 2004)—that is, as the difference between (a) the frequency of a word as estimated from a corpus of text read by (younger) children and (b) the frequency of a word as estimated from a corpus of text read by adults. As discussed in the introduction, recent evidence, including the analyses reported by Cortese and Khanna (2007) and the simulation results reported by P. Monaghan and Ellis (2010), suggests that AoA effects are observable independent of potentially confounding factors, including frequency, and that AoA effects stem from the point of entry of words in training. It should be noted that
P. Monaghan and Ellis (2010) calculated frequency trajectory values for the large sample of items in their training set (in the same manner as did Bonin et al., 2004; Zevin & Seidenberg, 2002, 2004) and found a small but significant effect in addition to the AoA effect. These results indicate that both effects of AoA and effects of frequency trajectory might be expected for large reading samples of reading in future analyses.

A related question asks what the results of analyses like those we report would be if one were to use objective instead of rated AoA. The advantages of using objective AoA include its relation to explicit behavioural data, in children’s picture naming (Morrison et al., 1997), and its lower correlation with other psycholinguistic factors (Bonin et al., 2004). We are unable to answer the question, however, because the presence of substantial numbers of abstract words among the Italian and the Experiment 2 Spanish stimuli would have restricted the use of objective AoA as a predictor to a relatively small subset of the items. Objective AoA values have been derived from picture naming data (Morrison et al., 1997), and abstract lexical concepts are (by definition) not pictureable in the same manner as object concepts.

Turning to another concern, we should consider the potential impact of variation in the derivation of frequency norms for the results of analyses like those we report. In recent research, Brysbaert and Cortese (2011) have shown that the characteristics of the measures of word frequency used in analyses have an impact on both the effect of frequency and the effect of factors like AoA that might additionally explain variation in reading. Brysbaert and Cortese (2011) found that the AoA effect endured whatever the frequency measure used. However, it is important to consider the potential impact of which frequency measures are used. We note that both the adult frequency estimates that we used derive from analyses of similar written text corpora: the Italian CoLFIS (Bertinetto et al., 2011) and the Spanish LEXESP (Sebastián et al., 2000), of about the same size (CoLFIS, 3.8 million words; LEXESP, 5 million words), and sampling newspapers, magazines, and books. We think, then, that our findings are not confounded by differences between the sources of the frequency values. However, work by Brysbaert and colleagues (Brysbaert & Cortese, 2011; Brysbaert & New, 2009), among others, shows that frequency estimates drawn from larger corpora (e.g., of film subtitle texts, the SUBTLEX database, Brysbaert & New, 2009) capture more variance due to frequency in reading behaviour. If more variance is captured by frequency, less may be available for capture by variables like AoA (Brysbaert & Cortese, 2011). This warrants the adoption of more up-to-date estimates in future research. Such estimates are now available in Spanish (the SUBTLEX-ESP, Spanish Subtitles Lexicon, database, Cuetos, González-Nosti, Barbón, & Brysbaert, 2011) but for a cross-linguistic analysis we think that future work should incorporate frequency estimates compiled from similar corpora in a similar manner, and an Italian version of the SUBTLEX approach has not been completed at the time of writing.

We conclude by focusing on our main finding. We propose that the interaction between language and AoA found in Experiment 1 is explained by assuming (a) that the key difference between CB06 and BBA02 is that the former presented higher imageability words; (b) the AoA effect, here captured by the effect of AoA for highly imageable words, has its locus in the flow of activation from semantic to phonological processing. That AoA effect is seen in addition to the effect of frequency, evident in reading across both languages. According to the results of connectionist simulations of learning (e.g., Ellis & Lambon Ralph, 2000; P. Monaghan & Ellis, 2010; Zevin & Seidenberg, 2002), the effect of frequency can be attributed principally to the cumulative impact of the frequency of occurrence of words on the weights of connections in the orthography-to-phonology mappings involved in reading.

The AoA effect found with highly imageable words in the CB06 Spanish study and replicated in the present Experiment 1 can be interpreted by assuming that such words benefited from a stronger semantic contribution to phonological output mappings (Balota et al., 2004). We would argue that imageable words are processed more semantically because they have richer semantic representations.
and connections and activate semantic features to a larger extent than lower imageability items (Jones, 1985; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008). Thus high–imageability words may induce access to semantics and activate semantic mediation in reading more than abstract/low–imageability words (Pexman et al., 2002). This would allow the AoA effect to appear in reading.

Further evidence in support of this view comes from another study in Spanish. Wilson et al. (2013) orthogonally manipulated frequency and AoA in three word naming studies. Frequency affected performance in all three studies. However, AoA effects were found only when the items were composed of only highly imageable words (similar to the Spanish items of Experiment 1 here). When stimuli contained words from a wider sample of imageability values (similar to the items in Spanish in Experiment 2 here), AoA did not affect word naming latencies.

Previous research has shown that an imageability effect (or effects of semantic indices like WordNet sense number) can be observed for large datasets for reading (Balota et al., 2004; Yap & Balota, 2009), and that if AoA is included in analyses alongside imageability there is an effect of AoA but not of imageability (Cortese & Khanna, 2007). Cortese and Khanna (2007) argue that the AoA effect that they observed in English is semantic. Our observations show that the AoA effect in Spanish, likewise, is a reflection of the semantic contribution to phonological coding in word naming. Together, our findings and those reported by Cortese and Khanna (2007) suggest that the characterization of the semantic influence in reading in terms of imageability (Strain, Patterson, & Seidenberg, 1995, 2002) rather than AoA (J. Monaghan & Ellis, 2002) may have been misleading. The association of more prominent AoA effects with less predictable mappings in simulations (Lambon Ralph & Ehsan, 2006; Mermillod et al., 2012) and the arbitrariness of mappings involving semantics, making them less predictable, furnish a persuasive logic for linking the AoA effect in reading with the influence of semantics. What our findings add is evidence that the semantic contribution is most likely to be detected when words are high in imageability, as for the Spanish items in Experiment 1 (and as in Wilson et al., 2013, Study 4). A semantic contribution, reflected in the AoA effect, is not detected when words are abstract or low in imageability, or when the experimental list includes many such items, as for the Italian items and the Spanish items in Experiment 2. Our findings extend the influence of semantics in word naming to reading in transparent orthographies but they also set limits on where that influence may be found.

The implications of our findings for computational models of reading are two-fold. First, we think that both AoA and frequency effects must be simulated for such models to be considered to be successful. At the time of writing, only the models reported by P. Monaghan and Ellis (2010) or Mermillod et al. (2012) can be argued to do so (but see, also, Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; J. Monaghan & Ellis, 2002; Zevin & Seidenberg, 2002). Both these models employ connectionist design features, and we think that the learning capacity of connectionist models is essential to the simulation of AoA and frequency effects. Thus, while the dual-route cascaded model (Coltheart, Perry, Ziegler, & Rastle, 2001) does simulate frequency but does not consider AoA effects, it could, in principle, be extended to simulate AoA effects through the incorporation of connectionist mappings (cf. the Connectionist Dual Process (CDP++) model, Perry, Ziegler, & Zorzi, 2007, 2010). Secondly, we think that the variation in the AoA effect that we have isolated requires the implementation of semantic representations. At the time of writing, only the simulations reported by Harm and Seidenberg (2004) implement realistic semantic representations. Future research may show whether a model with the same or similar architecture could, perhaps in combination with the phased training regime deployed by P. Monaghan and Ellis (2010), successfully simulate findings like those reported in the present article.
REFERENCES


