THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2013
Vol. 66, No. 5, 973–1000, http://dx.doi.org/10.1080/17470218.2012.723725

Routledge
Taylor & Francis Group

Does frequency trajectory influence word identification? A cross-task comparison

Bernard Lété[1], and Patrick Bonin[2]

[1]University of Lyon 2, Laboratoire EMC, Lyon, France
[2]Institut universitaire de France, University of Bourgogne, LEAD-CNRS, Dijon, France

In a series of six experiments, the influence of frequency trajectory in visual word recognition was investigated. In Experiment 1, frequency trajectory was found to exert a strong and reliable influence on age of acquisition (AoA) ratings. In word reading (Experiment 2), lexical decision (Experiments 3 and 6), proper name decision (Experiment 4), progressive demasking (Experiment 5), and a multiple regression analysis of lexical decision times taken from the French Lexicon Project, the effect of frequency trajectory was not reliable. In contrast, in all the experiments and in the multiple regression analysis, cumulative frequency had a strong and reliable influence on word recognition times. The findings firmly establish that in alphabetic languages such as French, age-limited learning effects do not surface readily in word recognition. In contrast, the total exposure to words across the lifetime is a strong determinant of word recognition speed. The implications of the findings are discussed.

Keywords: Word recognition; Frequency trajectory; Cumulative frequency; Lexical decision; Word reading; Progressive demasking.

Words acquired early in life are processed faster and more accurately than words acquired later. The so-called age of acquisition (AoA) effects have received strong empirical support in recent years. In effect, an impressive number of studies have reported these effects in a variety of lexical processing tasks such as object and face naming, lexical decision, and object identification, and also in different populations and languages (see Johnston & Barry, 2006; Juhasz, 2005, for reviews). A critical issue that remains in visual word recognition is to determine whether age-limited learning effects are universal effects or are limited in scope. In a study conducted in Turkish, a language with a very transparent orthographic system, Raman (2006) found AoA effects in word reading latencies and claimed that AoA effects were universal. However, Zevin and Seidenberg (2002, 2004) and Bonin, Barry, Méot, and Chalard (2004) have claimed that age-limited learning effects are not universal. On the basis of empirical or computational evidence, they have argued that age-limited

Correspondence should be addressed to Bernard Lété, Université de Lyon 2, Laboratoire EMC, 5 avenue Pierre Mendès France, 69676–Bron cedex, France. E-mail: bernard.lete@univ-lyon2.fr

The authors wish to thank very much Marc Brysbaert, Padraic Monaghan, Miguel Pérez, and an anonymous reviewer for helpful and challenging comments on a previous version of the paper. We are grateful to Sébastien Roux for conducting Experiment 1 and Experiment 2, and to Emma Forin and Clothilde Joly for conducting Experiment 4.

learning effects are observed under specific conditions. In particular, age-limited learning effects are found when lexical processing requires the mobilization of (specific) memorized links. This is typically the case in object or in face naming because these tasks require participants to produce an object to a proper (or an object) name in response to a face or object—that is, semantic–lexical links are recruited. In the current study, we examined the influence of age-limited learning effects in word recognition using a multitask approach and frequency trajectory as an index of these effects. As explained below, to investigate age-limited learning effects, we have used frequency trajectory and not the classical rated AoA norms. The use of frequency trajectory to investigate age-limited effects has not, as yet, given rise to many empirical studies in psycholinguistics. Investigating this issue in word recognition is very important due to its implication for word recognition models. As Zevin and Seidenberg (2002) put it: "the finding that AoA affects performance independent of frequency seems to present a challenge for models of word naming" (p. 2), and even more recently, Brysbaert and Cortese (2011, p. 545): "The extent to which frequency and age of acquisition (AoA) relate to naming and lexical decision performance has become a central issue for researchers interested in word recognition. It is important to understand this issue because of a genuine AoA effect has considerable impact on theoretical approaches to word recognition".

Age of acquisition refers to the age at which words are acquired in their written or spoken form. Traditionally, there have been two methods of obtaining AoA scores for words. The first method is to ask adults to evaluate the age at which they think they have acquired a given word in its spoken or written form by using point scales. The second method is to rely on children's performances. Morrison, Chappell, and Ellis (1997) used picture naming in children of various ages to collect (objective) AoA norms. Since rated AoA norms are easier to collect than objective AoA norms, this may explain why the former have been more extensively used than the latter. Researchers using rated norms have all acknowledged that, insofar as they are based on metalinguistic knowledge, they cannot truly reflect the *real age* at which words are acquired (Johnston & Barry, 2006). However, given the strong correlations that have been found between rated AoA scores and so-called objective measures of AoA obtained from children's performances (e.g., .75 in the Morrison et al., 1997, English study; .72 in the Pind, Jonsdottir, Tryggvadottir, & Jonsson, 2000, Icelandic study), researchers have continued to employ them to investigate age-limited learning effects (see Dent, Johnston, & Humphreys, 2008, for an example of a recent study based on rated AoA norms). However, as explained below, the use of either subjective or objective AoA norms to examine the influence of age-limited learning effects in lexical processing has been vigorously challenged by Zevin and Seidenberg (2002, 2004; see also Bonin et al., 2004).

As far as word naming is concerned, AoA effects on latencies have been reported in multiple regression studies (e.g., Colombo & Burani, 2002; Gilhooly & Logie, 1981; Morrison & Ellis, 2000) as well as in semifactorial or factorial studies (e.g., Brysbaert, Lange, & Van Wijnendaele, 2000; V. Coltheart, Laxon, & Keating, 1988; Gerhand & Barry, 1998, 1999a; Monaghan & Ellis, 2002). These effects have been found in alphabetic languages such as French or English but also in non-alphabetic languages (e.g., in Japanese kanji, Havelka & Tomita, 2006; in Chinese, Liu, Hao, Hua, Tan, & Weekes, 2008). Monaghan and Ellis (2002) have shown that AoA effects are stronger for inconsistent than for consistent words and, as mentioned earlier, Raman (2006) has even reported AoA effects in word reading in Turkish, a very transparent language. As mentioned above, this finding was taken to support the strong claim that these effects are universal. However, in Italian, which also has very regular letter-to-sound mappings, Burani, Arduino, and Barca (2007) did not find a reliable effect of AoA in word reading latencies when child written word frequencies were taken into account.

The reliability of AoA effects in word reading has been strongly challenged by Zevin and

Seidenberg (2002). These authors raised two main objections against the existence of AoA effects in word reading. The first objection was methodological. Zevin and Seidenberg (2002) closely examined previous reports of AoA effects in word reading studies in English and found that they did not properly control for cumulative frequency—namely, how often words are encountered throughout lifetime. (It should be recalled that other studies have gone a long way to showing that rated AoA effects in lexical processing tasks are not merely cumulative frequency effects in disguise; see Ghyselinck, Lewis, & Brysbaert, 2004; Morrison, Hirsh, Chappell, & Ellis, 2002; Stadthagen-Gonzalez, Bowers, & Damian, 2004). In effect, Zevin and Seidenberg (2002) showed that in these English studies (Gerhand & Barry, 1998, 1999a; Monaghan & Ellis, 2002; Morrison & Ellis, 1995; Turner, Valentine, & Ellis, 1998), the early- and late-acquired words were controlled for adult frequencies taken either from Kučera and Francis (1967) or from CELEX frequency counts (Baayen, Piepenbrock, & Gulikers, 1995), while child frequencies were not taken into account. The only case where the early- and late-acquired words did not differ on the Zeno frequency norms (Zeno, Ivens, Millard, & Duvvuri, 1995) was Monaghan and Ellis's (2002) consistent words, which exhibited an AoA effect of only 7 ms. According to Zevin and Seidenberg (2002), frequency norms such as Zeno et al.'s (1995) *Educator's Word Frequency Guide* permit a better control of cumulative frequency. In multiple regression analyses of word naming times taken from the the large-scale word naming studies of Seidenberg and Waters (1989) and Spieler and Balota (1997), and the large-scale lexical decision study by Balota, Pilotti, and Cortese (2001), Zevin and Seidenberg (2002) found that cumulative frequency, but not rated AoA, was a reliable predictor. The same pattern of results was found in the Bonin et al. (2004) word reading study conducted in French. These reports (which "go a little against the grain") are clearly at odds with the claim that AoA effects are universal. The second objection was from a theoretical point of view. According to their connectionist theory of word reading, when what is learned from initial (early) items can carry over to new items, generalization is possible, and, thus, age-limited learning effects are not predicted. This is precisely the case in word reading in alphabetic languages where the mappings between letter and sound units are quasiregular. In contrast, when learning the name of a particular object or a face, a specific link has to be memorized. In such cases, age-learning effects are predicted, and this is precisely what has been observed: Object naming gives rise to strong age-limited learning effects (Bonin et al., 2004; Ellis & Lambon Ralph, 2000).

One of the most important contributions of the Zevin and Seidenberg (2002) study was to point out the inadequacies of the classical AoA measures for the investigation of age-limited learning effects. For them, AoA measures are behavioural outcomes. They are the result of many factors, one of which is the frequency trajectory of the words. Frequency trajectory is the variation in frequency of the words across lifetime. Some words are more frequently encountered during childhood and less so in adulthood (high-to-low frequency words) whereas the reverse is true for other words (low-to-high frequency trajectory words). Thus, the rated AoA measures should not be considered as a genuine casual factor influencing naming speed (Bonin, Méot, Mermillod, Ferrand, & Barry, 2009). More objective measures should therefore be used to investigate age-limited effects, and frequency trajectory is such a measure.[1]

In the current study, we investigated the influence of frequency trajectory in word recognition in French. French, like English, is an alphabetic

[1] We accept that both objective (written and spoken) word frequency and frequency trajectory are also influenced by the behaviour of human beings (i.e., some words are more frequent because individuals produce or read them more often than others). In our view, objective or rated AoA scores are characterized as a behavioural outcome because the way they are measured directly depends on the performance of participants (naming accuracy in children, ratings in adults). In contrast, word frequency and frequency trajectory are derived from corpus analyses. They indicate the number of times a word is found in a corpus. This is the reason why we qualify the former measures, but not the latter, as "more objective".

language in which the relationships between orthographic and phonological sublexical units are quasisystematic (Peereman & Content, 1999; Peereman, Lété, & Sprenger-Charolles, 2007). There are several reasons why we consider it to be important to address the issue of the influence of frequency trajectory in word recognition in greater detail. First of all, although the influence of frequency trajectory has already been investigated in reading aloud in French (Bonin et al., 2004) and in English (Zevin & Seidenberg, 2002), only very few studies have focused on investigating the influence of frequency trajectory. Given the major implications these findings have with regard to the long-term influence of age-limited learning in word recognition, it is important to establish their robustness. Indeed, most influential models of word recognition do not take account of age-limited learning effects (e.g., M. Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 1999, 2004; Plaut, McClelland, Seidenberg, & Patterson, 1996). We return to this issue in the General Discussion. Moreover, the few studies that have used lexical decision have reported mixed findings. Bonin et al. (2004) found a reliable influence of both frequency trajectory and cumulative frequency in lexical decision times, whereas Caza and Moscovitch (2005) did not. Thus, more data are needed to determine the generality of the influence of frequency trajectory in lexical decision. Of course, many studies have reported AoA effects in lexical decision using classical subjective AoA ratings (e.g., Bonin, Chalard, Méot, & Fayol, 2001; Brysbaert, Lange, et al., 2000; Butler & Hains, 1979; Colombo & Burani, 2002; Cortese & Khanna, 2007; Gerhand & Barry, 1999b; Whaley, 1978). The strength of our study lies in the fact that we investigate, *with the same set of words*, the influence of frequency trajectory in different tasks that have often been used to study word recognition: lexical decision, word reading, and perceptual identification. This approach should help isolate task-independent and task-specific processes underlying frequency trajectory effects in visual word recognition. This rationale relates to the functional-overlap analysis proposed by Jacobs and Grainger (1994), Grainger and Jacobs (1996), and Carreiras, Perea, and Grainger (1997). If frequency trajectory produces the same pattern of results over tasks, then it can be inferred that it influences some process (es) that are common to all tasks and fundamental to visual word recognition. Moreover, the items used in the following experiments included not only high-to-low-frequency (HL) and low-to-high-frequency (LH) words, but also words having flat trajectories (high–high, HH, words and low–low, LL, words). These last two types of item have not been considered in studies manipulating frequency trajectory. HH words are words that are very frequently encountered during childhood and remain extremely frequent in adulthood, whereas LL words are words that are rarely found in children's books and remain rare in adulthood. According to the Zevin and Seidenberg (2002) theoretical approach (2002, 2004), the observation of frequency trajectory effects critically depends on the tasks used (see also Bonin et al., 2004). These effects are predicted in tasks mobilizing arbitrary links, but less so (if at all), in tasks in which the links are quasisystematic. In word naming in alphabetic languages such as French or English, the links that are mobilized to perform the task are quasisystematic. Bonin et al. (2004) did not find a reliable influence of frequency trajectory in a multiple regression analysis but found that high-cumulative-frequency words took less time to read than low-cumulative-frequency words. Thus, we predicted a reliable effect of objective word frequency with high-frequency words associated to shorter reaction times than low-frequency words (Experiment 2). In the lexical decision task, we did not make any specific predictions given the inconsistencies found in the literature (Experiment 3), but in a fourth experiment (Experiment 4), we used a variant of the lexical decision task that did not include nonwords—namely, the proper–common name decision task. This task has been found to be sensitive to age-limited learning effects (indexed by rated AoA; Brysbaert, Van Wijnendaele, & De Deyne, 2000). The rationale underlying the use of the progressive-demasking task in Experiment 5 is described later in the paper. Finally, we report the analyses from a virtual experiment (Experiment 6) and a multiple regression analysis of lexical decision times taken from the French Lexicon Project (FLP; Ferrand et al., 2010), which contains 38,840 words.

Before testing the influence of frequency trajectory in word naming and lexical decision, it was first necessary to assess whether our frequency trajectory manipulation had a reliable effect on AoA ratings as shown by both Zevin and Seidenberg (2002) and Bonin et al. (2004). In Experiment 1, two groups of participants had to estimate the age at which randomly presented words were acquired in their written or spoken form either by providing a numerical value corresponding to an age (Experiment 1A) or by putting a cross in a 5-point scale (Experiment 1B).

EXPERIMENT 1: INFLUENCE OF FREQUENCY TRAJECTORY ON AOA RATINGS

In Experiment 1, two groups of adults provided AoA ratings for the words used in the word recognition experiments. Because frequency trajectory has been found to influence the AoA ratings of words, it was necessary to check that our frequency manipulation worked in the predicted direction for the items used in the word recognition experiments. In Experiment 1A, participants were required to provide an age of acquisition for each word, whereas in Experiment 1B, they had to rate the age at which they thought they had acquired any given word in its written or spoken form on a 5-point scale.

EXPERIMENT 1A

Method

Participants

Twenty-nine students (three males) at the University of Lyon (France) were tested (mean age: 24 years). All were native speakers of French.

Stimuli

The target words were selected from *Manulex* (Lété, Sprenger-Charolles, & Colé, 2004). *Manulex-infra* (Peereman et al., 2007) was also used to control for the consistency of grapheme–phoneme mappings.[2] *Manulex* was compiled in the same manner as the *Word Frequency Book* by Carroll, Davies, and Richman (1971) and, more recently, the *Educator's Word Frequency Guide* (Zeno et al., 1995), which is used for child language studies in English. It is based on a corpus of 1.9 million words from 54 readers used in French primary schools between the first and fifth grades. The database contains two lexicons: the wordform lexicon (48,886 entries) and the lemma lexicon (23,812 entries). Each lexicon provides a grade-level-based list of words found in first-grade (hereafter G1), second-grade (G2), and third-to-fifth-grade (G3–5) readers.

Manulex-infra was developed later to describe the distributional characteristics of the sublexical and lexical units in *Manulex*. In *Manulex-infra*, the orthographic ambiguity of French words is estimated by calculating the consistency of grapheme-to-phoneme mappings (referred to as GP consistency below) and of phoneme-to-grapheme mappings (PG consistency below).

Four types of words from different grammatical categories (nouns, adjectives, and verbs) were selected. Two of these types consisted of high-to-low-frequency words (hereafter HL words) and low-to-high-frequency words (LH words). The other two consisted of flat frequency trajectories —that is, high-to-high-frequency words (HH words) and low-to-low-frequency words (LL words). There were 16 words in each category. The statistical characteristics of the words are shown in Table 1, and the full list of experimental stimuli is given in Appendix A.

The four categories of words were controlled for the number of letters and syllables. To ensure that the targets did not differ with regard to

[2] Free access to *Manulex* and *Manulex-infra* can be obtained online at http://www.manulex.org/

Table 1. *Statistical characteristics for the word targets used in Experiments 1–6 as a function of word type*

Word properties	*High-to-low*	*Low-to-high*	t *tests*	*High-to-high*	*Low-to-low*
Letter length	7.5 (1.2)	7.5 (1.2)		7.5 (1.2)	7.5 (1.2)
Grapheme length	6.1 (1.4)	6.0 (1.2)		6.3 (1.0)	6.3 (1.3)
Syllable length	2.4 (0.5)	2.3 (0.6)		2.3 (0.7)	2.6 (0.7)
Grade 1 *Manulex* frequency	203 (52)	28 (16)	12.84**	356 (189)	0 (0)
Grade 5 *Manulex* frequency	61 (23)	199 (69)	−7.53**	335 (128)	6 (8)
Lexique[a] frequency	26 (18)	159 (118)	−4.45**	174 (113)	11 (12)
Cumulative frequency (*Manulex* count)	264 (59)	226 (65)	1.68 *ns*	691 (254)	6 (8)
Cumulative frequency (*Lexique* count)	228 (51)	186 (124)	1.26 *ns*	530 (211)	11 (12)
GP and PG consistency (*Manulex-infra* count)					
GP consistency at initial position	94 (11)	98 (7)		93 (14)	94 (11)
GP consistency at middle position	80 (10)	80 (12)		79 (14)	81 (10)
GP consistency at end position	73 (26)	78 (24)		82 (23)	73 (24)
PG consistency at initial position	86 (19)	87 (15)		85 (21)	86 (19)
PG consistency at middle position	76 (16)	76 (16)		77 (16)	78 (14)
PG consistency at end position	53 (30)	53 (31)		59 (30)	53 (29)

Note: GP consistency: grapheme-to-phoneme consistency; PG consistency: phoneme-to-grapheme consistency. Standard deviations are provided in parentheses. The *t* tests are between high-to-low- and low-to-high-frequency trajectory. Frequency values are per million.

[a]New et al. (2004) database.

ns: $p > .10$. *$.01 < p < .05$. **$p < .01$.

orthographic consistency, they were controlled for on several GP and PG consistency indexes.[3]

As can be seen from Table 1, HL words had a frequency of 203 per million in the *Manulex* G1 lexicon and 61 in the G3–5 lexicon, $t(30) = 12.84$, $p < .01$; LH words had a frequency of 28 in G1 and 199 in G3–5, $t(30) = -7.53$, $p < .01$. The cumulative frequency was the same across HF and LF words—that is, 264 and 226, respectively, $t(30) = 1.68$, $p > .10$.[4] Turning to the flat trajectory conditions, HH words had a mean frequency of 356 in *Manulex* G1 and 335 in G3–5; LL words had a mean frequency of 0 in G1 and 6 in G3–5. The cumulative frequencies were 691 and 6, respectively. The frequency values found in *Lexique* (New, Pallier, Brysbaert, & Ferrand, 2004) were lower but they were maintained across the word categories.

Procedure

Participants rated the 64 word targets using an online web form. The words were presented in the same random order for each participant. Participants typed their AoA estimation in a text box under the word, which was displayed in lower case (e.g., for the word "*dragon*", the participant typed "3" if she or he thought that she or he learned "dragon" at the age of 3). Responses were automatically added to a spreadsheet for data analysis. The instructions followed the AoA-rating procedure of Ghyselinck,

[3] Lété, Peereman, and Fayol (2008) showed that GP and PG consistency values, calculated among words used in one and the same grade, were very similar across grades, thus indicating that the complexity of the French orthographic system is already captured in the early reader's vocabulary. Thus, words that are inconsistent for first-grade children remain inconsistent in higher grades, and there is no sharp modification in their levels of GP or PG ambiguity (see also Peereman et al., 2007, for a description of the French writing system at an infralexical level).

[4] Frequency values taken from *Lexique* (New, Pallier, Brysbaert, & Ferrand, 2004) instead of the *Manulex* G3–5 lexicon continued to reflect the critical difference between the two conditions. Level frequencies were 26 per million for LH words and 159 per million for LH words, $t(30) = -4.45$, $p < .01$, giving a cumulative-frequency value of 228 per million and 186 per million, respectively, $t(30) = 1.26$, $p > .10$.

Table 2. *Means and standard deviations for Experiments 1–6 as a function of of word type*

Tasks	*Experiment*		*High-to-high (2.8)*	*High-to-low (2.4)*	*Low-to-high (2.4)*	*Low-to-low (0.8)*
AoA rating	1					
Experiment 1A		Rating (years)	3.6 (0.8)	3.5 (0.9)	6.4 (1.4)	8.9 (1.5)
Experiment 1B		Rating (scale)	1.82 (0.26)	1.72 (0.33)	2.53 (0.40)	3.40 (0.63)
Immediate naming	2	Latency (ms)	510 (53)	520 (55)	519 (58)	557 (74)
		Errors (%)	3.9 (5.1)	1.2 (2.4)	1.4 (3.2)	6.0 (6.8)
Delayed naming	2	Latency (ms)	376 (83)	383 (84)	380 (77)	376 (76)
		Errors (%)	1.4 (2.6)	3.5 (5.7)	4.6 (5.3)	5.0 (7.3)
Lexical decision task	3	RT (ms)	503 (50)	521 (52)	517 (57)	599 (80)
		Errors (%)	1.8 (2.9)	1.0 (2.3)	0.8 (2.7)	16.8 (10.5)
Semantic task	4	RT (ms)	644 (85)	662 (87)	641 (79)	692 (91)
		Errors (%)	1.3 (2.6)	2.5 (4.0)	1.0 (2.3)	2.0 (3.0)
Progressive demasking	5	RT (ms)	1,241 (190)	1,244 (203)	1,289 (198)	1,441 (257)
		Errors (%)	3.0 (5.8)	0.5 (1.6)	3.2 (4.8)	6.4 (6.9)
Virtual experiment	6	RT (ms)	630 (28)	635 (51)	626 (52)	698 (77)
(by-item analysis)		Errors (%)	1.3 (2.0)	1.8 (2.1)	0.3 (1.0)	5.4 (5.8)

Note: Standard deviations in parentheses. Cumulative frequencies (log transformed) are in parentheses following word type. AoA = age of acquisition. RT = reaction time.

De Moor, and Brysbaert (2000; see also Ferrand et al., 2008; Stadthagen-Gonzalez & Davis, 2006, for recent uses of this rating procedure) and are reported in full in Appendix B.

Results

Mean AoA ratings for words were calculated across items for the by-participant analyses (F_1) and across participants for the by-item analyses (F_2). Analyses of variance (ANOVAs) were run with word type treated as a within factor in the by-participant analyses and as a between factor in the by-item analyses. In each analysis, a partial eta squared (η_p^2) was computed to estimate the strength of the effect size. Next, pairwise comparisons were performed with Bonferroni adjustment to the confidence intervals and significance. A Cohen's *d* (Cohen, 1988) was also computed in the by-participant analysis to estimate the strength of the difference. Age-limited learning effects were assessed by comparing the HL condition to the LH condition. We also compared HH and HL words and LH and LL words.

Table 2 shows the mean AoA ratings and the standard deviations corresponding to the four experimental conditions.

There was a significant main effect of word type on AoA ratings, $F_1(3, 84) = 301.70$, $p < .001$, $\eta_p^2 = .92$; $F_2(3, 60) = 61.92$, $p < .001$, $\eta_p^2 = .76$. Participants estimated HH words and HL words as being learned earlier than LH words and LL words. Ratings for HH words did not differ reliably from HL words. Ratings for LH words were significantly lower than those for LL words, both $p < .000$, $d = 1.75$ (i.e., huge effect in Cohen's, 1988, categorization). Importantly, the difference between HL and LH words was reliable, both $p < .001$, $d = 2.51$ (i.e., huge effect).

EXPERIMENT 1B

Method

Participants

Twenty psychology students from the University of Bourgogne (Dijon, France) took part in the experiment in exchange for course credits.

Stimuli
The words were the same as those used in Experiment 1A.

Procedure
The same instructions were given to the participants concerning the ratings of the AoA of the words except that they had to use a 5-point scale to provide their evaluation. A 5-point scale was printed below each word, and the different age bands were inserted in the scale instead of the numerical values 1–5.

Results

The same analyses as those described for Experiment 1A were performed for the data from Experiment 1B. The mean AoA ratings and their standard deviations for the different types of words are presented in Table 2.

A main effect of word type was found on the subjective AoA ratings, $F_1(3, 57) = 131.70$, $p < .0001$, $\eta^2_p = .87$; $F_2(3, 60) = 55.12$, $p < .0001$, $\eta^2_p = .73$. HH words and HL words were estimated to be learned earlier than LH words and LL words. Ratings for HH words did not reliably differ from HL words. Ratings for LH words were significantly lower than those for LL words, both $p < .001$, $d = 1.69$ (i.e., huge effect). As in Experiment 1A, the difference between HL and LH words was significant, both $p < .001$, $d = 2.27$ (i.e., huge effect).

Discussion

The results from Experiment 1 were clear-cut. Whatever the methods used to collect AoA ratings—that is, by providing a numerical value corresponding to an age of acquisition of the word (Experiment 1A) or by putting a cross on a 5-point scale (Experiment 1B)—HL words were judged to be acquired earlier than LH words. Indeed, the similarity of the results obtained with these two methods is striking as indicated by the Pearson's correlation, $r(64) = .95$, $p < .001$. We were therefore able to replicate the findings from the Zevin and Seidenberg (2002, 2004) and the Bonin et al. (2004) studies by showing that frequency trajectory (the difference between HL and LH words) has a strong and reliable influence on the AoA ratings provided by adults. Establishing such a relationship[5] was a necessary step before investigating the influence of age-limited learning in word recognition using word naming and lexical decision, respectively, in the following two experiments.

EXPERIMENT 2: INFLUENCE OF FREQUENCY TRAJECTORY IN WORD READING

As reviewed in the introduction, word reading studies in English have generally shown a reliable influence of AoA on word reading latencies (Johnston & Barry, 2006, for a review). However, the reanalyses conducted by Zevin and Seidenberg (2002) of previous word reading studies revealed that most of them did not properly control for cumulative frequency. Moreover, the investigation of both frequency trajectory and cumulative frequency on word reading latencies in factorial or regression studies has shown that, in the alphabetic languages of English and French, cumulative frequency, but not frequency trajectory, has a reliable influence on word reading latencies (Bonin et al., 2004; Zevin & Seidenberg, 2002, 2004). In Turkish, a very transparent orthographic system, Raman (2006) found a strong effect of AoA on reading latencies. However, in this study, subjective frequency, and not objective cumulative frequency, was controlled for. In contrast, in

[5] As far as the relationship between frequency trajectory and rated AoA is concerned, it is possible to ask why subsequent changes in frequency over time would affect the AoA of words. We accept that the frequency of exposure during childhood is undoubtedly the most important factor if a word is to be learned. One advantage of using frequency trajectory rather than directly using child frequency to investigate age-limited learning effects lies in the fact that frequency trajectory provides information concerning the way words have been encountered during childhood compared to adulthood for a given level of cumulative frequency. This relative information would be lost if child frequency was used. We thank Miguel Pérez very much for pointing out this interesting issue.

Italian, which is also a transparent orthographic system, Burani et al. (2007) found that frequency, but not AoA, affected word reading latencies. Thus, we have an inconsistent picture of the role of age-limited learning effects in word reading. However, all the studies that have found age-limited learning effects used classical AoA measures and did not properly control for cumulative frequency. In Experiment 2, we predicted that cumulative frequency, but not frequency trajectory, should have a reliable influence on word naming times. A delayed word naming production task was included in Experiment 2 to control for the impact of the articulatory characteristics of initial phonemes in immediate naming times, since it has been found that these account for a large proportion of the variance (Bonin et al., 2004; Morrison & Ellis, 2000).

Method

Participants

A group of 30 students (7 males) at the University of Bourgogne were tested (mean age: 23 years).

Stimuli

The experimental words were the same as those in Experiment 1.

Apparatus

The presentation of the stimuli was controlled by a Macintosh (iMac) computer running PsyScope v.1.2.5 software (Cohen, MacWhinney, Flatt, & Provost, 1993). The latencies were recorded via a microphone connected to the PsyScope button box.

Procedure

The participants were required to read aloud each word presented on the screen. An experimental trial had the following structure. First a cross was presented centred on the screen for 1,000 ms and was followed by a word presented in Chicago 48. The participants had to say aloud the word as fast as possible. The intertrial interval was set to 1,500 ms. In the delay word reading task, the same sequence of events took place except that the participants had to wait for a ready signal ("???") before saying aloud the word. From the onset of the word, the ready signal was presented after a random delay taken in the interval 1,000–1,300 ms. The experiment started with 10 warm-up trials.

Results

Trials with latencies longer than 1,500 ms and smaller than 200 ms were considered as outliers and were removed from further analyses (none in immediate naming and 5 trials in delayed naming, 0.3%). Next, any latency more than three standard deviations above a participant's mean was replaced by that value (26 trials out of 1,920 in immediate naming, 1.4% of trials; 19 trials out of 1,900 in delayed naming, 1.3% of trials). Table 2 shows mean latencies and standard deviations for the four conditions of the two naming tasks.

There was a main effect of word type on error rates, $F_1(3, 87) = 8.81$, $p < .001$, $\eta_p^2 = .23$; $F_2(3, 60) = 5.47$, $p < .002$, $\eta_p^2 = .22$. Pairwise comparisons showed that there was no significant difference between HH and HL. LH words produced fewer naming errors than LL words, both $p < .01$, $d = 0.88$ (i.e., large effect). Finally, there was no reliable difference between HL and LH.

As far as latencies are concerned, the main effect of word type was significant, $F_1(3, 87) = 36.56$, $p < .001$, $\eta_p^2 = .56$; $F_2(3, 60) = 5.36$, $p = .002$, $\eta_p^2 = .21$. Pairwise comparisons showed that HH words were named faster than HL words but significantly so only in the by-participant analysis, $p < .05$, $d = 0.19$ (i.e., small effect). LH words were named faster than LL words, both $p < .05$, $d = 0.58$ (i.e., medium effect). Finally, HL words did not differ reliably from LH words, both $p > .90$.

The delayed naming task revealed a main effect of word type on error rates, $F_1(3, 87) = 3.34$, $p < .05$, $\eta_p^2 = .10$; $F_2(3, 60) = 3.36$, $p = .05$, $\eta_p^2 = .14$. Pairwise comparisons yielded no significant differences between mean error rates. The mean latencies were flat over the four frequency trajectory conditions, both $Fs < 1$, both $\eta_p^2 < .03$, and none of the pairwise comparisons of interest reached significance.

Discussion

In line with previous word reading studies, we observed a reliable influence of cumulative frequency, but not of frequency trajectory (HL words compared to LH words), on word reading latencies (Bonin et al., 2004; Zevin & Seidenberg, 2002, 2004). It therefore now seems to be well established that, at least in alphabetic languages where the relationships between orthographic and sound units are quasisystematic, there are few, if any, age-limited learning effects in word reading latencies. As Zevin and Seidenberg (2002) suggested, when what is learned can assist the learning of new items, as is the case when learning the relationships between orthographic and phonological units in alphabetic systems, there is no long-lasting influence of early learning because generalization is possible. An important aspect of our findings is that a strong influence of cumulative frequency was observed in line with previous studies (Bonin et al., 2004; Zevin & Seidenberg, 2002, 2004).

EXPERIMENT 3: LEXICAL DECISION TASK

As mentioned in the introduction, studies investigating the influence of age-limited learning effects in lexical decision have consistently found an effect of this variable when indexed with AoA ratings (e.g., Bonin et al., 2001; Brysbaert, Lange, et al., 2000; Colombo, & Burani, 2002; Gerhand & Barry, 1999b; Whaley, 1978). However, the few studies that have considered frequency trajectory have yielded inconsistent findings (e.g., Bonin et al., 2004; Caza & Moscovitch, 2005). Therefore, it is important to reconsider the influence of frequency trajectory in lexical decision. Moreover, this task is one of the most popular tasks employed to investigate word recognition.

Method

Participants

Twenty-five students (two males; mean age 22 years old) from the University of Lyon took part.

Materials

The 64 words from in Experiments 1 and 2 were used together with 64 legal nonwords, which were 7, 8, and 9 letters long. The nonwords were created from a set of words by substituting a minimum of two letters. All respected the phonotactic constraints of French.

Procedure

Each participant was seated at a fixed distance of 60 cm in front of a 17″ colour monitor connected to a Pentium III laptop computer running DMDX software (Forster & Forster, 2003).[6] The stimuli were displayed in lower case in 24-point Courier font with a 640 × 480 resolution. The participants were tested individually in single 10-minute sessions.

Each trial consisted of the following sequence of events. The participant was first instructed to look at a fixation point ("+") at the beginning of each trial. After 1,000 ms, the fixation point was replaced by a target centred on the screen. The target remained on the screen until the participant responded by selecting either the word response (right shift key) or the nonword response (left shift key on the keyboard). He or she was instructed to respond as quickly as possible, while avoiding errors. There was one block of 10 practice trials followed by four blocks of 32 experimental trials (16 nonwords and 16 words with four targets in each

[6] It could be asked whether the use of different softwares (PsyScope versus DMDX) or different computers or keyboards might influence the precision of the collected reaction time (RT) data and the comparability of different results obtained across experiments. However, as shown by Damian (2010), given the variability of human performance in standard behavioural tasks, such as those used in the current study, the measurement errors related to the imprecision of input devices (e.g., computer keyboards) are bound to be very small.

experimental condition). In each block, the targets were presented in a different random order to each participant.

Results

One trial with a reaction time (RT) below 300 ms was considered as an outlier and was discarded from the analyses. Any RT above three standard deviations of a participant's mean was replaced by that value (18 trials over 1,600, 1.2% of trials). Table 2 shows the mean RTs, their standard deviations, and error rates for the four conditions in the lexical decision task.

The error rates were very low for HH, HL, and LH words (2%, 1%, 1%, respectively) but increased for LL words (16%) and resulted in a main effect of word type, $F_1(3, 72) = 48.10$, $p < .000$, $\eta_p^2 = .67$; $F_2(3, 60) = 9.77$, $p < .000$, $\eta_p^2 = .33$. Pairwise comparisons showed a significant difference between LH and LL words only, both $p < .001$, $d = 2.13$ (i.e., huge effect).

On RT data, there was a significant main effect of word type, $F_1(3, 72) = 73.64$, $p < .001$, $\eta_p^2 = .75$; $F_2(3, 60) = 23.92$, $p < .001$, $\eta_p^2 = .55$. Pairwise comparisons showed that the difference between HH and HL words was significant in the by-participants analysis only ($p < .001$), $d = 0.36$ (i.e., small effect). The difference between LH words and LL words was significant in both types of analyses, $p < .000$, $d = 1.20$ (i.e., very large effect). Finally, there was no hint of an influence of frequency trajectory since the difference between HL and LH words was not reliable, both $p > .90$.

Discussion

As in Experiment 2, a reliable influence of cumulative frequency, but not of frequency trajectory, was found on lexical decision times for words. This finding contrasts with what Bonin et al. (2004) observed in their regression analysis of lexical decision times in French. However, the words used in Bonin et al. (2004) corresponded to object names, and the nonwords were wordlike nonwords. It is also possible to conjecture that the use of object names rendered semantics more salient. As a result, frequency trajectory, which is assumed to index semantic–lexical links, might emerge more readily when object names are used. The words used in our study were less concrete than those used in the Bonin, Van Wijnendaele et al. (2004) study since in the latter research the items all referred to concrete objects. Although this type of account is tempting, it remains speculative. Nevertheless, the fact that recent studies have shown that the involvement of semantic codes in visual lexical decision is subject to context modulations lends it some plausibility (i.e., the type of nonwords; see, for example, Evans, Lambon Ralph, & Woollams, 2012). Our experiment is not able to address the issue of how, precisely, the type of words and nonwords used influences lexical decision latencies. Logically, to avoid any influence of nonwords on word decisions, we need to conduct a decision task in which nonwords are not used. In Experiment 4, we used a task of this type. This task is referred to as "proper/common name" decision task.

EXPERIMENT 4: PROPER/COMMON NAME DECISION

In the "proper/common name" decision task, participants are presented with both proper names and common names and have to categorize them according to whether they refer either to a category of words with a definable meaning ("*door*") or to the category of first names ("*Patrick*"). This task has already been used by Brysbaert, Van Wijnendaele et al. (2000) to investigate AoA effects in word recognition. Using this task, the authors found AoA effects and accounted for them by assuming that they stemmed from the semantic level and not from the phonological level since Taft and van Graan (1998) had previously found no phonological effects with this task (i.e., regularity was used as an index of phonological activation). However, Brysbaert, Van Wijnendaele et al. (2000) did not find a reliable effect of imageability in the proper/common name decision task, an observation that they attributed to their use of a rather restricted range of imageability values in their stimuli. The logical prediction is therefore that if this task is able to tap into the semantic system, frequency trajectory effects should be found. This task can be seen as a variant of the lexical decision task. However, it has the advantage that it is not impaired by the potential influence of nonwords when participants are required to make binary decisions.

Method

Participants

Twenty-five students (seven males; mean age 21 years) from the University of Lyon took part.

Materials

The 64 experimental words were used together with 64 first names taken from *Lexique* (New et al., 2004) and had a mean frequency of occurrence of 21 per million words. First names were matched to the experimental words for length (mean of 7.5 letters).

Procedure

Each participant was seated at a fixed distance of 60 cm in front of a 17″ colour monitor connected to a Pentium III laptop computer running DMDX software (Forster & Forster, 2003). The stimuli were displayed in lower case in 24-point Courier font with a 640 × 480 resolution. The participants were tested individually in single 10-minute sessions. Each trial consisted of the following sequence of events. The participant was first instructed to look at a fixation point ("+") at the beginning of each trial. After 1,000 ms, the fixation point was replaced by a target centred on the screen. Participants were instructed to decide whether the target belonged to the category "word with definable meaning" or to the category "first name". The target remained on the screen until the participant responded by selecting either the "definable meaning" response (right shift key) or the "first name" response (left shift key on the keyboard). (The keys were inverted for the left-handed participants.) He or she was instructed to respond as quickly as possible, while trying not to make any mistakes. There was one block of 10 practice trials followed by four blocks of 32 experimental trials (16 first names and 16 experimental words with four targets in each experimental condition). In each block, the targets were presented in a different random order to each participant.

Results

Since no extreme scores were found in the RT data (responses faster than 300 ms or slower than 3,000 ms), any RT more than three standard deviations above a participant's mean was replaced by the mean (43 trials over 1,600, 2.7% of trials). Table 2 shows the mean RTs, their standard deviations, and error rates for the four frequency trajectory conditions. The error rates were very low across conditions and revealed no significant main effect of word type.

On RT data, there was a significant main effect of word, $F_1(3, 72) = 9.16$, $p < .001$, $\eta_p^2 = .28$; $F_2(3, 60) = 2.93$, $p < .05$, $\eta_p^2 = .13$. Pairwise comparisons indicated no significant difference between HH and HL words. The difference between LH and LL words was significant in the by-participants analysis only ($p < .01$), $d = 0.56$ (i.e., medium effect). Finally, there was no significant difference between HL and LH words, both $p > .18$.

Discussion

Once again, we did not find a reliable influence of frequency trajectory on RTs in a task that indexes word recognition. In contrast, as in Experiments 1–3, cumulative word frequency was reliable. The proper/common name decision task was chosen because in the same way as the lexical decision task, it is a binary decision task, which has the advantage that it does not require nonwords that can influence the type of processing performed on words. Also, since Brysbaert, Van Wijnendaele et al. (2000) thought that the AoA effects observed in this task were due to semantic code activation, we thought it appropriate to use this task to index the influence of frequency trajectory. In effect, age-learning effects are clearly observed in tasks where semantic codes are obligatory such as in spoken or in written naming. It should be remembered, however, that imageability, which has been thought to be a genuine semantic variable, was not reliable in the Brysbaert, Van Wijnendaele et al. (2000) findings. Finally, it could be argued that relying only on the lexical decision task (or on a variant such as the proper/common name decision task) to investigate word recognition is problematic given that this task can be performed on the basis of global lexical activation (Carreiras et al., 1997; Grainger & Jacobs, 1996) and not only through the activation of individual lexical entries. What is needed is a visual word recognition task that unambiguously makes use of the characteristics of individual lexical entries. In the word recognition literature, it has been assumed that the progressive-demasking task is performed on the basis of individual lexical entries (Dufau, Stevens, & Grainger, 2008). Moreover, this task has been found to be more sensitive to word frequency and neighbourhood effects (Dunabeitia, Avilés, & Carreiras, 2008). We therefore used this task to further test whether there is an influence of frequency trajectory in word recognition.

EXPERIMENT 5: PROGRESSIVE-DEMASKING TASK

As described by Dufau et al. (2008), the progressive-demasking task manipulates the display duration of a masking stimulus of hash marks (#####) and a target stimulus. Each trial involves several cycles consisting of the successive presentation of the mask (i.e., the noise) followed by the target (i.e., the signal). On successive cycles within a given trial, the signal-to-noise ratio is increased by increasing target duration and decreasing mask duration, giving the feeling that the target emerges from the mask. At the beginning of a trial (i.e., on the first cycle), the target duration is typically set at a minimal value close to zero, and the mask duration is set at some upper limit (200 ms in our study). As the cycles follow each other, the mask duration decreases while the target duration increases by the same proportion—that is, Cycle 1: 195–15; Cycle 2: 180–30; Cycle 3: 165–45, and so on until the target is identified by the participant. Responses are made by pressing the space bar and by speaking the target word aloud.

Method

Participants

A group of 25 students (12 males) from the University of Lyon were recruited (mean age: 21 years).

Stimuli

The words were the same as those in the previous experiments.

Procedure

The participants were seated at a fixed distance of 60 cm in front of a 17″ colour monitor connected to a Pentium III laptop computer running the PDM software (Dufau et al., 2008). The stimuli were displayed in lower case in 24-point Courier font with a 640 × 480 resolution. The participants were tested individually in a single 15-minute session.

Each trial consisted of the following sequence of events. The participant was first instructed to look at a fixation point ("+") at the beginning of each trial. After 1,000 ms, the fixation point was replaced by the pair [######-word] centred on the screen (the number of hash marks was equal to the number of letters in the target). The cycles continued until the participant pushed the space bar when the target was thought to have been recognized. The experimenter typed the answer on a second keyboard connected to the laptop computer. The participants were instructed to respond as quickly as possible—namely, as soon as "the word comes to mind".

In the experimental session, the participants were invited to rest between two trials when they felt the need. The 64 words were presented in a different random order to each participant. There were 20 practice trials.

Results

Trials with hesitations, stutters, and responses that did not correspond to the expected word were classified as errors. After finding no extreme scores in the RT data (responses faster than 300 ms or slower than 3,500 ms), any RT more than three standard deviations above a participant's mean was replaced by that value (19 trials over 1,600; 1.2% of trials). Table 2 shows mean RTs and error rates for the four conditions in the progressive-demasking task.

The analyses of the error rates revealed a main effect of word type, $F_1(3, 72) = 8.07$, $p < .001$, $\eta_p^2 = .25$; $F_2(3, 60) = 4.72$, $p < .005$, $\eta_p^2 = .19$. Pairwise comparisons showed no significant difference over the three comparisons of interest.

As far as the RT data are concerned, the main effect of word type was significant, $F_1(3, 72) = 42.01$, $p < .001$, $\eta_p^2 = .64$; $F_2(3, 60) = 12.95$, $p < .001$, $\eta_p^2 = .39$. Pairwise comparisons showed that the difference between HH and HL words was not significant, but that the difference between LH and LL words was, both $p < .001$, $d = 0.68$ (i.e., medium effect). Finally, no significant difference was found between HL and LH words, both $p > .14$.

Discussion

In Experiment 5, again, we did not find a reliable influence of frequency trajectory (the comparison between HL and LH words). However, the cumulative frequency of the words had a reliable influence on the time taken to recognize them. The progressive-demasking task was used in this experiment to test for frequency trajectory because other studies of word recognition have shown that frequency effects and neighbourhood effects are larger in the former task than in lexical decision or in word naming (Dunabeitia et al., 2008). Thus, it seems difficult to argue that our null results concerning the impact of frequency trajectory reflect a lack of sensitivity of the tasks we used to investigate word recognition. Moreover, and importantly, the progressive-demasking task has been employed to investigate the influence of number of associates in word recognition (Dunabeitia et al., 2008). The number-of-associates variable is assumed to index semantic code activation in word recognition. Dunabeitia et al. (2008) found that these effects were reliably observed in the progressive-demasking task, and a close examination of these effects reveals that they are larger than those observed in reading or in lexical decision. Thus, since the progressive-demasking task is sensitive to semantics, as suggested by the reliable effects of number of associates, it cannot be argued that the lack of frequency trajectory effects is due to the choice of tasks that are not sufficiently sensitive to detect the activation of semantic codes. If age-limited learning effects are genuine effects in word recognition (when cumulative frequency is taken into account), they should have emerged in some, if not all, the tasks we used to study word recognition.

VIRTUAL LEXICAL DECISION EXPERIMENT AND MULTIPLE REGRESSION ANALYSIS OF LEXICAL DECISION TIMES TAKEN FROM THE FRENCH LEXICON PROJECT (Ferrand et al., 2010)

Advocates of the use of AoA ratings to investigate age-limited learning effects in lexical processing might claim that our demonstration falls short of the stated goal of the current research. Such readers will therefore be sceptical of our findings because of the factorial approach we have employed so far. Indeed, this approach has sometimes been qualified as being too crude to investigate such effects (P. Monaghan, personal communication, January 26, 2011). In effect, in our experiments, we used relatively small numbers of items per condition due to the demands of experimental designs that make it necessary to control for a large number of variables. However, it is important to stress that the approach we have followed is a traditional one in the psycholinguistic field. For many years researchers have relied on experiments that manipulate factorially word properties while attempting to hold other factors constant. A potential limitation of the factorial approach to the study of word processing is that most experiments use monosyllabic words even though these words only represent a small part of the lexicon. Another way to investigate age-limited learning effects in word recognition is the simulation approach. Recently, Monaghan and Ellis (2010) designed a number of simulations relating to this issue in word reading and showed that the order of acquisition of the items does have a small effect if measured precisely (our assumption is that simulations are of use when they are substantiated by empirical data). Finally, another way to search for frequency trajectory effects in word recognition is to use larger stimulus samples, which include words of lower frequency than those used here. As Keuleers, Lacey, Rastle, and Brysbaert (2012) pointed out, in the last decade, there has been a steady increase in the design of so-called "megastudies" in which RTs are collected for an impressively large number of words. Fortunately, as far as French is concerned, a study of this type has been designed by Ferrand et al. (2010) who collected lexical decision times in French for a set of 38,840 words: the French Lexicon Project (FLP). Likewise, it was possible to examine the influence of frequency trajectory using a multiple regression approach and a large set of words. According to Sibley, Kello, and Seidenberg (2009), small-scale experiments make it possible to identify the *existence of an effect* but do not provide any information about its size. Megastudies are the most suitable way to obtain information about the *size of an effect* (Sibley et al., 2009). As far as frequency trajectory is concerned, although the size of this effect in word recognition is certainly interesting, it seems less important than the fact of its existence because, as reviewed in the introduction, it is the question of whether or not age-limited learning effects actually exist in visual word recognition that is disputed. Most influential models of visual word recognition do not take this effect into account (we return to this issue in the General Discussion). However, some researchers still feel uncertain about effects that are obtained using small groups of words, as is the case here (M. Brysbaert, personal communication, January 26, 2011). Thus, one compromise solution is to use both approaches, as we have done in the current research. Below, we report a multiple regression analysis that aimed at determining the influence of frequency trajectory in lexical decision times taken from the FLP. However, before reporting the results from the multiple regression analysis, we examine whether the pattern of results found in the lexical decision task used in Experiment 3 can be replicated in a virtual experiment performed using the RTs taken from the FLP database.

EXPERIMENT 6: VIRTUAL EXPERIMENT: ARE THE FINDINGS FROM EXPERIMENT 3 REPLICATED IN A VIRTUAL EXPERIMENT?

To design factorial experiments, it is necessary to select stimuli that vary on dimensions of interest and which are at the same time controlled on certain other important dimensions and, then, to collect behavioural data. Virtual experiments are interesting in that they are less time consuming because they do not require the collection of behavioural data given that it is possible to design an experiment by selecting items directly from the corpus of RT data. However, virtual experiments have been criticized because inconsistent findings are obtained when compared to the more classical approach involving "real" experiments. Indeed, Sibley et al. (2009) found that five major studies of word naming in English would have yielded inconclusive results with the use of this experimental strategy. According to these authors, the virtual experiment methodology has limited utility because megastudies are too noisy to allow fine-grained analyses. In contrast, the opposite claim has recently been made by Keuleers et al. (2012), who found a high correlation between two English megastudies. This finding suggests that a high percentage of the variance in megastudies is systematic rather noise induced. According to these authors, the recourse to virtual experiments is not problematic but is instead useful! We cannot rule out the possibility that the failure to detect frequency trajectory effects in our "small-scale" experiments was due to the naming of this specific of small set of words as well as to the observation of the size of the difference between high- and low-frequency words. What happens, therefore, when this set of words is processed within a large set of words? This question was addressed in the following analysis by comparing the effects of both word frequency and frequency trajectory on lexical decision times obtained in the factorial experiment (Experiment 3) with those obtained on the basis of the lexical decision times taken from the FLP (Ferrand et al., 2010).

The RTs of our set of 4×16 experimental words were extracted from the French Lexicon Project database. Four words from the LL condition (low-to-low-frequency words) were not found in the database (*astronome*, *baudet*, *obélisque*, *poivron*). Given that this set of words produced very low RTs, and that this condition was not critical for our purposes, we considered that the nonbalanced plan did not impair the robustness of the present analysis. The by-item ANOVA (F_2) revealed a main effect of word type, $F_2(3, 56) = 5.30$, $p = .003$, $\eta^2_p = .22$. Pairwise comparisons showed no significant difference between HH and HL words. The difference between LH and LL words was significant ($p < .01$), $d = 1.13$ (i.e., very large effect). Finally, there was no significant difference between HL and LH words, $p > .10$.

Thus, exactly the same pattern of results was found for the items in both the virtual and the real lexical decision experiment (Experiment 3).

MULTIPLE REGRESSION ANALYSES ON 26,394 WORDS FROM THE FRENCH LEXICON PROJECT

RTs for the 38,840 words of the French Lexicon Project were merged with word-frequency measures taken from the grade-level databases *Manulex* (Lété et al., 2004) and *Manulex-infra* (Peereman et al., 2007) and from the adult database *Lexique* (New et al., 2004). Cumulative frequency and trajectory frequency were computed as in Bonin et al. (2004). Cumulative frequency was calculated as the *sum* of the *z* scores of the *Manulex* child measure corresponding to all grades (1st grade to 5th grade) and of the adult measure taken from *Lexique*. Frequency trajectory was computed as the difference between the *z* scores associated with the two measures of frequency (*Lexique* minus *Manulex*). As explained in Bonin et al. (2004), *z* scores instead of raw (or log-transformed) frequencies were used because the sizes of *Lexique* and *Manulex* databases are not the same, a fact that might otherwise led to discrepancies between the two measures of word frequency.

Apart from the cumulative-frequency and frequency trajectory variables, three other classical variables characterizing word processing difficulty were included in the regression model: the number of letters in the words, the N-size of the words (number of orthographic neighbours taken from *Manulex-infra* and computed across the five grade levels), and the grapheme–phoneme (GP) consistency of the words.

The correlation matrix between the set of predictors (see Table 3) showed that they were well suited for describing RTs. The rule of thumb generally adopted for regression studies is that no pairwise correlation should be higher than .70 (Baayen, Feldman, & Schreuder, 2006), which was the case in our correlation matrix. It is worth noting that the two predictors of interest—that is, frequency trajectory and cumulative frequency—were uncorrelated ($r = .003$).

Table 4 presents the results of the stepwise regression (forward method). Frequency trajectory did not contribute significantly to the model. Cumulative frequency was entered first in the model with about 21% of explained variance. The number of letters significantly slowed down RTs,

Table 3. *Pairwise correlations between the predictors used in the analysis*

	LET	*NEI*	*GPC*	*TRAJ*
LET	—			
NEI	−.17	—		
GPC	−.13	.05	—	
TRAJ	−.04	−.03	.01	—
CUM	−.35	.28	.04	−.00

Note: $N = 26{,}394$. CUM = cumulative frequency; GPC = grapheme-to-phoneme consistency; LET = number of letters; NEI = neighbourhood size; TRAJ = frequency trajectory.

Table 4. *Summary of the stepwise regression analysis*

Step	*Predictor*	β	t	p	*Adjusted* η^2	η^2 *change*
1	CUM	−.45	−82.57	.000	.205	.205
2	LET	.21	36.42	.000	.243	.038
3	NEI	−.08	−13.36	.000	.248	.005
4	GPC	−.02	−4.09	.000	.249	.000

Note: $N = 26{,}394$. Probability of F at .05 for entry and at .10 for removal. CUM = cumulative frequency; GPC = grapheme-to-phoneme consistency; LET = number of letters; NEI = neighbourhood size.

and the N-size significantly speeded up RTs. The GP consistency had a small but significant facilitatory effect on RTs.

To summarize, the findings from both the virtual experiment and the large-scale regression analysis led to the same conclusion as the previous small-scale visual word recognition experiments. Cumulative frequency made a reliable contribution whereas frequency trajectory did not.

GENERAL DISCUSSION

In the current study, we tested age-limited learning effects in word recognition using four different tasks: word naming (immediate and delayed word naming), lexical decision, proper/common name decision, and perceptual identification. Four categories of words were tested across the four tasks with different participants: HL and LH words as well as HH and LL words. In contrast to previous studies that used only words having low-to-high- and high-to-low-frequency trajectories to investigate word recognition, we included words that had flat frequency trajectories. Importantly, across the frequency trajectory manipulation corresponding to HL and LH frequency words, the cumulative frequency of the words was controlled for. The results were clear-cut. Strong and reliable effects of cumulative frequency, but not of frequency trajectory (i.e., the frequency trajectory manipulation corresponding to HL and LH frequency words), were found on reaction times across the tasks used to investigate word recognition. The lack of an effect of frequency trajectory in these tasks cannot be due to the fact that the frequency trajectory contrast we used was not large enough, since in Experiment 1, we found strong and significant effects of frequency trajectory on AoA ratings in two different groups of participants who used two different methods to provide their AoA ratings (either by indicating a numerical value corresponding to an age or by putting a cross on a 5-point scale). Moreover, the pattern of results found in lexical decision was replicated in a virtual experiment (Experiment 6) using the RTs from the FLP (Ferrand et al., 2010), thus establishing that these effects are robust and are not due to a particular processing strategy induced by the use of highly selected words. Importantly, when large sets of items were used, the multiple regression analysis did not provide evidence for an influence of frequency trajectory in lexical decision whereas a strong effect of cumulative frequency was found.

As far as the word recognition experiments are concerned, since the same words were used across the different tasks, we performed a combined analysis of the data using a z score transformation to gain an overall picture of the effect of frequency trajectory across tasks. However, the data corresponding to the proper/common name decision in Experiment 4 were not taken into account in these analyses because (a) this task is rarely used in word recognition studies, and (b) this approach simplifies the presentation of the overall picture revealed by these analyses. It is important to note that the omission of this task does not alter the general conclusion that can be derived from this analysis.

As can be seen from Figure 1, the effect of frequency trajectory varied as a function of the tasks used to assess its influence. The interaction effect

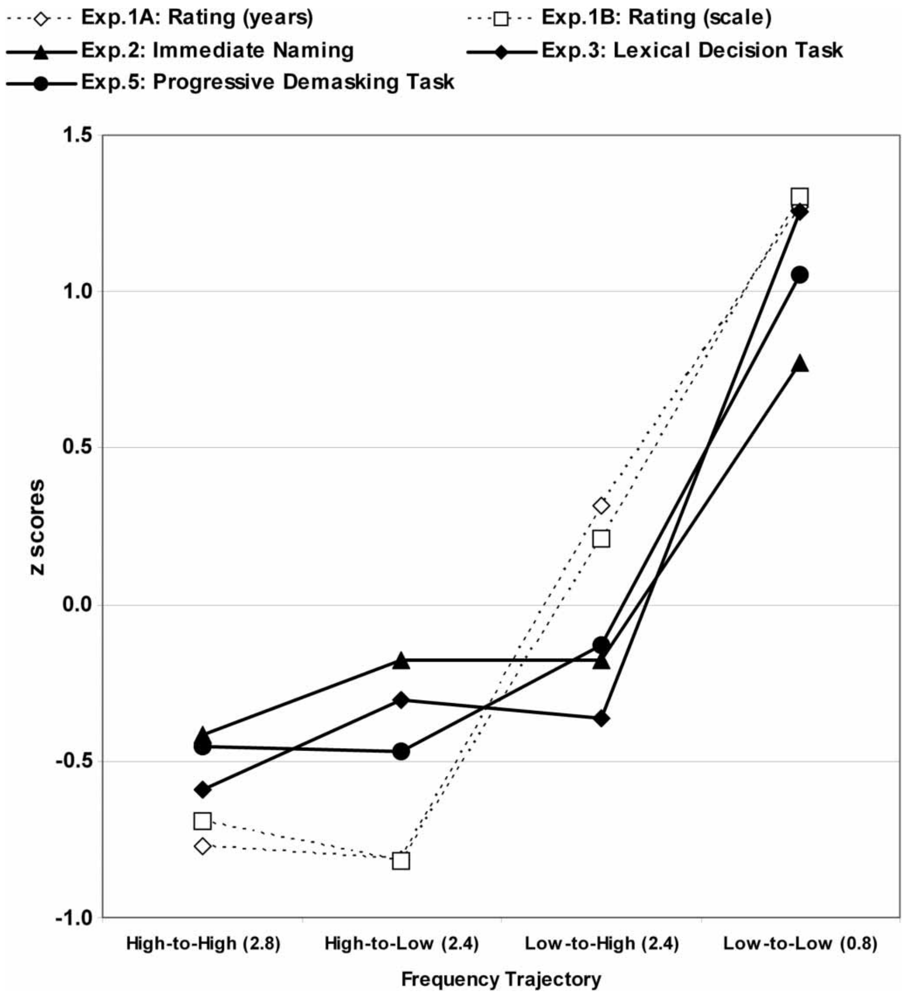

Figure 1. *The z scores for Experiments 1, 2, 3, and 5. Cumulative frequency (log transformed) is in parentheses.*

between type of tasks and frequency trajectory was reliable, $F_2(12, 240) = 3.49$, $p < .001$. The analysis on the z scores confirms that an influence of frequency trajectory was found in the AoA rating tasks but not in the online word recognition tasks (word naming, lexical decision, and perceptual identification). Moreover, this pattern of results was confirmed when the analysis on the z scores is restricted to HL and LF words, $F_2(4, 120) = 6.51$, $p < .001$.

How do we relate our findings to those in previous reports of AoA effects in word naming? It should be remembered that previous reports of AoA effects in word naming in English did not properly control for the cumulative frequency of early- and late-acquired words. When these data were reanalysed with cumulative frequency introduced as a factor, AoA was no longer reliable (Zevin & Seidenberg, 2002). The only study that has claimed that universal AoA effects occur in lexical processing based on word reading data is that of Raman (2006). In effect, because Turkish has very transparent orthography, the observation of AoA effects in this language was taken to argue that these are widespread and not confined to low-frequency irregular words as has been observed in English (Monaghan & Ellis, 2002). However, in the Raman study, AoA effects were found with the use of AoA ratings rather than with a frequency trajectory manipulation. More importantly, Raman (2006) employed subjective frequency ratings to control for word frequency. Subjective frequency ratings, such as AoA ratings, are behavioural outcomes and are, therefore, not objective measures of the frequency with which the words are encountered. Subjective frequency estimations are subject to the same kind of criticisms as those levelled at AoA ratings. These, too, are based on some kind of metalinguistic knowledge and, in our view, should not be used to predict reaction times.

The idea that semantics are involved in lexical decision, and to a greater extent than in word reading, has been argued by Cortese and Khanna (2007) based on the observation that imageability (a putative semantic variable) constitutes a strong determinant of lexical-decision latencies but not of word-reading latencies. Cortese and Khanna also found a stronger effect of the AoA variable in lexical decision than in word reading. It should be remembered that a large number of studies have reported AoA effects in lexical decision when subjective AoA ratings were used (e.g., Bonin et al., 2001; Gerhand & Barry, 1999b; Morrison & Ellis, 1995). However, the picture is different when frequency trajectory is used to index age-limited learning effects. Caza and Moscovitch (2005) did not find a reliable effect of frequency trajectory on word decision times but did obtain a cumulative-frequency effect in both healthy older adults and patients with Alzheimer's disease. In contrast, Bonin et al. (2004) found a reliable influence of both frequency trajectory and cumulative frequency on decision times in their regression analysis. As discussed above, the discrepancy between the findings relating to frequency trajectory in lexical decision in the current study and that conducted by Bonin et al. might be due to the use of concrete object names in the latter study. In effect, because the words in Bonin et al.'s study were chosen to be able to be pictured for spoken/written naming, they were all concrete and referred to objects. Concrete object names might render semantic information more salient, and this information could be used to perform the task. The nature of the nonword stimuli used could also be responsible for the discrepancy. Whatever the reason that is ultimately found to account for the different pattern of results observed in lexical decision, our findings speak for the use of a multitask approach to the study word of recognition. Finally, in Experiment 4 a decision task that did not require the use nonwords was used (proper/common name decision), and again we did not find a reliable influence of frequency trajectory.

In Experiment 5, we used the progressive-demasking task to test whether the lack of an effect of frequency trajectory was due to the fact that the word naming and lexical decision tasks were not sufficiently sensitive to detect such an effect. Since the number of associates has been found to exert an effect in the progressive-demasking task (Dunabeitia et al., 2008), we thought it appropriate to use this task to reveal the activation of semantic codes. Because frequency trajectory is assumed to index the activation of lexical–semantic links, an effect of this variable should have been observed in the progressive-demasking task. However, once again, we did not observe a reliable effect of frequency trajectory on identification times.

Since we did not control for the imageability of our stimuli, it could be argued that the lack of an effect of frequency trajectory on lexical decision, word reading, and perceptual identification times was due to the fact that imageability values for HL and LH words washed our any effect of frequency trajectory. If this were indeed the case, it would be difficult to explain why there was a reliable frequency trajectory effect on AoA ratings, except if we submit that AoA ratings are not sensitive to imageability. However, AoA ratings have been found to be reliably predicted by imageability as well as by other factors such as conceptual familiarity and cumulative frequency (Bonin et al., 2004). However, in order to ascertain that the lack of an effect of frequency trajectory was not due to a confound with imageability, we collected imageability scores on our stimuli. The instructions used for the collection of imageability scores were the same as those used in the Bonin et al. (2003) study. A group of 20 adults were involved for the collection of the norms (none of them had participated in the previous experiments). HL words had a mean imageability score of 6.23, and LH words a score of 3.71 (for the HH and LL words, the scores were 5.27 and 4.56, respectively). An ANOVA run on imageability scores yielded a main effect of word type, $F_1(3, 57) = 76.43$, $p < .001$, $\eta_p^2 = .80$; $F_2(3, 60) = 9.37$, $p < .001$, $\eta_p^2 = .32$. Pairwise comparisons showed that the difference between HL (6.23) and LH words (3.71) was reliable, both $p < .001$, $d = 2.69$ (i.e., huge effect). The differences between HH words (5.27) and HL words (6.23), and between LH words (3.71) and LL words (4.56) were significant only in the by-participant analysis, both $p < .001$, $d = 1.44$ and 0.77, respectively (i.e., large effect for both). Thus, the lack of an effect of frequency trajectory in lexical decision, word reading, and perceptual identification cannot be due to imageability washing out any frequency trajectory influence. Moreover, following Ghyselinck et al. (2004), it should be recalled that word recognition studies in which both AoA and imageability were manipulated consistently found robust effects of rated AoA when stimulus words were controlled for imageability but very small or no reliable imageability effects when stimuli were matched on AoA (e.g., Brysbaert, Lange, et al., 2000; Brysbaert, Van Wijnendaele, et al., 2000). Finally and importantly, when the combined analysis reported above with *z* scores and the factors type of tasks and frequency trajectory was performed with the imageability factor introduced as a covariate factor, the pattern of results reported previously remained the same.

We wish to emphasize that the criticisms that have been raised against the use of adult AoA ratings to predict reaction times in a number of lexical tasks also apply to the use of imageability scores. Just like AoA ratings, imageability ratings are behavioural outcomes. Although it has been assumed that imageability scores are a valid index of the activation of semantic codes in visual word recognition (Cuetos & Barbon, 2006; Shibahara, Zorzi, Hill, Wydell, & Butterworth, 2003; Strain, Patterson, & Seidenberg, 2002), no systematic studies have attempted to determine whether these scores are a true index of semantic code activation. In a recent study, Bonin, Méot, Ferrand, and Roux (2011) found that the correlation between imageability scores and the number of semantic features (the latter being assumed to reflect semantic richness), though reliable, was not large. We suggest that the task of rating words for imageability is not based on exactly the same type of information as the task of rating words for AoA, and that the former task is easier to figure out than the latter. It is clear that this issue requires further research.

Did we really miss any frequency trajectory influence in visual word recognition?

Strong claims have been made in the past in an attempt to persuade theorists of word recognition to modify their models in a way that integrates the influence of AoA effects. However, we hope that we have established, on the basis of the small-scale experiments and large-scale multiple regression analysis reported here, that any effect of frequency trajectory, if it exists, is at the very least not easy to detect (since it is logically impossible to prove the nonexistence of an effect, it cannot be concluded that this effect does not exist!). A number of previous AoA effects in word recognition studies have been due to differences in cumulative frequency (Zevin & Seidenberg, 2002). As far as our stimuli are concerned, we have shown that the frequency trajectory manipulation covaries with rated AoA. Could the lack of an effect of frequency trajectory be due to the fact that our frequency trajectory/AoA manipulation was not large enough to detect such effects in lexical decision, word reading, and so on? If we consider the earlier study conducted in French by Bonin et al. (2001), strong AoA effects were found on RTs in a lexical decision task in which adult word frequency was controlled for, and the difference between the early- and late-acquired words was very close to that observed here with our present stimuli. However, in Bonin et al.

(2001), adult frequency and not cumulative frequency was controlled for. Nevertheless, if the lack of a reliable effect was due to the selection of a restricted set of stimuli, then one question that remains to answered is why the effect was still not reliable when we ran a multiple regression analysis on more than 20 thousand words?

What are the implications of our findings for word recognition models?

The finding of AoA effects in word recognition was long taken as representing a challenge to the most influential word recognition models and, more specifically, word reading models (Zevin & Seidenberg, 2002). As Sibley et al. (2009) claim, small-scale experiments, such as those used in the current study, make it possible to identify the existence of an effect. At a theoretical level, the existence of true age-limited learning effects in visual word recognition is critical for most influential models of word recognition, in the same way that the existence of true regularity effects are necessary to distinguish between dual-route models and single-route models of word reading (M. Coltheart et al., 2001; Plaut et al., 1996; see Sibley et al., 2009). Across a series of six experiments reported here (five real-time and one virtual experiment), we were unable to find a reliable effect of frequency trajectory in visual word recognition when cumulative frequency was taken into account. The use of large sets of words and a multiple regression approach did not change the outcome reached by the factorial approach. Again, we did not find a reliable impact of frequency trajectory. It should be recalled that establishing the existence of age-limited learning effects in visual word recognition has been considered more important than determining the size of these effects given that most influential models of visual word recognition do not take account of them at all. Proponents have vigorously criticized the way AoA effects have been ignored in word reading (e.g., Monaghan & Ellis, 2010) and have been vociferous in claiming that most word recognition models have to be altered to account for them.

In effect, AoA effects have been taken into account neither in the dual-route view (the dual-route cascaded, DRC, model of M. Coltheart et al., 2001), nor in the connectionist single-route view of word recognition (e.g., Harm & Seidenberg, 1999; Plaut et al., 1996; Seidenberg & McClelland, 1989). In the DRC model, there are two routes, the lexical and nonlexical routes, which are required for the processing of familiar words, unknown words, and nonwords. The nonlexical route applies grapheme-to-phoneme correspondence rules in a serial left-to-right manner. As far as the lexical procedure is concerned, there are indeed two lexical routes: a direct lexical route, which maps orthographic lexical representations to phonological lexical representations, and an indirect lexical route, in which semantic mediation takes place. The connectionist single-route view of word recognition assumes the existence of two pathways from spelling to sound: One pathway is a direct mapping from orthographic to phonological representations, whereas the second pathway maps from print to sound by means of the representation of word meanings. Since a phonological locus has been proposed to explain AoA effects in word recognition (an orthographic locus has been put forward to account for frequency effects in lexical decision and word reading, Gerhand & Barry, 1998, 1999a, 1999b), the DRC model, which includes both an orthographic input lexicon and an phonological output lexicon, could in principle be modified to account for these effects in both lexical decision and word reading. However, the DRC model was unable to simulate the main effect of AoA in word reading or the AoA × Consistency interaction (Monaghan & Ellis, 2002).

The connectionist single-route view of word recognition was believed to be unable to account for AoA effects because of catastrophic interference (Johnston & Barry, 2006)—that is to say, the observation that newly learned items overwrite earlier learned ones when one set of items entirely replaces another during the training of the network, which is the opposite of an AoA effect. In an influential paper, Ellis and Lambon Ralph (2000) showed that connectionist models were able to exhibit AoA effects when learning of input–output mappings was cumulative and interleaved. Lambon Ralph and Eshan (2006) found that these effects were larger when the mappings were arbitrary (such as those between objects and their names) than when they were systematic (such as those between orthographic and phonological units). Finally, it is important to stress that certain recent theoretical accounts of word recognition do not even discuss age-limited learning effects (e.g., Grainger & Ziegler, 2011; Perry, Ziegler, & Zorzi, 2007). This is particularly clear in the recent connectionist dual-process model (CDP+) of Perry et al. (2007). Although the paper describing this model contains extensive discussions about how the model accounts for important findings in the literature such as consistency or neighbourhood effects, the issue of AoA effects in word reading is not even mentioned.

Our findings strongly suggest that age-limited learning effects are not as widespread in visual word recognition as generally thought when these are properly indexed with frequency trajectory and not with classical subjective AoA measures. As we have argued in the introduction, rated AoA measures are problematic for the study of age-limited learning effects since they are behavioural measures that are in part determined by the trajectory of the words, as clearly evidenced in Experiments 1A and 1B. As a result, the recommendation made by proponents of AoA that (adult) word-recognition models should be modified to account for the influence of AoA is certainly too strong and premature. But do our findings mean that frequency trajectory *never* influences word identification? To reiterate our position, we assume that the observation of frequency trajectory crucially depends on the task used, and, more specifically, on the type of links that are mobilized to perform the task at hand (Bonin et al., 2004; Johnston & Barry, 2006). It should be remembered that in the Bonin et al. (2004) study, we found strong effects of frequency trajectory on both written and spoken naming latencies when using the same set of items but not in word reading or spelling-to-dictation latencies. Also, this factor has a reliable influence on the learning of artificial patterns (Stewart & Ellis, 2008). Finally, and again in laboratory settings, Izura et al. (2010) found genuine *order of acquisition* effects (independent of other factors such as cumulative frequency or frequency trajectory) on the learning of a foreign vocabulary. Order of acquisition effects were observed in picture naming, lexical decision, and semantic categorization and were found to persist for several weeks after the end of training. Thus, order of acquisition stricto sensu also plays a role in the formation, and then the retrieval, of lexical representations (see also Tamminen & Gaskell, 2008).

In word reading, Havelka and Tomita (2006) also observed that age-limited learning effects (with the use of subjective AoA norms) on word reading latencies were stronger with kanji words than with kana words (indeed the effect for the latter was reliable only in the by-participants analysis). The kanji orthographic system is logographic whereas kana is syllabic. Thus, these findings indicate that in word reading, the size of the AoA effect is modified by the degree of overlap between orthography and phonology. Because the mappings between orthography and phonology in kanji are arbitrary and unpredictable, early-acquired words will not help the assimilation of later acquired ones when learning to read kanji words. Indeed, as predicted by Zevin and Seidenberg (2002), strong age-limited learning effects are predicted when the input–output mappings are arbitrary, as is the case in kanji script. In contrast, in the kana script, the relationships between orthographic and phonological units are more systematic, and, as a result, the AoA effect is smaller. Moreover, in English orthography, Monaghan and Ellis (2002) found effects of AoA on irregular/exception words but not on regular/consistent words. For regular/consistent words, early-acquired items can assist in the assimilation of later acquired ones, whereas in the case of irregular/inconsistent items, the learning patterns generated by early-acquired items will not be helpful for the assimilation of late-acquired items. Thus, we cannot exclude the possibility that an age-limited learning effect might emerge with the use of irregular/inconsistent words in word naming in French. Of direct relevance to this issue is the fact that we did not find a reliable frequency trajectory effect in spelling-to-dictation latencies when adults performed a spelling-to-dictation task using the same words as those used in Experiments 1–6. This finding is especially interesting since the words used in Experiments 1–6 are more inconsistent in the phoneme-to-grapheme (PG) direction than in the grapheme-to-phoneme (GP) direction and in particular in the case of the sublexical units located at the end of the words (see Table 1). Thus, if the lack of an effect of frequency trajectory in the word recognition experiments was solely due to the fact that our words were too regular/consistent in the GP direction, there should have been some sign of a reliable frequency trajectory effect in spelling-to-dictation latencies since these words were more irregular/inconsistent in the PG direction. It should be noted that the observation that frequency trajectory has no reliable influence in spelling to dictation is clearly consistent with Zevin and Seidenberg's (2002) view. Since spelling to dictation, like word reading, involves quasiregular mapping relationships in alphabetic languages, what is learned about the spelling of one word can carry over to other words. Therefore, in the same way as in word reading, there is no need to memorize individual patterns when learning to spell most familiar words, whereas this is necessary for highly irregular words. Our study makes it clear that, contrary to Raman's (2006) claim, these effects are not universal in word recognition and may well be restricted to only certain categories of words, perhaps consisting of highly irregular words since it has been demonstrated at both empirical and computational levels that inconsistent or exceptional words result in stronger age-limited learning effects than consistent words (Monaghan & Ellis, 2002, 2010). However, it is important to stress that rated AoA measures and not frequency trajectory measures were used in these studies.[7] At a general level, we assume that the use of frequency trajectory will allow us to gain a better understanding of age-limited learning effects in lexical processing.

Finally, frequency trajectory was found to have a reliable influence only in the AoA rating tasks. This finding has already been observed in multiple regression analyses (Bonin et al., 2004; Zevin & Seidenberg, 2002). It suggests that adults possess knowledge about how words vary in frequency throughout life and that this information is used to perform the task. However, adult AoA ratings are also performed by relying on different types of knowledge as suggested by the observation that imageability, cumulative frequency, and conceptual familiarity are reliable determinants of the AoA ratings (Bonin et al., 2004). In our view, the issue concerning the types of information that are used, and their weight when performing AoA ratings, requires further studies. Even if AoA ratings have been consistently found to be strongly correlated with objective AoA norms (Morrison et al., 1997; Pind et al., 2000), this issue of whether adults' AoA ratings are truly based on the chronology of acquisition of the words remains unsettled. In effect, we agree with Barbarotto, Laiacona, and Capitani (2005) that these correlations, although high, do not demonstrate that the two variables are equivalent. Interestingly, the findings from the Barboretto et al. (2005) study suggest that adults AoA ratings are strongly influenced by the frequency and the familiarity of the words to be rated and that the "real" AoA remains in the background.

To conclude, our study using both a small-scale factorial and a multiple regression approach makes a valuable contribution in showing that age-limited learning effects do not surface easily in visual word recognition in French when frequency trajectory is used to investigate these effects. In contrast, cumulative frequency has a strong and reliable influence in word reading, lexical decision, and progressive demasking, confirming previous findings in the

[7] Indeed, using the FLP corpus (Ferrand et al., 2010), we ran a regression analysis restricted to very inconsistent words. Frequency trajectory was reliable but acted in the opposite direction to that predicted. There is no straightforward way to account for this reversed frequency trajectory effect, and it may possibly be due to strategic factors involved in lexical decision when participants are confronted with very inconsistent items. However, it should be remembered that while the prediction of age-limited effects on very inconsistent items is undoubtedly critical for word naming, this is not the case for lexical decision because GP effects are stronger in word naming than in lexical decision.

word recognition literature and reinforcing the importance of taking this factor into account when building models of word recognition.

Original manuscript received 6 April 2012
Accepted revision received 25 June 2012
First published online 3 October 2012

REFERENCES

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *55,* 290–313.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database [CD-ROM]. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2938 monosyllabic words. *Memory & Cognition*, *29,* 639–647.

Barbarotto, R., Laiacona, M., & Capitani, E. (2005). Objective versus estimated age of word acquisition: A study of 202 Italian children. *Behavior Research Methods*, *37,* 644–650.

Bonin, P., Barry, C., Méot, A., & Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory and Language*, *50,* 456–476.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2001). Age-of-acquisition and word frequency in the lexical decision task: Further evidence from the French language. *Current Psychology of Cognition*, *20,* 401–443.

Bonin, P., Méot, A., Aubert, L., Malardier, N., Niedenthal, P., & Capelle-Toczek, M.-C. (2003). Normes de concrétude de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots [Norms of concreteness, imageability, subjective frequency and emotional valence for 866 words]. *L'Année Psychologique [Imageability: Norms and relationships with other psycholinguistic variables]*, *103,* 655–694.

Bonin, P., Méot, A., Ferrand, L., & Roux, S. (2011). L'imageabilité: Normes et relations avec d'autres variables psycholinguistiques [Imageablity: norms and relationships with other psycholingwistic variables]. *L'Année Psychologique*, *111,* 327–357.

Bonin, P., Méot, A., Mermillod, M., Ferrand, L., & Barry, C. (2009). The effects of age of acquisition and frequency trajectory on object naming: Comments on Pérez (2007). *Quarterly Journal of Experimental Psychology*, *2,* 1132–1140.

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology*, *64,* 545–559.

Brysbaert, M., Lange, M., & van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, *12,* 65–85.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104,* 215–226.

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

Butler, B., & Hains, S. (1979). Individual differences in word recognition. *Memory and Cognition*, *7,* 68–76.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *23,* 857–871.

Carroll, J. B., Davies, P., & Richman, B. (1971). J.B. Carroll, P. Davies, & B. Richman (Eds.), *The American heritage word-frequency book*. Boston, MA: Houghton Mifflin.

Caza, N., & Moscovitch, M. (2005). Effects of cumulative frequency, but not of frequency trajectory, in lexical decision times of older adults and patients with Alzheimer's disease. *Journal of Memory and Language*, *53,* 456–471.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods Instruments & Computers*, *25,* 257–271.

Colombo, L., & Burani, C. (2002). The influence of age of acquisition, root frequency, and context availability in processing nouns and verbs. *Brain and Language*, *81,* 398–411.

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

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology*, *79,* 1–12.

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

Cuetos, F., & Barbon, A. (2006). Word naming in Spanish. *European Journal of Cognitive Psychology*, *18,* 415–436.

Damian, M. (2010). Does variability in human performance outweigh imprecision in response devices such as computer keyboards? *Behavior Research Methods*, *42,* 205–211.

Dent, K., Johnston, R. A., & Humphreys, G. W. (2008). Age of acquisition and word frequency effects in picture naming: A dual-task investigation. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *34,* 282–301.

Dufau, S., Stevens, M., & Grainger, J. (2008). Windows executable software for the progressive demasking task (PDM). *Behavior Research Methods*, *40,* 33–37.

Dunabeitia, J. A., Avilés, A., & Carreiras, M. (2008). NoA's ark: Influence of the number of associates in visual word recognition. *Psychonomic Bulletin & Review*, *15,* 1072–1077.

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

Evans, G. A. L., Lambon Ralph, M. A., & Woollams, A. M. (2012). What's in a word? A parametric study of semantic influences on visual word recognition. *Psychonomic Bulletin & Review*, *19,* 325–331.

Ferrand, F., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., et al. (2008). Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods*, *40,* 1049–1054.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42,* 488–496.

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

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 & Cognition*, *24,* 267–283.

Gerhand, S., & Barry, C. (1999a). Age-of-acquisition and frequency effects in speeded word naming. *Cognition*, *73,* B27–B36.

Gerhand, S., & Barry, C. (1999b). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition*, *27,* 592–602.

Ghyselinck, M., De Moor, W., & Brysbaert, M. (2000). Age-of-acquisition ratings for 2,816 Dutch four- and five-letter nouns. *Psychologica Belgica*, *40*, 77–98.

Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychologica*, *115,* 43–67.

Gilhooly, K. J., & Logie, R. H. (1981). Word age-of-acquisition, reading latencies and auditory recognition. *Current Psychological Research*, *1,* 251–262.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103,* 518–565.

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2,* 1–13.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106,* 491–528.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111,* 662–720.

Havelka, J., & Tomita, I. (2006). Age of acquisition in naming Japanese words. *Visual Cognition*, *13,* 981–991.

Izura, C., Pérez, M., Agallou, E., Wright, V. C., Marín, J., Stadthagen-Gonzalez, H., et al. (2010). Age/order of acquisition effects and cumulative learning of foreign words: A word training study. *Journal of Memory and Language, 64*, 32–58.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art.

Journal of Experimental Psychology: Human Perception and Performance, *20,* 1311–1334.

Johnston, R., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, *13,* 789–845.

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

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44,* 287–304.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

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.

Lété, B., Peereman, R., & Fayol, M. (2008). Phoneme-to-grapheme consistency and word-frequency effects on spelling among first- to fifth-grade French children: A regression-based study. *Journal of Memory and Language*, *58,* 952–977.

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, and Computers*, *36,* 156–166.

Liu, Y. Y., Hao, M. L., Hua, S., Tan, L.-H., & Weekes, B. S. (2008). Age of acquisition effects on oral reading in Chinese. *Psychonomic Bulletin and Review*, *15,* 344–350.

Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling–sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *28,* 183–206.

Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, *63,* 506–525.

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*, *50,* 528–559.

Morrison, C. M., & Ellis, A. W. (1995). The roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *21,* 116–133.

Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology*, *91,* 167–180.

Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology*, *14,* 435–459.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, and Computers*, *36,* 516–524.

Peereman, R., & Content, A. (1999). Lexop. A lexical database with orthography–phonology statistics for French monosyllabic words. *Behavior Research Methods, Instruments, and Computers*, *31,* 376–379.

Peereman, R., Lété, B., & Sprenger-Charolles, L. (2007). Manulex-infra: Distributional characteristics of grapheme–phoneme mappings, infra-lexical and lexical units in child-directed written material. *Behavior Research Methods*, *39,* 593–603.

Perry, C., Ziegler, J., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114,* 273–315.

Pind, J., Jonsdottir, H., Tryggvadottir, H. B., & Jonsson, F. (2000). Icelandic norms for the Snodgrass and Vanderwart (1980) pictures: Name and image agreement, familiarity, and age of acquisition. *Scandinavian Journal of Psychology*, *41,* 41–88.

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

Raman, I. (2006). On the age-of-acquisition effects in word naming and orthographic transparency: Mapping specific or universal? *Visual Cognition*, *13,* 1044–1053.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, *96,* 523–568.

Seidenberg, M. S., & Waters, G. S. (1989). Word recognition and naming: A mega study. *Bulletin of the Psychonomic Society*, *27,* 489.

Shibahara, N., Zorzi, M., Hill, M. P., Wydell, T., & Butterworth, B. (2003). Semantic effects in word naming: Evidence from English and Japanese kanji. *Quarterly Journal of Experimental Psychology*, *56,* 263–286.

Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (2009). Error, error everywhere: A look at megastudies of word reading. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 2009 Meeting of the Cognitive Science Society* (pp. 1036–1041). Austin, TX: Cognitive Science Society.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8,* 411–416.

Stadthagen-Gonzalez, H., Bowers, J. S., & Damian, M. F. (2004). Age-of-acquisition effects in visual word recognition: Evidence from expert vocabularies. *Cognition*, *93,* B11–B26.

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

Stewart, N., & Ellis, A. W. (2008). Order of acquisition in learning perceptual categories: A laboratory analogue of the age of acquisition effect? *Psychonomic Bulletin and Review*, *15*, 70–74.

Strain, E., Patterson, K., & Seidenberg, M. S. (2002). Theories of word naming interact with spelling–sound consistency. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *28,* 207–214.

Taft, M., & van Graan, F. (1998). Lack of phonological mediation in a semantic judgement task. *Journal of Memory and Language*, *38,* 203–224.

Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *Quarterly Journal of Experimental Psychology*, *61,* 361–371.

Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory & Cognition*, *26,* 1282–1291.

Whaley, C. P. (1978). Word–nonword classification times. *Journal of Verbal Learning and Verbal Behavior*, *17,* 143–154.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

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.

Zevin, J. D., & Seidenberg, M. S. (2004). Age of acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory & Cognition*, *32,* 31–38.

APPENDIX A

Words used in Experiments 1–6

High-to-low	*Low-to-high*	*High-to-high*	*Low-to-low*
Accrocher (to hang)	Article (article)	Apprendre (to learn)	Abaisser (to reduce)
Assiette (plate)	Auteur (author)	Classe (classroom)	Astronome (astronomer)
Banane (banana)	Aventure (adventure)	Content (happy)	Baudet (donkey)
Bouquet (bouquet)	Chasse (hunting)	Dessiner (to draw)	Bouclier (shield)
Camion (truck)	Contraire (opposite)	Éléphant (elephant)	Compas (compass)
Chanson (song)	Français (french)	Exercice (exercice)	Coupable (guilty)
Chaussure (shoe)	Indiquer (to point to)	Garçon (boy)	Coutume (custom)
Crayon (pencil)	Lecture (reading)	Journal (newspaper)	Frontière (frontier)
Crocodile (crocodile)	Manière (manner)	Papier (paper)	Inaugurer (to unveil)
Déguiser (disguise)	Mauvais (bad)	Personne (nobody)	Maintien (carriage)
Fantôme (ghost)	Présenter (to introduce)	Retourner (to turn over)	Maudire (to damn)
Farine (flour)	Raison (reason)	Retrouver (to find again)	Nomade (nomad)
Fromage (cheese)	Regard (a look)	Soleil (sun)	Obélisque (obelisk)
Mercredi (Wednesday)	Singulier (singular)	Travail (job)	Pichet (flagon)
Papillon (butterfly)	Souligner (to underline)	Traverser (to cross)	Poivron (pepper)

(*Continued overleaf*)

APPENDIX B

Instructions used to collect age of acquisition (AoA) ratings in Experiment 1A

We acquire words throughout our lives. Some words are acquired at a very early stage, some are acquired later, and others fall in between. The purpose of this study is to determine the approximate age at which words have been acquired (in their written or spoken form). By “learning a word” we mean when you understood that word when somebody used it in front of you, even if you did not use, read, or write it at that time. Your task is to type, in years, the age at which you learned each of the words presented on the screen. An approximate age is good enough for this rating. For instance, if you think you learned the word “dragon” at the age of 3 years, then you type “3” in the text box below this word. If you think you learned the word “tax” at the age of 16, then type “16”. If you do not know the meaning of a word, type “NO” in the text box below the word. When making your ratings, try to be as accurate as possible, but do not spend too much time on any one word.