Bilingualism: Language and Cognition

cambridge.org/bil

An "Author Fluency Task": Semantic fluency as predictor of L2 vocabulary knowledge

Sean P. McCarron[1,2], Victoria A. Murphy[3] and Kate Nation[1]

[1]Department of Experimental Psychology, University of Oxford, Oxford, UK; [2]Department of Psychology, University of Waterloo, Waterloo, ON, Canada and [3]Department of Education, University of Oxford, Oxford, UK

Research Article

Cite this article: McCarron, S.P., Murphy, V.A. and Nation, K. (2026). An "Author Fluency Task": Semantic fluency as predictor of L2 vocabulary knowledge. *Bilingualism: Language and Cognition* **29**, 809–822. https://doi.org/10.1017/S136672892510045X

Received: 14 August 2024
Revised: 21 June 2025
Accepted: 30 July 2025
First published online: 27 August 2025

Keywords:
print exposure; formulaic language; assessment; Author Recognition Test; semantic fluency

Corresponding author:
Sean P. McCarron;
Email: sean.mccarron@psy.ox.ac.uk

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

Reading experience provides critical input for language learning. This is typically quantified via estimates of print exposure, such as the Author Recognition Test (ART), although it may be unreliable in L2. This study introduces the Author Fluency Task (AFT) as an alternative measure, comparing with ART for assessing knowledge of English discourse connectives and collocations among 60 bilingual French/English speakers, and a comparison sample of 60 L1 English speakers. Participants completed AFT, ART, and LexTALE in both languages. Analysis of L2 measures showed AFT more accurately predicted L2 vocabulary knowledge than ART, even when controlling for proficiency (LexTALE). Conversely, ART was more effective for L1 speakers, showing a striking dissociation between the measures across language groups. Additionally, data showed limited contributions from L1 proficiency and print exposure on L2 vocabulary. These findings recommend AFT as a valuable tool for quantifying the role of L2 print exposure for language learning.

Paper Highlights

- Introduces the "Author Fluency Task" (AFT) to assess print exposure
- AFT is more correlated with L2 vocabulary than the Author Recognition Test (ART)
- Print exposure measures vary in utility across L1/L2 groups
- L2 performance only partially explained by verbal fluency skill
- Recommends AFT for measuring L2 print exposure

1. Background

1.1. Assessment of L2 print exposure

Print exposure predicts individual differences in component skills of reading in L1 speakers (Mol & Bus, 2011; Moore & Gordon, 2015), yet the amount an individual reads for pleasure is almost necessarily a function of the pleasure derived from reading. This leads to the "Matthew effect": a growing gap between the rich and poor in reading skills (Cunningham & Stanovich, 1998; Mol & Bus, 2011; Stanovich, 1986). An adept reader may find the practice both enjoyable and rewarding, reaping the benefits of increased exposure. Less proficient readers, however, may find reading more frustrating than gratifying, and avoid picking up books in their free time – consequently, their skills may stagnate, making reading even less enjoyable.

Reading difficulties (and a concomitant skill gap) may be further compounded in L2, as learners often struggle with more obscure vocabulary than they encounter in daily speech. Surmounting this difficulty is crucial for second language acquisition (SLA), however, because a significant portion of L2 vocabulary is acquired incidentally through reading (Restrepo-Ramos, 2015). Naturally, learners need significant language exposure to reach their full potential in L2, meaning researchers require precise psychometric instruments to quantify L2 speakers' exposure to print.

The Author Recognition Test (ART; Stanovich & West, 1989) is the standard test of print exposure, in which participants are asked to select authors from a checklist of names. Foils (i.e., fake names) are usually included to discourage guessing. As a proxy measure for reading experience, ART is well-validated in L1 populations as a predictor of individual differences in a variety of reading skills (McCarron, 2026), including vocabulary (Dąbrowska, 2018), word recognition (Chateau & Jared, 2000), spelling (Stanovich & West, 1989), reading frequency (Acheson et al., 2008; Moore & Gordon, 2015), sentence processing (Acheson et al., 2008), oral language skills (Acheson et al., 2008; Mol & Bus, 2011), reading comprehension, and academic achievement (Mol & Bus, 2011).

Despite this, ART faces concerns about its reliability and validity in L2 populations (McCarron & Kuperman, 2021; Vermeiren & Brysbaert, 2023). Essentially, L2 speakers generally know very few authors on ART – the question is whether they do not read enough in L2, or if they

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

are simply not reading the kinds of authors on ART. This distinction is not trivial, as there is substantially more variation in the quantity and kind of L2 compared to L1 exposure (Flege, 2008, 2019; Gullifer & Titone, 2020), meaning the selection of authors who are representative of L1 reading experience may not index the same latent variable in L2. If so, it may be that a valid and reliable measure of reading experience in L2 must acknowledge and exploit this variation (in the parlance of computer programmers) as a *"feature, not a bug"*.

Alternatively, it may be that extensive L2 reading for pleasure does not materially contribute to second language proficiency. Given the vital role of reading for L1 proficiency, this may seem unlikely. However, due to the preponderance of language input online, one might wonder whether Internet text exposure is a better predictor of language skill than print exposure for most speakers. In fact, recent L1 research suggests the opposite – when comparing effects of print exposure, years of postsecondary education, reading attitudes and website exposure in a study of fluent Norwegian speakers, Strømsø (2024) found only print exposure predicted reading comprehension scores. Moreover, a high degree of online exposure negated the positive effects of print exposure for participants with a high degree of both. Although not yet replicated in L2, this finding demonstrates that book reading is a uniquely beneficial kind of exposure.

There are reasons to expect book reading to be a better source of language input compared to the Internet. Online texts found on social media and discussion forums tend to have more in common with spoken rather than written language, being more conversational and informal (Johns et al., 2020; Snow, 2010), and there are critical distinctions between the two modalities. Compared to speech, written language features greater lexical density and diversity (Berman & Nir, 2010; Roland et al., 2007), as well as longer sentences, and correspondingly, more complex syntax (Biber, 1988), including more passive constructions (Dąbrowska & Street, 2006) and relative clauses (Roland et al., 2007). This follows from the disembodied nature of text, which must construct context and meaning *ex nihilo*, whereas spoken language can create meaning through reciprocity and shared context (Clark, 2020; Snow, 2010). Corpus studies of children's books have also revealed they contain more relative clauses, more complex syntax and greater lexical density and diversity compared to both child-directed speech (Dawson et al., 2021; Hsiao et al., 2022; Nation et al., 2022) and adult television transcripts (Cunningham & Stanovich, 1998). Books are thus not only qualitatively different from other sources of input – they also distinguish themselves in the very early stages of language development.

1.2. A semantic fluency measure of L2 print exposure

The primary advantage of proxy measures like ART is that they avoid potential social desirability biases associated with self-report measures such as reading surveys (West et al., 1993). Yet a standard ART does not indicate whether recognising an author's name reflects personal reading *experience* or general reading *exposure* (*"primary versus secondary print knowledge"*; Martin-Chang & Gould, 2008). Additionally, some research suggests ART reflects general cultural knowledge rather than reading experience specifically (Moore & Gordon, 2015; Vermeiren et al., 2022). Ideally, researchers would use a test that measures the latter rather than the former, to the extent that these concepts can be extricated.

The fundamental assumption of ART is that knowledge of author names offers a reliable proxy for print exposure. But because L2 speakers have different kinds of cultural exposure to their target language, and consequently may encounter different authors when reading in L2, a potential alternative might be to simply ask second language speakers to name L2 authors who come to mind. Such measures of semantic fluency (SF) involve listing as many items as possible from a given category in a set time, with one point for each unique and valid item. SF tasks are often used in estimating the advancement of neurodegenerative diseases such as Alzheimer's and dementia (Macoir et al., 2006; Troyer et al., 1997, 1998). However, SF tasks have also been used in L2 studies, where bilinguals typically generate fewer category items and proper names than monolinguals (e.g., Gollan et al., 2002). This undoubtedly relates partially to the speed of lexical access, but also to fewer encounters with L2 exemplars; naturally, the two concepts are interrelated. Yet this group difference is not deterministic, as evidence shows that proficient bilingual adults can perform equivalently to monolinguals on SF tasks (Friesen et al., 2015). An "Author Fluency Task" (AFT) would thus rely on the assumption that individuals with greater L2 print exposure can also access more author names extemporaneously, consistent with the *"principle of likely need"* (Jones et al., 2017).

What is the benefit of developing a new proxy measure of print exposure, as opposed to simply creating an ART for L2 populations? Although an L2 ART may be more reliable than one that has been validated with L1 speakers, it would nevertheless require calibration for each L2 population evaluated, and the scores would not be directly comparable between groups. Another advantage of AFT over ART is that it might level the playing field for L2 speakers, providing all participants equal time to demonstrate their print knowledge, whereas L2 speakers may not have encountered as many authors on ART. Author fluency and recognition also rely on very different skills, with AFT requiring an extensive search of explicit memory surrounding reading experience, and ART arguably a less demanding task, given that it requires participants only to identify familiar author names as opposed to retrieving them independently. Comparatively, AFT might be more difficult, but tasks that target productive rather than receptive use of language are more useful for advanced learners of English (Webb & Kagimoto, 2009), our primary population of interest. Granted, just like recognising an author on ART, naming an author does not necessarily reflect personal reading experience. Nevertheless, there is reason to expect that author names that are recalled could be more indicative of primary print exposure than those that are recognised. Recognition tasks like ART have been argued to reflect *"marginal knowledge"*, or information that is stored in memory but inaccessible unless presented (Berger et al., 1999; Cantor et al., 2015), suggesting it is not deeply encoded. Semantic fluency, in contrast, primarily indexes the semantic organisation of memory (Lehtinen et al., 2023), and by necessity, this requires a substantial body of well-integrated information. To evaluate AFT as a measure of L2 print exposure – and to compare with ART – it would need to be validated using outcome measures for L2 vocabulary that are often acquired through extensive reading experience. For this reason, we decided to use measures of formulaic language.

1.3. Formulaic (and functional) language in L2

Some vocabulary is especially difficult for L2 speakers to acquire and use naturally. Discourse connectives, which link ideas across sentence clauses, are one such example. They are often associated with written language (in particular, academic writing; Biber, 2006) and may be composed of either single ("consequently", "nevertheless") or multiple words ("as long as", "in addition", "on the other hand", etc.). In the latter case, such connectives are clear

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

examples of formulaic language, typically defined as expressions comprised of at least two words that are processed as a single unit (Wray, 2002, 2006). We contend, however, that single-word connectives may also be considered formulaic in the sense that they encode a set of "operating instructions" for interpreting the coherence relations linking separate clauses (Andersson, 2016; Li et al., 2017). Many connectives also blur the line between single and multi-word items, as they often begin life as lexical bundles but become lexicalised as single words over time due to their frequent co-occurrence and entrenchment (e.g., "indeed", "furthermore", "moreover", "nevertheless", etc.). We argue that this highlights how the distinction between single and multi-word processing is largely arbitrary, as put forth by "single-system" models of language (e.g., Arnon & Christiansen, 2017; Bybee, 2007). Therefore, we consider connectives to be formulaic language under a broader, usage-based perspective that emphasises their pragmatic function as linguistic "prefabs" (Bybee, 2006). Connectives, like other constructions, are "partially schematic—that is, they have positions that can be filled by a variety of words or phrases" (Bybee, 2010, p. 25). For connectives, these "open positions" are clauses that must fulfil the conditions of their coherence relations. For example, "whereas" requires a contrastive clause, e.g., "*whereas x* [statement], *y* [contrast]", and "although" requires a concession relation, e.g., "*although x* [statement], *y* [concession]". Similar to other kinds of formulaic language, connectives require speakers to "chunk" together words or phrases into meaningful sequences, and this is often developed through substantial implicit learning (Ellis, 1996).

Because they lack a strict lexical definition (Van Silfhout et al., 2015; Zufferey et al., 2015), connectives can be difficult to acquire through explicit teaching. One explicit approach for teaching L2 connectives is simply to provide an approximate L1 equivalent; yet the closest translation in L1 may not necessarily encode the same relations in L2 in all cases (Zufferey & Gygax, 2017). This can be problematic for second language speakers who often filter L2 through the lens of L1 – particularly during early stages of acquisition – relying on an inconsistent equivalence between L1 and L2 vocabulary items (Ringbom, 2016). Consequently, connectives pose a serious challenge, with even very advanced L2 speakers often struggling to understand how and when to use them (Wetzel et al., 2020).

Similarly, word collocations (e.g., *"weak tea"* preferred over *"feeble tea"*) are another obstacle for L2 learners. Although these can be learned through explicit instruction, they are more challenging than learning single words (Peters, 2014, 2016), and their virtually endless number means they are likely an inefficient use of targeted language instruction, which largely focuses on individual words (Schmitt, 2010). However, collocations can be acquired incidentally through statistical learning from input, both in L1 and L2 (Pellicer-Sánchez, 2017; Sonbul & Schmitt, 2013; Webb et al., 2013), and the more language input one receives, the more these associations are formed. Accordingly, L2 speakers process L2 collocations more slowly than L1 speakers (Siyanova & Schmitt, 2008) and use fewer collocations in L2, which also tend to be congruent with how words are paired in their L1 (Granger, 1998).

Whereas the importance of selecting the correct connective may be clear, the significance of collocation knowledge may be less evident. After all, what difference is there between *"raise prices"* and *"lift prices"*? If "raise" and "lift" are essentially synonymous, surely either one will serve the same purpose. But all word pairings are not created equal, and set phrases are subject to certain preferential selection constraints. Indeed, although speakers of a language may correctly infer the meaning of a novel expression, formulaic language is processed more quickly and accurately (Ellis et al., 2008; Hallin & Van Lancker Sidtis, 2017). Therefore, researchers have posited that this preference for formulaic language ostensibly functions to ease processing burdens between communicators (Wray, 2002).

What constitutes as "formulaic", however, is largely (though not solely) a matter of frequency of occurrence (Siyanova-Chanturia et al., 2011), and L2 learners are even more sensitive to frequency for formulaic language compared to natives (Ellis et al., 2008). Corpus studies reveal that written and spoken language are also distinct in their use of formulaic language; certain collocations are more common in writing, whereas others appear more frequently in speech (Gablasova et al., 2017; Shin, 2007), connective frequencies vary by modality and register (Andersson & Sundberg, 2021) and connectives use is more varied in writing (Tskhovrebova et al., 2022). Given that L2 learners have comparatively less exposure, and are more likely to interpret formulaic language in L2 serially (i.e., word-by-word) rather than processing into meaningful "chunks" as in L1 (Conklin & Schmitt, 2012), connectives and collocations present a significant hurdle. Accordingly, L2 writing and speech is often characterised by an overreliance on certain connectives (Wetzel et al., 2020), and features less formulaic language in general (Granger, 1998; Pérez-Llantada, 2014). Formulaic expressions, however, are often less about the meaning of individual words than understanding how words relate to each other. As J.R. Firth put it, echoing Wittgenstein, *"you shall know a word by the company it keeps"* (1957, p. 11).

Although we classify both connectives and collocations as formulaic language, we reiterate that there are critical distinctions between the two which are important for interpreting our results. If collocations are a more canonical example of formulaic language, connectives are perhaps more "functional" than formulaic. This is due to the kinds of meanings they convey. Connectives encode procedural meaning (Blakemore, 2002) and guide inferences between clauses, whereas collocations encode conceptual meaning, reflecting learned associations between co-occurring words. Connectives are further complicated by their polyfunctionality, as a particular connective may perform a different role depending on semantic or pragmatic context (as in the French *"en effet"*, which can convey causal or confirmational coherence relations; Zufferey & Gygax, 2017). Further complicating matters, temporal prepositions such as "since" or "while" – already challenging for many L2 speakers – can double as discourse connectives; compare, for example, *"since he had surgery, he hasn't come hiking"* and *"since he had surgery, he can't come hiking"*. This polyfunctionality requires speakers to distinguish subtle gradations in relational meaning that are not required for collocations, and which span across separate clauses. However, both connectives and collocations require substantial experience to master, and print exposure likely helps speakers attune to the statistical regularities that inform their use. For these reasons, we use both in this study as a validation of AFT.

1.4. Contributions of L1- and L2-specific skills for L2 learning

Although the importance of L1 input is well-accepted, the degree of influence of L1- versus L2-specific skills in SLA remains a matter of debate. Language transfer theories (Baker et al., 2011; Cummins, 1979; Sparks et al., 2012) posit that greater L1 proficiency affords a proportionate degree of linguistic knowledge in L2, and while there is considerable evidence for this (Sparks, 2023), some have argued it is limited to more general language skills such as phonology and pragmatics rather than syntax and vocabulary (Verhoeven, 1994). For our present discussion, the role of L1 print exposure is

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

particularly relevant, and there is evidence of its influence on L2 reading skills, including decoding and comprehension (Sparks et al., 2012). One study showed that while L1 German print exposure (as measured by a German ART) predicted L2 French connectives knowledge, a French ART did not (Wetzel et al., 2020). Although this may be attributable to language interdependence, we contend that the findings are expected for this population of adolescent beginner L2 speakers, who generally have little L2 exposure – as the authors point out, their participants knew very few of the second-language authors on ART. Since the effect of print exposure is cumulative over a lifetime, a more interesting case might be to compare L1 and L2 print exposure measures in an older, more proficient L2 population. This is what we endeavoured to do in the present study.

1.5. Present study

This study received ethics approval [reference R77364/RE002] and was pre-registered (https://osf.io/nsduz/). We tested whether L1 French/L2 English print exposure (assessed by AFT and ART in both languages) is associated with individual differences in knowledge of English connectives and collocations, even when accounting for a standard proficiency measure in both languages. Our research questions were intended to assess the utility of AFT as a novel measure of print exposure:

1) Does an L2 AFT outperform ART as a predictor of L2 vocabulary knowledge? Does either measure explain additional variance not accounted for by proficiency?
2) Do AFT/ART perform differently by vocabulary measure (collocations versus connectives)?
3) Does L1 or L2 print exposure better predict performance on L2 vocabulary tasks?

For our L2 English cohort, we hypothesised that

1) L1/L2 LexTALE scores would both positively predict connectives and collocations scores.
2) L2 (but not L1) AFT scores would positively predict connectives scores when controlling for LexTALE.
3) L1/L2 ART (but not AFT) scores would positively predict collocations scores when controlling for LexTALE.

For comparison, we hypothesised the same pattern for our L1 English cohort; i.e., that the English ART would predict collocation scores, and AFT would predict connectives. Essentially, we predicted that L2 print exposure, measured by AFT, would reliably predict connectives scores when controlling for LexTALE, but only ART would predict additional variance for collocations scores. Our rationale was that recognising authors might recruit similar skills as those required for recognising collocations. In contrast, we posited that the L2 English AFT, reflecting explicit memory of L2 reading experience, would be associated with English connectives, which require careful consideration to evaluate their functions. The L1 French AFT, however, which reflects L1 reading, was not anticipated to predict L2 vocabulary. In this way, we aimed to determine how L1 and L2 print exposure variously contribute to L2 language skills.

2. Methods

2.1. Participants

Prior to data collection, power analysis was carried out using G*Power (Faul et al., 2007). For 0.8 power to detect a small effect size of .15 at a .05 alpha error probability, we obtained a recommended sample size of $n = 55$. Sixty L1 French/L2 English participants (M_{age} = 31.13, 32 women) were recruited through Prolific (2024) to complete a single session on the online experimental research platform Gorilla (Anwyl-Irvine et al., 2020). Participants who provided informed consent and completed the study were reimbursed £6.67 each. Recruitment was limited to participants between 18 and 75 years old, who spoke French natively and currently lived in France, who spoke and read English fluently at an intermediate-to-advanced level and had normal or corrected-to-normal vision. Self-rated L2 proficiency (0–5 Likert scale) was high, with the majority scoring themselves between 3 ("Professional Working Proficiency") and 4 ("Full Professional Proficiency"), see Supplementary Figure S2.

We also recruited 60 L1 English speakers (M_{age} = 39.42, 37 women) through Prolific. Selection criteria mirrored those of the L2 group, but with native English speakers living in the United Kingdom. Below, we primarily restrict our analyses to the L2 cohort, permitting us to compare the relative contributions of L1 and L2 measures. However, we also include models from L1 English speakers to illustrate the differential predictions made by our print exposure measures.

2.1.1. Procedure

Participants completed 1) the demographics questionnaire, followed by 2) the English and French AFT, 3) the English and French LexTALEs and ARTs, 4) the connectives and collocations tasks and 5) a motivation survey. Task order was counterbalanced for levels 2, 3 and 4 due to task similarities. The L1 participant procedure was identical, excluding French tasks.

2.2. Materials

2.2.1. Author fluency task

For the English AFT, participants listed as many author names as possible in three minutes. Instructions asked participants to provide names of authors who had been published in English and who were known primarily for their writing. Participants typed names into a text field. Due to the reportedly difficult nature of the task, names were scored leniently by the first author. Each name was verified using an online database of over 6,000,000 author names (Internet Archive, 2022) and Google to determine possible misspellings, which were corrected. Validated author names were rated 1, non-authors −1 and indeterminate names 0. The coded ratings were then summed for each participant's list of names. For example, a hypothetical participant listing, "J.R.R. Tolkien, Margaret Atwood, Kurt Vonnegut, Conan O'Brien, J. Smith" (3 authors, 1 non-author, 1 indeterminate), would receive a score of 3–1−0 = 2. Selection statistics are shown in Supplementary Tables S7/S8, and group differences in score distributions are visualised in Supplementary Figure S4.

The French AFT was identical in procedure and scoring, but with French instructions. Accordingly, participants were asked to provide names of authors who had been published in French. Selection statistics are shown in Supplementary Table S9.

2.2.2. Author recognition test

The English ART was taken from Vermeiren et al. (2022), featuring 60 author names and 30 foil names. Participants were randomly shown each name serially and were asked to indicate whether each was an author or not with keyboard responses. Correct author selections increased scores by 1 point, incorrect selections decreased

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

scores by 1 and no penalty was incurred for not indicating an existing author. A full list of author names, mean response times and selection statistics is found in Supplementary Tables S10/S11, and is illustrated in Supplementary Figures S6/S7. Group differences in score distributions are visualised in Supplementary Figure S5.

The French ART followed the same procedure and scoring logic, but participants were provided instructions in French. This version was taken from Zufferey and Gygax (2020), and features 40 author names and 40 foil names (Supplementary Table S12).

2.2.3. LexTALE

The English LexTALE (Lemhöfer & Broersma, 2012) is a lexical decision task containing 40 words, 20 non-words and 3 filler words. Participants were shown each item randomly and responded with keyboard presses to indicate whether each was an English word or not. Scores were calculated as the percentage of correct selections for words and non-words out of the total (Supplementary Table S13). Supplementary Figure S8 visualises group score differences.

The French LexTALE (Brysbaert, 2013) followed the same procedure and scoring logic, but contained 56 French words and 28 non-words. Instructions were provided in French, and scores were calculated as the percentage of correct selections out of the total (Supplementary Table S14).

2.2.4. Discourse connectives task

This task was adapted and translated to English from the original version in Wetzel et al. (2020), which was presented in French to L1 German French learners. This is a sentence cloze task that asks participants to complete a coherent sentence by selecting the appropriate connective from six options. For example:

Nadir likes his job, ____ his sister would like to change her career.

1)	therefore	2)	since	3)	**whereas**
4)	furthermore	5)	nevertheless	6)	as long as

Each connective falls into one of six coherence relations denoting the logical relationships specified by each connective, e.g., "whereas" encodes a "contrast" relation. For each sentence, competitors were selected from each of the other relations. High and low frequency connectives were selected using the corpus English Web 2020 ("enTenTen20") in corpus software *SketchEngine* (Kilgarriff et al., 2014). The full list of stimuli (Supplementary Table S1) and corpus frequency statistics (Supplementary Table S2) can be found in the Supplementary Materials.

2.2.5. Collocations task

The *Words That Go Together* task was used to assess knowledge of English collocations (Dąbrowska, 2014). Participants read a list of five word pair phrases and were instructed to select the one that was most familiar or natural. Accuracy scores were calculated as percentages of correct selections. The full list of stimuli is provided in Supplementary Table S3.

2.2.6. Semantic fluency

After reviewing the initial findings, we conducted an additional analysis (pre-registered in an update) using a test of general semantic fluency in English. This followed the same format as AFT, but with three different categories of items: "animals", "grocery items" and "public figures" (i.e., famous people, including celebrities, politicians, etc.). Of the original 60 L2 participants, 48 returned two months later. Participants were given one minute for each category, for a total of three minutes, equivalent to AFT. Unlike AFT, participants were unable to complete the task early, which may have increased the number of items provided. Items were scored by the first author and calculated as the sum of unique and valid items per category. Score distributions by sub-task are visualised in Supplementary Figure S11.

2.2.7. Additional variables

Details on additional variables, including motivation (Supplementary Table S4) and demographics, can be found in the Supplementary Materials. Supplementary Table S6 provides summary statistics for the motivation measure, and Supplementary Figures S1, S2 and S3 visualise demographic information of L2 participants, including age of acquisition, self-reported proficiency and ratings of perceived importance of reading for learning English.

3. Results

Summary statistics and sample sizes per task are presented in Table 1. L1 participant scores exceeded L2, most notably for LexTALE and collocations. Outliers were identified as those falling below Q1−1.5×IQR or above Q3 + 1.5×IQR within their cohort on each task. While not pre-registered, this step was taken due to very low scores for some tasks. Outliers on each task were removed, leading to slightly lower sample sizes in some measures. Correlations for all measures in the L2 group are shown in Table 2, and an analogous table for L1 speakers is provided in Supplementary Table S5. Author name selection statistics are also provided for AFT (Supplementary Tables S7/S8/S9) and ART (Supplementary Tables S10/S11/S12).

Analysis was performed in R (version 2023.12.1, R Core Team, 2024). Generalised linear mixed effects models (GLMER) were constructed using the package lme4 (Bates et al., 2015), *p*-values were extracted using the package lmerTest (Kuznetsova et al., 2017) and model assumptions of overdispersion, normality and outliers were checked using the package DHARMa (Hartig, 2022). To counter problems with multicollinearity, continuous predictors were first standardised before being entered into GLMERs, and we iteratively compared model performance with likelihood ratio tests using the maximal effects structure justified by the design (Barr et al., 2013).

3.1. Connectives

We begin by describing performance by connective type and language group before considering models that demonstrate the relative strengths of each predictor for both language groups. Group differences in score distributions are illustrated in Supplementary Figure S9. Scores for each connective, by coherence relation, frequency and language group are presented in Table 3, and performance by coherence relation and group is illustrated in Figure 1. Higher frequency connectives, unsurprisingly, were responded to more accurately than lower-frequency alternatives. A notable exception was "indeed", where performance was poorer compared to even the lowest frequency connective. This may be because we were specifically interested in its use as a subordinating conjunction, which could not be uniquely captured with our search terms – although "indeed" is very common in the corpus, its use as a connective is substantially lower relative to alternative uses.[1]

[1]Consequently, it may seem reasonable to re-code "indeed" as a low frequency connective rather than a high one. To decide, we hand-coded a random sample of 500 instances of "indeed" on *SketchEngine* and determined it appears

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

Table 1. Summary statistics for each task by cohort. Mann–Whitney U test *p*-values were Bonferroni-corrected for multiple comparisons. Bold *p*-values indicate $p < .05$.

Task	Group	n	Min	Median	Mean	Max	SD	IQR	U	*p*-value
AFT	L1	59	3	13	12.97	29	5.80	7.50	2474.5	**< .001*** **
	L2	60	1	8.50	8.92	22	5.32	9.00		
AFT-FR	L2	60	1	11	11.77	28	6.18	8.25		
ART	L1	60	5	25.50	27.23	53	11.18	14.75	2278	.08
	L2	60	0	20	21.35	46	10.40	15.25		
ART-FR	L2	60	–3	7	9.50	29	7.90	10.50		
Collocations (%)	L1	59	59.00	42.50	77.50	76.06	95.00	11.47	3136	**< .001*** **
	L2	60	60.00	20	47.50	48.38	82.50	17.23		
Connectives (%)	L1	59	61.67	86.67	85.28	98.33	8.68	13.33	2339	**< .05***
	L2	60	20.00	79.17	71.97	100	21.77	35.42		
LexTALE (%)	L1	57	81.67	95.00	95.03	100	4.44	5.00	2834.5	**< .001*** **
	L2	59	60.00	95.00	83.11	100	11.25	20.00		
LexTALE-FR (%)	L2	60	65.48	88.10	87.26	98.81	7.71	11.90		
Semantic Fluency	L2	47	14	34	35.57	60	11.42	16.00		

Table 2. Spearman correlation matrix for all measures, L2 English cohort. Significant correlations are in **bold**; * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

	AFT	ART	AFT-FR	ART-FR	Coll.	Conn.	LT	LT-FR	SF
AFT	1								
ART	**.686*** **	1							
AFT-FR	**.676*** **	**.536*** **	1						
ART-FR	**.439*** **	**.675*** **	**.464*** **	1					
Collocations	**.603*** **	**.486*** **	**.324*** *	**.300*** *	1				
Connectives	**.612*** **	**.400*** **	**.374*** **	.248	**.782*** **	1			
LexTALE	**.605*** **	**.567*** **	**.357*** **	**.363*** **	**.820*** **	**.734*** **	1		
LexTALE-FR	.152	.240	.132	**.388*** **	.097	−.070	.202	1	
Semantic Fluency	**.616*** **	**.467*** **	**.486*** **	−.002	**.510*** **	**.622*** **	**.530*** **	.007	1

Curiously, L2 speakers outperform L1 participants on "indeed", the sole exception of its kind. This could be due to familiarity with the French connective *"en effet"*, which functions similarly to "indeed" (albeit with important differences; Zufferey & Gygax, 2017), yet it is unclear why English natives struggle with this high-frequency connective.

For L2 English speakers, comparisons favoured a linear regression model with only the English AFT as a predictor of connectives scores over one with ART alone, as indicated by a significant Vuong test ($z = 2.54$, $p < .01$) and a lower AIC ($\Delta AIC = -17.76$). Our most comprehensive model ($F(2, 57) = 39.86$, $p < .001$, Adj-$R^2 = .57$) showed effects for both LexTALE ($F(1, 57) = 73.61$, $p < .001$) and AFT ($F(1, 57) = 6.11$, $p < .05$). The English ART did not significantly predict connectives when considering either of the other variables. The contributions of each L2 predictor are illustrated in the standardised partial residuals presented in Figure 2A, and a model with LexTALE and AFT as predictors is provided in Supplementary Table S16.

Performance was also evaluated using L1 French measures. Comparing regression models with predictors of the French AFT and ART alone preferred the AFT model ($\Delta AIC = -5.20$). The best-fitting model ($F(1, 58) = 9.40$, $p < .01$, Adj-R^2: .13) identified the French AFT (AFT-FR) as a significant predictor (Supplementary Table S17). However, this model did not satisfy the assumption of normally distributed residuals (Shapiro–Wilk $p < .01$), and attempts to address this issue through data transformation and robust regression methods were unsuccessful. This model's findings are thus interpreted with caution, but evidently, the explanatory power of L1 print exposure appears modest. For comparison across the French LexTALE, ART and AFT, Supplementary Figure S12 shows standardised partial residual plots from a model with all predictors.

For L1 English speakers, separate regression models showed a marginal effect of AFT on connectives (Adj-R^2: .05, $p = .05$), whereas ART performed modestly ($F(1,55) = 7.65$, Adj-R^2: .11, $p < .01$). Figure 2B shows standardised partial residual plots from a model with all predictors. To illustrate the differences in the two

as a connective in 39.8% of instances. While this is low, the proportionally adjusted value is 29.69 ppm, which is still slightly higher than "furthermore". We have opted to leave the original coding intact.

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

Table 3. Accuracy scores as percentages per connective, by frequency (high/low) and cohort

			L1		L2	
Relation	Connective	Frequency	Mean	SD	Mean	SD
Addition	Indeed	High	0.52	0.50	0.58	0.49
	Furthermore	Low	0.83	0.37	0.76	0.43
Cause	Since	High	0.91	0.29	0.75	0.43
	Given that	Low	0.94	0.25	0.71	0.45
Concession	Despite this	High	0.87	0.34	0.80	0.40
	Nevertheless	Low	0.90	0.30	0.80	0.40
Condition	As long as	High	0.93	0.25	0.83	0.38
	Provided	Low	0.90	0.30	0.61	0.49
Consequence	Therefore	High	0.90	0.30	0.79	0.41
	Hence	Low	0.79	0.41	0.62	0.49
Contrast	Whereas	High	0.91	0.29	0.77	0.42
	Conversely	Low	0.85	0.36	0.62	0.49
Global	*High frequency*		0.84	0.37	0.75	0.43
	Low frequency		0.87	0.34	0.69	0.46
	Overall		**0.85**	**0.35**	**0.72**	**0.45**

print exposure predictors across language groups, we constructed exploratory GLMER models. Our final model included fixed effects of AFT, ART, connective frequency and coherence relation, and their interactions with language group, as well as random intercepts for participants and items (Marginal-R^2 = .16, Conditional-R^2 = .34; Table 4). Contrasts were dummy-coded, with the baseline set to "low" for connective frequency, "addition" for coherence relation and "L1" for group. Main effects for all coherence relations were significant (*OR*s = 2.59–5.74), though confidence intervals varied widely when comparing across groups. There was also a significant negative interaction for the L2 group for all coherence relations except for "concession". The main effect of frequency was non-significant, but interacted with language group such that L2 speakers showed significantly increased odds in the high frequency condition compared to L1 speakers (*OR* = 1.64, $p < .001$). Main effects for ART and AFT were also non-significant, but there was a significant interaction between AFT and language group, such that AFT predicted increased odds ratios in L2 (*OR* = 2.12, $p < .001$). Thus, for each 1 SD increase in AFT (5.32 author names in L2), the odds of correct selections increased by 112% for L2 compared to L1 speakers. Fixed effects are visualised in Figure 3.

3.2. Collocations

For the collocations task, L2 trial accuracy was as low as 6.67% for *"refuse an application"* (due to competition from *"deny an application"*) to as high as 80% for *"fair share"*. Detailed statistics on the full list of items by language group are provided in Supplementary Table S15, and group differences in score distributions are illustrated in Supplementary Figure S10.

Comparing non-nested linear regression models with predictors of the English AFT and ART separately favoured the model with AFT (ΔAIC = −10.92), and our best fitting model ($F(2, 57) = 63.39$, $p < .001$, Adj-R^2: .68) included both LexTALE ($F(1, 57) = 123.46$, $p < .001$) and AFT, although this was marginal ($F(1, 57) = 3.32$, $p = .07$) (Supplementary Table S18). ART was not significant when accounting for either of the other variables. To illustrate the differential contributions of each predictor, standardised partial residual plots from a model with all predictors are shown in Figure 4A.

Using L1 French predictors, separate linear regression models showed the French AFT ($F(1, 58) = 6.79$, $p < .05$, Adj-R^2: .09) (Supplementary Table S19) and ART ($F(1, 58) = 5.75$, $p < .05$, Adj-R^2: .07) each modestly predicted collocations scores, with negligible differences in model fit (ΔAIC = −0.97), indicating limited explanatory power for L1 print exposure. The French LexTALE was not associated with L2 collocations scores. For comparison across the French LexTALE, ART and AFT, Supplementary Figure S13 shows standardised partial residual plots from a model with all predictors.

For L1 English speakers, individual regression models predicting collocations scores showed a null effect of AFT, but a significant albeit small effect of ART ($F(1, 55) = 7.65$, $p < .01$, Adj-R^2: .11). As with the connectives task, the English ART was a better predictor compared to AFT – an opposite finding to L2 speakers. However, our best model included LexTALE alone ($F(2, 55) = 10.78$, $p < .001$, Adj-R^2: .15). For comparison with L2, we provide residual plots from a model including all predictors in Figure 4B.

We also constructed an exploratory GLMER predicting the odds of correct collocation selections. Our final model included fixed effects of AFT, ART and collocation frequency (as a continuous measure, using values from Dąbrowska, 2014), and their interactions with language group, with random intercepts for participants and items (Marginal-R^2: .15, Conditional-R^2: .35; Table 5). Significant main effects were found for ART (OR = 1.55, $p = .001$) and language group (OR = 0.24, $p < .001$), but AFT and frequency were non-significant. However, there were significant interactions with language group, with AFT predicting increased odds ratios in L2 compared to L1 (OR = 1.67, $p < .01$), translating into 67% higher odds per 1 SD in AFT score; and for frequency and language group, predicting increased odds for higher-frequency collocations in L2 compared to L1 speakers (OR = 1.17, $p < .05$). ART also marginally predicted lower odds in L2 compared to L1 (OR = 0.73, $p = .08$). Fixed effects are visualised in Figure 5.

3.3. Mediating effects of semantic fluency

To evaluate whether verbal fluency generally could moderate the effect of AFT, we re-recruited participants for a test of semantic fluency with three different item categories: "animals", "grocery items" and "public figures". Some participants interpreted the instructions incorrectly, providing names of French supermarket chains instead of grocery items, and categories of public figures (e.g., actor, musician) instead of proper names, but we opted to keep these observations. We removed one participant who entered all items in French. Below, we compare both a combined measure with the sum of all scores, as well as the individual subtasks.

A regression model showed that AFT ($F(1, 44) = 5.81$, $p < .05$) and the SF sum score ($F(1,44) = 31.46$, $p < .001$) co-predicted L2 connectives ($F(2, 44) = 18.64$, $p < .001$, Adj-R^2: .43). For L2 collocations, only AFT predicted the outcome ($F(1, 44) = 28.43$, $p < .001$, model Adj-R^2: .38), whereas SF was non-significant.

Analysis by subtask revealed divergent outcomes. A model predicting connectives with AFT and animal naming ($F(2, 44) = 22.15$, $p < .001$, Adj-R^2: .48) showed effects of AFT ($F(1, 44) = 31.16$, $p < .001$) and animals ($F(1, 44) = 13.15$, $p < .001$); another model comparing AFT and groceries ($F(2, 44) = 15.50$, $p < .001$, Adj-R^2:

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

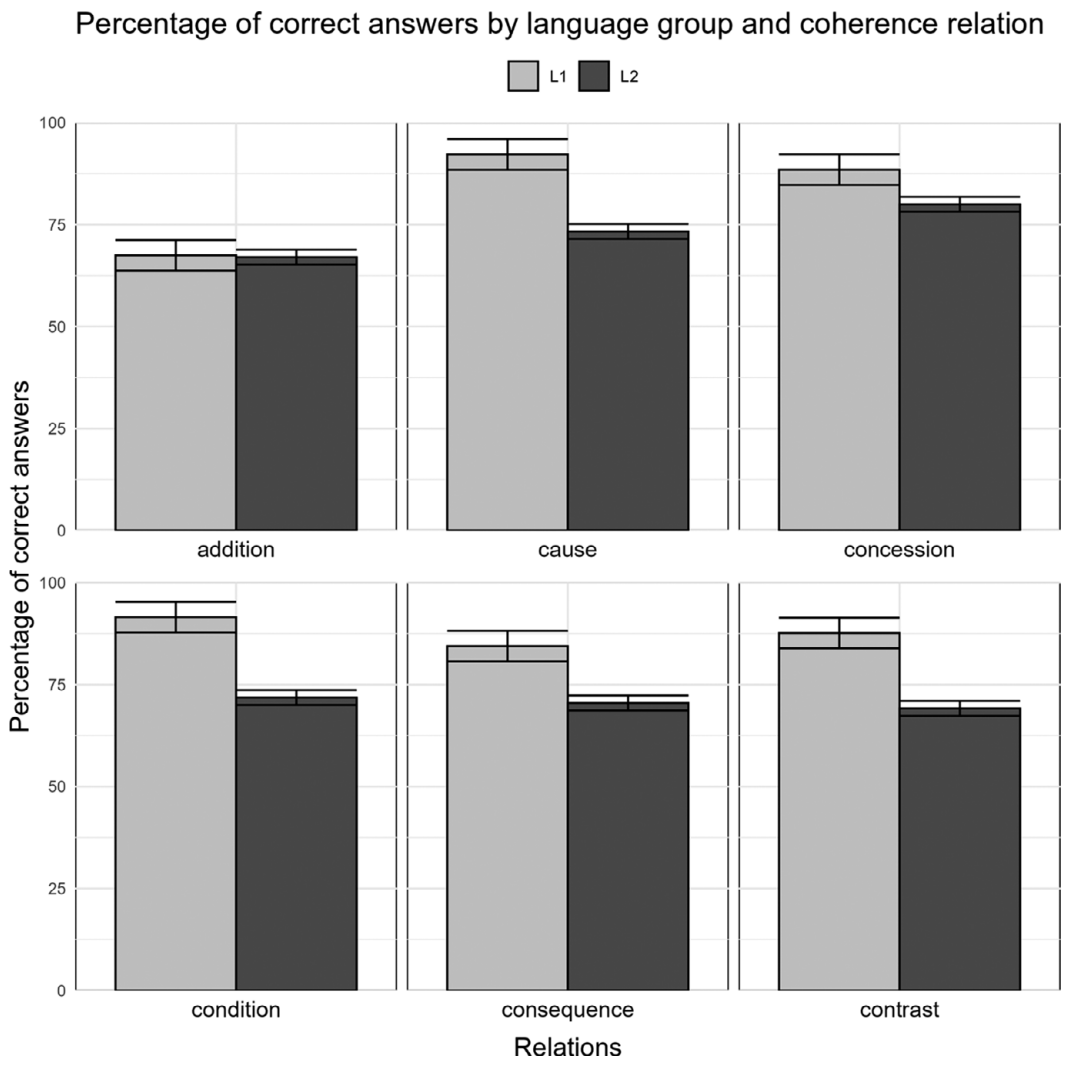


Figure 1. Percentage of correct answers by language group and coherence relation.

.39) showed effects of AFT ($F(1, 44) = 26.46$, $p < .001$) and groceries ($F(1, 44) = 4.54$, $p <. 05$); and a model with AFT and public figures ($F(2, 44) = 12.57$, $p < .001$, Adj-R^2: .33) showed an effect of AFT ($F(1, 44) = 24.39$, $p < .001$) but a null effect for public figures. Analogous models predicting collocations from AFT and animals, groceries and public figures only showed effects of AFT ($F(1, 44) = 27.21–28.85$), all ps < .001.

4. General discussion

We sought to validate a semantic fluency task for author names in L2 (AFT) as a measure of print exposure using outcome measures of formulaic vocabulary, and to determine the relative contributions of L1 and L2 print exposure for L2 vocabulary knowledge. We hypothesised that when controlling for LexTALE, ART would predict collocations, whereas AFT would predict connectives. The rationale for this was that the two print exposure measures might reflect the different kinds of memory required for each task. That is, evaluating the correct use of connectives requires not only word recognition but also knowledge of their function; conversely, evaluating collocations is a far more automatic process – either you know which words tend to co-occur more than others, or you do not. In fact, however, we found that AFT was more positively correlated with both L2 connectives and collocations compared to ART.

The finding that L2 AFT scores predict significant additional variance beyond LexTALE for connectives scores, and marginally so for collocations, further underscores the importance of reading for acquiring L2 vocabulary. Moreover, this was not the case for ART. Given the high variability in L2 exposure and proficiency, and the restrictive nature of ART, this is not entirely surprising. That an open-ended measure like AFT performs well in this regard, however, even when accounting for L2 proficiency, is the primary contribution of the present research. Second language research is replete with discussions about how to access L2 learners' *"cultural capital"* (Bourdieu, 1986; Tunmer et al., 2006), yet when evaluating the role of print exposure in these populations, researchers have not always acknowledged that the language experiences of L2 speakers rarely mirror those of English natives. Consequently, an effective and reliable proxy measure of L2 print exposure may not be the same as one used for L1. This is precisely what we demonstrate, with interactive models showing ART is most effective in L1, and AFT exceeding in L2.

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

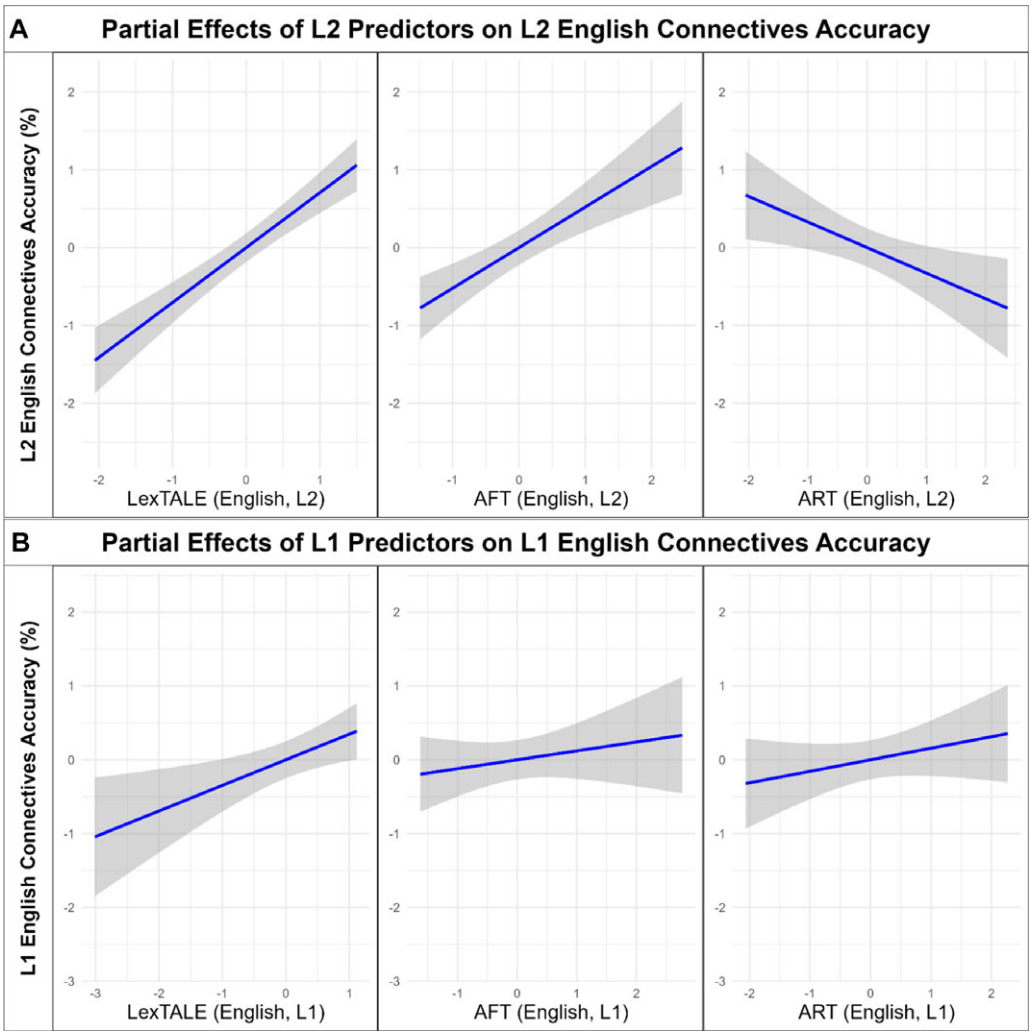


Figure 2. (A) Partial effects of L2 predictors on L2 English connectives accuracy. (B) Partial effects of L1 predictors on L1 English connectives accuracy.

Furthermore, measures of L2 proficiency and print exposure outperformed analogous L1 measures as predictors of L2 vocabulary. Logically, one's degree of exposure to a particular language should explain more about vocabulary knowledge in that language, compared with exposure to another. Yet L1 experience is generally considered fundamental, laying the groundwork for learning additional languages (Sparks et al., 2012). Again, our connectives measure is an adapted version of a task from a study in which L2 proficiency was predicted by an L1, but not L2 ART (Wetzel et al., 2020). We maintain that this was due to limited L2 exposure, which makes ART unlikely to be useful for L2 beginners. Granted, L1 proficiency is undoubtedly a limiting factor for L2 novices, but the question of what distinguishes advanced L2 speakers is a separate one. Once a speaker becomes relatively proficient in a target language, it follows that more extensive and naturalistic L2 exposure becomes critical. However, we acknowledge that our participants were also older than those recruited by Wetzel and colleagues, and since print exposure increases with age, their effects are difficult to disentangle. Similarly, years of exposure to L2 and the age of acquisition also influence print exposure and, consequently, proficiency. Most likely, both age and print exposure are implicated to some degree in explaining our results.

Despite the criticisms surrounding ART, it is interesting to note that author recognition still correlated with L2 vocabulary in our study – although considerably less than AFT. While ART may index L2 print exposure in advanced L2 speakers such as these, its overlap with proficiency measures might lead researchers to infer null effects for print exposure when controlling for other tasks. However, we acknowledge that AFT is unlikely to be useful for novice learners either, given their limited L2 reading experience.

Additionally, we found that a semantic fluency (SF) aggregate measure also correlated positively with connectives and collocations (Table 2). For predicting connectives scores, this SF measure moderated some, but not all variance explained by AFT. For collocations, SF scores were non-significant when paired with AFT. This also varied by outcome measure and by SF subtask. For connectives, "animals" and "groceries" remained significant when controlling for AFT, but "public figures" became non-significant. This distinction is likely because lexical access pathways are distinct for common and proper nouns (Proverbio et al., 2001; Semenza, 2009). Semantic fluency for authors and celebrities both requires recall of proper nouns, and we observe the expected outcome that author names are more informative than public figures, as the former index reading experience whereas the

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

Table 4. Fixed effects and their interactions with language group, and random effects of participant/item on odds of correct connectives selections. Bold *p*-values indicate $p < .05$.

Predictors	Odds ratios	*CI*	*p*
(Intercept)	2.828	1.397–5.725	**.004****
ART	1.258	0.923–1.713	.146
AFT	1.086	0.797–1.479	.602
Frequency [high]	0.931	0.553–1.568	.789
Relation [cause]	5.741	2.327–14.161	**<.001*****
Relation [concession]	3.596	1.481–8.730	**.005****
Relation [condition]	5.299	2.153–13.037	**<.001*****
Relation [consequence]	2.588	1.074–6.240	**.034***
Relation [contrast]	3.290	1.357–7.974	**.008****
Group [L2]	0.753	0.484–1.172	.209
ART × group [L2]	0.800	0.514–1.245	.324
AFT × group [L2]	2.118	1.357–3.306	**.001*****
Freq [high] × group [L2]	1.639	1.250–2.149	**<.001*****
Relation [cause] × group [L2]	0.251	0.158–0.401	**<.001*****
Relation [concession] × group [L2]	0.644	0.413–1.004	.052
Relation [condition] × group [L2]	0.259	0.163–0.412	**<.001*****
Relation [consequence] × group [L2]	0.483	0.317–0.736	**.001*****
Relation [contrast] × group [L2]	0.345	0.224–0.530	**<.001*****
Random effects			
σ^2	3.29		
$\tau_{00\ Participant}$	0.69		
$\tau_{00\ Item}$	0.18		
ICC	0.21		
$N_{Participant}$	117		
N_{Item}	12		
Observations	7020		
Marginal R^2/conditional R^2	0.163/0.337		

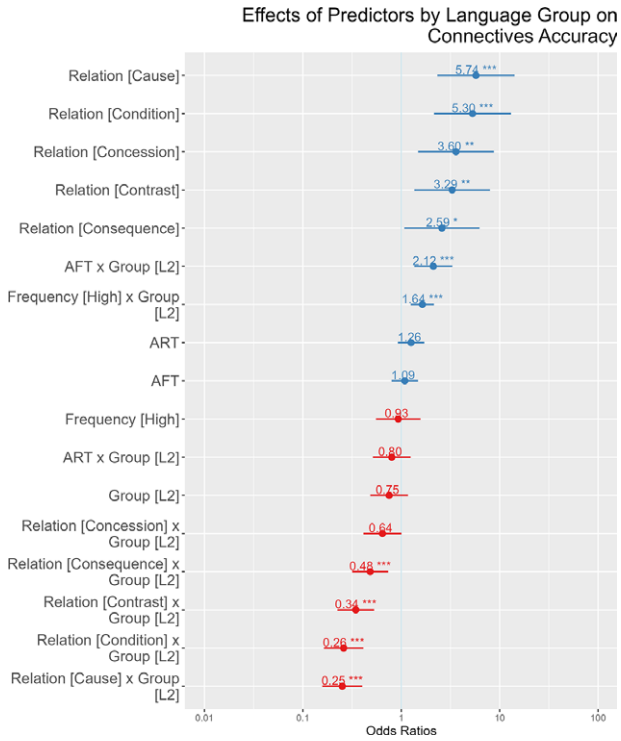


Figure 3. Effects of predictors by language group on connectives accuracy.

latter reflect general cultural exposure. Conversely, no semantic fluency subtask predicted collocations when paired with AFT. These divergent outcomes suggest that semantic fluency, or something associated with it, plays a larger role in the processing of connectives, and print exposure is more important for acquiring collocations. We suspect that if the variance explained by AFT in L2 were simply due to differences in fluency alone, first, we would also observe some effect of semantic fluency for the collocations task when paired with AFT. Second, a similar effect for AFT would likely also be seen in the L1 English population. But in fact, we see a sort of "inverted picture", where ART is the better predictor for L1 speakers, and AFT outperforms in L2. It is possible that the role of semantic fluency is simply stronger in L2 than in L1, given the wider range of L2 skill generally, and research demonstrating L1 and L2 speakers are primarily differentiated by fluency rather than comprehension (Kuperman et al., 2023; Siegelman et al., 2024). Yet it is unclear why such variance would not have been sufficiently captured by LexTALE, which also measures lexical access. We contend that the limiting factor for AFT is familiarity with authors (and consequently, is a reliable proxy for reading experience) rather than verbal fluency generally, as indicated by the null effect of public figure naming when paired with AFT. Thus, we argue that AFT reflects the additional engagement with reading required to become highly proficient in L2.

Before concluding, we note some limitations to our study. First, we calculated AFT scores using one point for each author, with no weighting for authors who are perceived to be more (or less) valuable to the reader. Perhaps more popular author names are more likely to represent general cultural knowledge rather than personal reading experience – after all, one need not have read any of Stephen King or Jane Austen's books for them to come readily to mind when thinking of authors, and they may be associated with Hollywood adaptations of their works rather than the original material. Developing weights for author names is a complex and delicate issue, but one that bears consideration.

Although LexTALE was a robust predictor of vocabulary knowledge, it also has limitations. LexTALE is a word recognition measure, and word knowledge is a multidimensional construct, with depth of word knowledge and meaning a better metric than knowledge or recognition of form (Jeon & Yamashita, 2022). As a lexical decision task, LexTALE only indexes knowledge of word form (and correspondingly, processing speed). Moreover, some evidence suggests that although the LexTALE is a robust measure of vocabulary knowledge, it may not be reliable as a global proficiency measure in L2 (Puig-Mayenco et al., 2023). Thus, a more sensitive measure may be required to separate the effects of L2 proficiency, semantic fluency and L2 print exposure.

The study may also have benefited from a larger sample size, as online studies generally require more observations due to increased variability in testing conditions (Rodd, 2024). To ensure these findings are robust, AFT will need to be replicated in greater numbers, and participants should complete the task on multiple

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

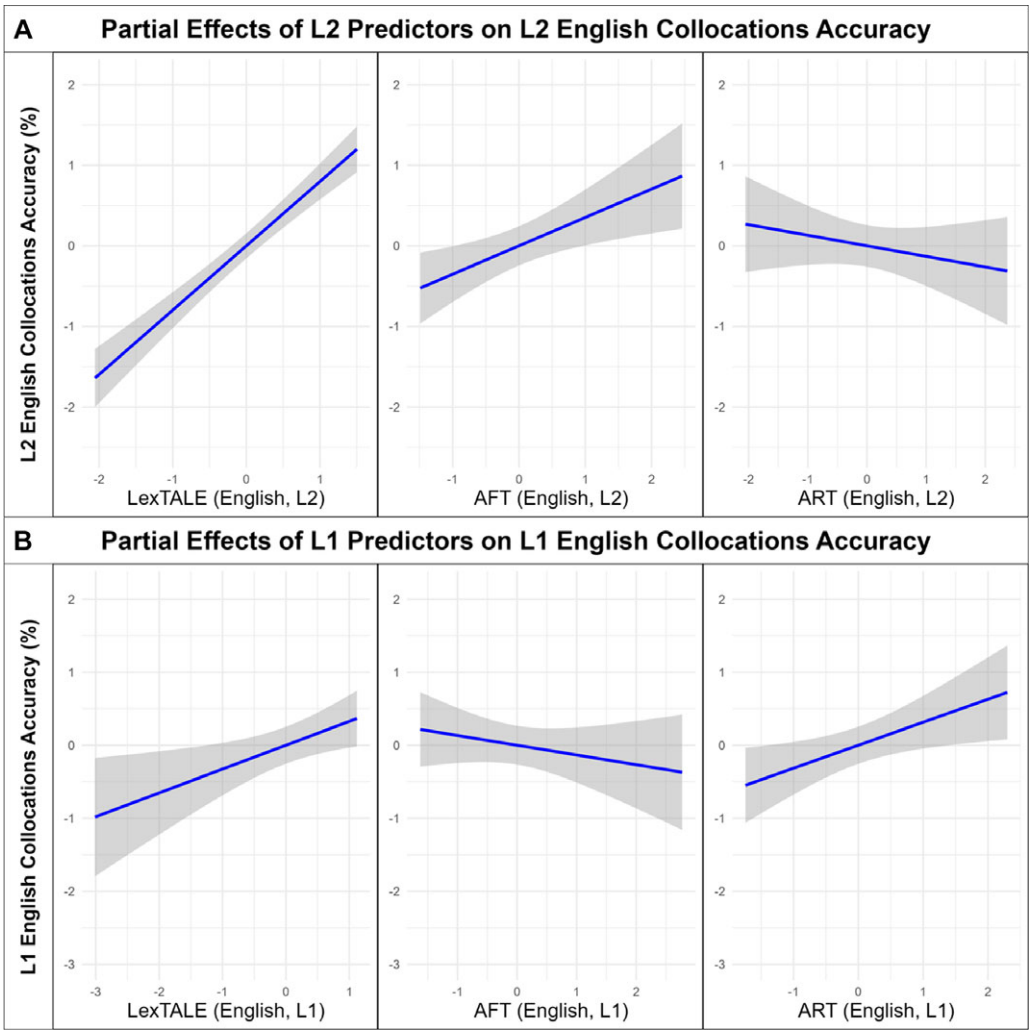


Figure 4. (A) Partial effects of L2 predictors on L2 English collocations accuracy. (B) Partial effects of L1 predictors on L1 English collocations accuracy.

occasions to determine its test–retest reliability. Although its reliability is likely comparable to other semantic fluency tasks, this metric is an important dimension of a test's utility and would provide additional insight into its use across language groups. AFT will also require replication in diverse language populations, since our findings may be partially related to the close linguistic distance between English and French. However, we suspect this is unlikely to completely explain the results, since the similarities between these two languages might be expected to instead diminish the importance of L2 reading experience. Similarly, it is possible our findings may apply primarily to English L2 speakers due to the global spread and influence of English, which means many English authors benefit from the language's broad reach and market dominance. Future studies will determine if these findings generalise well to other target languages, although we expect AFT will be most effective in languages with a similar culture of readership to English.

L2 learners may also know many English authors, but their personal experience with them could be primarily through L1 translations. This first iteration of AFT did not ask participants which L2 authors they provided were read in an L1 translation, and this could be an important modification for subsequent studies. We attempted to diminish this by instructing participants to only name L2 authors who had been published in English. An alternative phrasing for these instructions could have asked participants to instead name authors they had read personally, but we considered this to be too restrictive, especially given that this restriction was not present for ART. Instead, we determined it was more important to allow participants to name whichever authors came most readily to mind when they thought about L2 reading generally. Undoubtedly, some of their actual encounters with these authors will have been through translation. Yet we assert that although an increase in author names may not directly reflect primary print exposure, just as for ART, it nevertheless indicates increased familiarity with authors associated with the target language. Additionally, as our participants self-reported to have intermediate or greater proficiency in English, we suspect that they would have ample opportunity and interest to explore these works as they originally appeared, rather than reading translations. Therefore, we reason that the potential impact of reading translations of these works is unlikely to explain the robust correlations between the L2 AFT and L2 connectives/collocations scores.

Finally, this study reinforces previous findings that indicate that explicit recall tasks more accurately assess L2 language proficiency. Unlike self-report surveys or the ART, a semantic fluency task for author names allows second language learners to demonstrate their print knowledge directly, while reducing concerns about social

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

Table 5. Fixed effects and their interactions with language group, and random effects of participant/item on odds of correct collocations selections. Bold *p*-values indicate $p < .05$.

Predictors	Odds ratios	CI	*p*
(Intercept)	3.909	2.857–5.347	**< .001*** **
ART	1.554	1.207–2.000	**< .001*** **
AFT	0.912	0.711–1.171	.470
Collocation frequency	1.047	0.804–1.364	.733
Group [L2]	0.237	0.180–0.310	**< .001*** **
ART × group [L2]	0.725	0.507–1.037	.079
AFT × group [L2]	1.673	1.172–2.387	**.005****
Collocation frequency × group [L2]	1.173	1.024–1.345	**.022***
Random effects			
σ^2	3.29		
$\tau_{00\ Participant}$	0.40		
$\tau_{00\ Item}$	0.61		
ICC	0.24		
$N_{Participant}$	117		
N_{Item}	40		
Observations	4680		
Marginal R^2/conditional R^2	0.148/0.349		

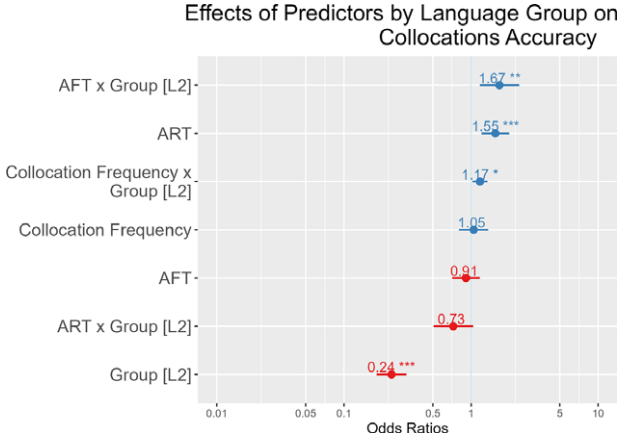


Figure 5. Effects of predictors by language group on collocations accuracy.

desirability bias or guessing. As a practical and intuitive measure, the AFT offers a useful alternative or complement to existing print exposure assessments, and may help to refine our understanding of how reading experience contributes to second language learning.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S136672892510045X.

Data availability statement. The data that support the findings of this study are openly available on OSF at https://osf.io/q62mt/.

Acknowledgements. Thank you to Hui Zhu for contributing to the French–English translations of the connectives task derived from Wetzel et al. (2020).

Competing interests. The authors declare none.

References

Acheson, D. J., **Wells, J. B.**, & **MacDonald, M. C.** (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, **40**(1), 278–289. https://doi.org/10.3758/BRM.40.1.278.

Andersson, M. (2016). *The architecture of result relations: Corpus and experimental approaches to result coherence relations in English [Doctoral dissertation]*. Department of English, Stockholm University.

Andersson, M., & **Sundberg, R.** (2021). Subjectivity (re)visited: A corpus study of English forward causal connectives in different domains of spoken and written language. *Discourse Processes*, **58**(3), 260–292. https://doi.org/10.1080/0163853X.2020.1847581.

Anwyl-Irvine, A. L., **Massonnié, J.**, **Flitton, A.**, **Kirkham, N.**, & **Evershed, J. K.** (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, **52**(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x.

Arnon, I., & **Christiansen, M. H.** (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science*, **9**(3), 621–636. https://doi.org/10.1111/tops.12271.

Baker, D. L., **Stoolmiller, M.**, **Good, R. H.**, & **Baker, S. K.** (2011). Effect of reading comprehension on passage fluency in Spanish and English for second-grade English learners. *School Psychology Review*, **40**(3), 331–351. https://doi.org/10.1080/02796015.2011.12087702.

Barr, D. J., **Levy, R.**, **Scheepers, C.**, & **Tily, H. J.** (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., **Mächler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berger, S. A., **Hall, L. K.**, & **Bahrick, H. P.** (1999). Stabilizing access to marginal and submarginal knowledge. *Journal of Experimental Psychology: Applied*, **5**(4), 438–447. https://doi.org/10.1037/1076-898X.5.4.438.

Berman, R. A., & **Nir, B.** (2010). The lexicon in writing–speech-differentiation. *Written Language & Literacy*, **13**(2), 183–205. https://doi.org/10.1075/wll.13.2.01ber.

Biber, D. (1988). *Variation across speech and writing* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511621024.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. **23**). John Benjamins Publishing Company. https://doi.org/10.1075/scl.23.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511486456.

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood Press.

Brysbaert, M. (2013). Lextale_FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica*, **53**(1), 23. https://doi.org/10.5334/pb-53-1-23.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, **82**(4), 711–733. https://doi.org/10.1353/lan.2006.0186.

Bybee, J. (2007). The emergent lexicon. In *Frequency of use and the organization of language* (1st ed.). Oxford University Press; New York. https://doi.org/10.1093/acprof:oso/9780195301571.003.0013

Bybee, J. (2010). *Language, usage and cognition* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511750526.

Cantor, A. D., **Eslick, A. N.**, **Marsh, E. J.**, **Bjork, R. A.**, & **Bjork, E. L.** (2015). Multiple-choice tests stabilize access to marginal knowledge. *Memory & Cognition*, **43**(2), 193–205. https://doi.org/10.3758/s13421-014-0462-6.

Chateau, D., & **Jared, D.** (2000). Exposure to print and word recognition processes. *Memory & Cognition*, **28**(1), 143–153. https://doi.org/10.3758/BF03211582.

Clark, E. V. (2020). Conversational repair and the acquisition of language. *Discourse Processes*, **57**(5–6), 441–459. https://doi.org/10.1080/0163853X.2020.1719795.

Conklin, K., & **Schmitt, N.** (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, **32**, 45–61. https://doi.org/10.1017/S0267190512000074.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, **49**(2), 222. https://doi.org/10.2307/1169960.

Cunningham, A. E., & **Stanovich, K. E.** (1998). What reading does for the mind. *American Educator*, **22**(1–2), 8–15.

Dąbrowska, E. (2014). Words that go together: Measuring individual differences in native speakers' knowledge of collocations. *The Mental Lexicon*, **9**(3), 401–418. https://doi.org/10.1075/ml.9.3.02dab.

Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, **178**, 222–235. https://doi.org/10.1016/j.cognition.2018.05.018.

Dąbrowska, E., & **Street, J.** (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences*, **28**(6), 604–615. https://doi.org/10.1016/j.langsci.2005.11.014.

Dawson, N., **Hsiao, Y.**, **Tan, A. W. M.**, **Banerji, N.**, & **Nation, K.** (2021). Features of lexical richness in children's books: Comparisons with child-directed speech. *Language Development Research*, **1**(1), 9–53. https://doi.org/10.34842/5WE1-YK94

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, **18**(1), 91–126. https://doi.org/10.1017/S0272263100014698.

Ellis, N. C., **Simpson-Vlach, R.**, & **Maynard, C.** (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, **42**(3), 375–396. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x.

Faul, F., **Erdfelder, E.**, **Lang, A.-G.**, & **Buchner, A.** (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, **39**(2), 175–191. https://doi.org/10.3758/BF03193146.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-55. In *Studies in linguistic analysis* (pp. 1–31). Basil Blackwell.

Flege, J. E. (2008). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Multilingual Matters. https://doi.org/10.21832/9781847691118-012.

Flege, J. E. (2019). A non-critical period for second-language learning. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A Sound Approach to Language Matters: In Honor of Ocke-Schwen Bohn* (pp. 501–541). Aahus University Library. https://doi.org/10.7146/aul.322.218

Friesen, D. C., **Luo, L.**, **Luk, G.**, & **Bialystok, E.** (2015). Proficiency and control in verbal fluency performance across the lifespan for monolinguals and bilinguals. *Language, Cognition and Neuroscience*, **30**(3), 238–250. https://doi.org/10.1080/23273798.2014.918630.

Gablasova, D., **Brezina, V.**, & **McEnery, T.** (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, **67**(S1), 155–179. https://doi.org/10.1111/lang.12225.

Gollan, T. H., **Montoya, R. I.**, & **Werner, G. A.** (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, **16**(4), 562. https://doi.org/10.1037/0894-4105.16.4.562.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145–160). Oxford University Press; Oxford. https://doi.org/10.1093/oso/9780198294252.003.007.

Gullifer, J. W., & **Titone, D.** (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, **23**(2), 283–294. https://doi.org/10.1017/S1366728919000026.

Hallin, A. E., & **Van Lancker Sidtis, D.** (2017). A closer look at formulaic language: Prosodic characteristics of Swedish proverbs. *Applied Linguistics*, **38**(1), 68–89. https://doi.org/10.1093/applin/amu078.

Hartig, F. (2022). *DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models* [Computer software]. https://doi.org/10.32614/CRAN.package.DHARMa

Hsiao, Y., **Dawson, N. J.**, **Banerji, N.**, & **Nation, K.** (2022). The nature and frequency of relative clauses in the language children hear and the language children read: A developmental cross-corpus analysis of English complex grammar. *Journal of Child Language*, **50**(3), 1–26. https://doi.org/10.1017/S0305000921000957.

Jeon, E. H., & **Yamashita, J.** (2022). L2 reading comprehension and its correlates: An updated meta-analysis. In E. H. Jeon & Y. In'nami (Eds.), *Bilingual processing and acquisition* (Vol. **13**, pp. 29–86). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.13.03jeo.

Johns, B. T., **Dye, M.**, & **Jones, M. N.** (2020). Estimating the prevalence and diversity of words in written language. *Quarterly Journal of Experimental Psychology*, **73**(6), 841–855. https://doi.org/10.1177/1747021819897560.

Jones, M. N., **Dye, M.**, & **Johns, B. T.** (2017). Context as an organizing principle of the lexicon. In *Psychology of learning and motivation* (Vol. **67**, pp. 239–283). Elsevier. https://doi.org/10.1016/bs.plm.2017.03.008

Kilgarriff, A., **Baisa, V.**, **Bušta, J.**, **Jakubíček, M.**, **Kovář, V.**, **Michelfeit, J.**, **Rychlý, P.**, & **Suchomel, V.** (2014). The sketch engine: Ten years on. *Lexicography*, **1**(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9.

Kuperman, V., **Siegelman, N.**, **Schroeder, S.**, **Acartürk, C.**, **Alexeeva, S.**, **Amenta, S.**, **Bertram, R.**, **Bonandrini, R.**, **Brysbaert, M.**, **Chernova, D.**, **Da Fonseca, S. M.**, **Dirix, N.**, **Duyck, W.**, **Fella, A.**, **Frost, R.**, **Gattei, C. A.**, **Kalaitzi, A.**, **Lõo, K.**, **Marelli, M.**, … **Usal, K. A.** (2023). Text reading in English as a second language: Evidence from the Multilingual Eye-movements Corpus. *Studies in Second Language Acquisition*, **45**(1), 3–37. https://doi.org/10.1017/S0272263121000954.

Kuznetsova, A., **Brockhoff, P. B.**, & **Christensen, R. H. B.** (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, **82**(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lehtinen, N., **Kautto, A.**, & **Renvall, K.** (2023). Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: An extended analysis of verbal fluency task performance in an L1 language attrition population. *International Journal of Bilingualism*, **28**(5), 884–906. https://doi.org/10.1177/13670069231193727.

Lemhöfer, K., & **Broersma, M.** (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, **44**(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0.

Li, F., **Mak, W. M.**, **Evers-Vermeul, J.**, & **Sanders, T. J. M.** (2017). On the online effects of subjectivity encoded in causal connectives. *Review of Cognitive Linguistics*, **15**(1), 34–57. https://doi.org/10.1075/rcl.15.1.02li.

Macoir, J., **Sylvestre, A.**, & **Turgeon, Y.** (2006). Classical tests for speech and language disorders. In *Encyclopedia of Language & Linguistics* (2nd ed., pp. 439–445). Elsevier. https://doi.org/10.1016/B0-08-044854-2/04191-2

Martin-Chang, S., & **Gould, O. N.** (2008). Revisiting print exposure: Exploring differential links to vocabulary, comprehension and reading rate. *Journal of Research in Reading*, **31**(3), 273–284. https://doi.org/10.1111/j.1467-9817.2008.00371.x.

McCarron, S. P. (2026). Author recognition tests. In H. Nesi & P. Milin (Eds.), *Encyclopedia of Language & Linguistics* (3rd ed.). Elsevier. https://doi.org/10.1016/B978-0-323-95504-1.00497-X

McCarron, S. P., & **Kuperman, V.** (2021). Is the author recognition test a useful metric for native and non-native English speakers? An item response theory analysis. *Behavior Research Methods*, **53**(5), 2226–2237. https://doi.org/10.3758/s13428-021-01556-y.

Mol, S. E., & **Bus, A. G.** (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, **137**(2), 267–296. https://doi.org/10.1037/a0021890

Moore, M., & **Gordon, P. C.** (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. *Behavior Research Methods*, **47**(4), 1095–1109. https://doi.org/10.3758/s13428-014-0534-3.

Nation, K., **Dawson, N. J.**, & **Hsiao, Y.** (2022). 'Book language' and its implications for children's language, literacy, and development. *Current Directions in Psychological Science*, **31**(4), 375–380. https://doi.org/10.1177/09637214221103264.

Internet Archive. (2022). OpenLibrary [Website]. https://openlibrary.org/search/authors

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, **21**(3), 381–402. https://doi.org/10.1177/1362168815618428.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, **14**, 84–94. https://doi.org/10.1016/j.jeap.2014.01.002.

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, **18**(1), 75–94. https://doi.org/10.1177/1362168813505384.

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, **20**(1), 113–138. https://doi.org/10.1177/1362168814568131.

Prolific. (2024). *Prolific* [Website]. London, UK. https://www.prolific.co/

Proverbio, A. M., **Lilli, S.**, **Semenza, C.**, & **Zani, A.** (2001). ERP indexes of functional differences in brain activation during proper and common names retrieval. *Neuropsychologia*, **39**(8), 815–827. https://doi.org/10.1016/S0028-3932(01)00003-3.

Puig-Mayenco, E., **Chaouch-Orozco, A.**, **Liu, H.**, & **Martín-Villena, F.** (2023). The LexTALE as a measure of L2 global proficiency: A cautionary tale based on a partial replication of Lemhöfer and Broersma (2012). *Linguistic Approaches to Bilingualism*, **13**(3), 299–314. https://doi.org/10.1075/lab.22048.pui.

R Core Team. (2024). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Restrepo-Ramos, F. D. (2015). Incidental vocabulary learning in second language acquisition: A literature review. *PROFILE Issues in Teachers' Professional Development*, **17**(1), 157–166. https://doi.org/10.15446/profile.v17n1.43957.

Ringbom, H. (2016). Comprehension, learning and production of foreign languages: The role of transfer. In R. A. (Ed.), *Crosslinguistic influence in second language acquisition* (pp. 38–52). Multilingual Matters. https://doi.org/10.21832/9781783094837.

Rodd, J. M. (2024). Moving experimental psychology online: How to obtain high quality data when we can't see our participants. *Journal of Memory and Language*, **134**, 104472. https://doi.org/10.1016/j.jml.2023.104472.

Roland, D., **Dick, F.**, & **Elman, J. L.** (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, **57**(3), 348–379. https://doi.org/10.1016/j.jml.2007.03.002.

Schmitt, N. (2010). *Researching vocabulary*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230293977.

Semenza, C. (2009). The neuropsychology of proper names. *Mind & Language*, **24**(4), 347–369. https://doi.org/10.1111/j.1468-0017.2009.01366.x.

Shin, D. (2007). The high frequency collocations of spoken and written English. *English Teaching*, **62**(1), 199–218. https://doi.org/10.15858/engtea.62.1.200703.199.

Siegelman, N., **Elgort, I.**, **Brysbaert, M.**, **Agrawal, N.**, **Amenta, S.**, **Arsenijević Mijalković, J.**, **Chang, C. S.**, **Chernova, D.**, **Chetail, F.**, **Clarke, A. J. B.**, **Content, A.**, **Crepaldi, D.**, **Davaabold, N.**, **Delgersuren, S.**, **Deutsch, A.**, **Dibrova, V.**, **Drieghe, D.**, **Filipović Đurđević, D.**, **Finch, B.**, … **Kuperman, V.** (2024). Rethinking first language–second language similarities and differences in English proficiency: Insights from the ENglish Reading Online (ENRO) project. *Language Learning*, **74**(1), 249–294. https://doi.org/10.1111/lang.12586.

Siyanova, A., & **Schmitt, N.** (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, **64**(3), 429–458. https://doi.org/10.3138/cmlr.64.3.429.

Siyanova-Chanturia, A., **Conklin, K.**, & **Van Heuven, W. J. B.** (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **37**(3), 776–784. https://doi.org/10.1037/a0022531.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, **328**(5977), 450–452. https://doi.org/10.1126/science.1182597.

Sonbul, S., & **Schmitt, N.** (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, **63**(1), 121–159. https://doi.org/10.1111/j.1467-9922.2012.00730.x.

Sparks, R. L. (2023). The linguistic coding differences hypothesis (LCDH) and L2 learning: A thirty-year retrospective. In Z. (Edward) Wen, P. Skehan, & R. L. Sparks (Eds.), *Language aptitude theory and practice* (1st ed., pp. 275–301). Cambridge University Press. https://doi.org/10.1017/9781009076463.015.

Sparks, R. L., **Patton, J.**, **Ganschow, L.**, & **Humbach, N.** (2012). Do L1 reading achievement and L1 print exposure contribute to the prediction of L2 proficiency? *Language Learning*, **62**(2), 473–505. https://doi.org/10.1111/j.1467-9922.2012.00694.x.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, **21**(4), 360–407. https://doi.org/10.1598/RRQ.21.4.1.

Stanovich, K. E., & **West, R. F.** (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, **24**(4), 402–433. https://doi.org/10.2307/747605.

Strømsø, H. I. (2024). Does students' exposure to websites moderate the positive relationship between print exposure and text comprehension? *Reading and Writing*, **37**(8), 2151–2171. https://doi.org/10.1007/s11145-023-10468-6.

Troyer, A. K., **Moscovitch, M.**, & **Winocur, G.** (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, **11**(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138.

Troyer, A. K., **Moscovitch, M.**, **Winocur, G.**, **Leach, L.**, & **Freedman, M.** (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of the International Neuropsychological Society*, **4**, 137–143. https://doi.org/10.1017/s1355617798001374

Tskhovrebova, E., **Zufferey, S.**, & **Gygax, P.** (2022). Individual variations in the mastery of discourse connectives from teenage years to adulthood. *Language Learning*, **72**, 412–455. https://doi.org/10.1111/lang.12481.

Tunmer, W. E., **Chapman, J.**, & **Prochnow, J.** (2006). Literate cultural capital at school entry predicts later reading achievement: A seven year longitudinal study. *New Zealand Journal of Educational Studies*, **41**, 183–204.

Van Silfhout, G., **Evers-Vermeul, J.**, & **Sanders, T.** (2015). Connectives as processing signals: How students benefit in processing narrative and expository texts. *Discourse Processes*, **52**(1), 47–76. https://doi.org/10.1080/0163853X.2014.905237.

Verhoeven, L. T. (1994). Transfer in bilingual development: The linguistic interdependence hypothesis revisited. *Language Learning*, **44**(3), 381–415. https://doi.org/10.1111/j.1467-1770.1994.tb01112.x.

Vermeiren, H., & **Brysbaert, M.** (2023). How useful are native language tests for research with advanced second language users? *Bilingualism: Language and Cognition*, **27**(1), 204–213. https://doi.org/10.1017/S1366728923000421

Vermeiren, H., **Vandendaele, A.**, & **Brysbaert, M.** (2022). Validated tests for language research with university students whose native language is English: Tests of vocabulary, general knowledge, author recognition, and reading comprehension. *Behavior Research Methods*, **55**(3), 1036–1068. https://doi.org/10.3758/s13428-022-01856-x.

Webb, S., & **Kagimoto, E.** (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, **43**(1), 55–77. https://doi.org/10.1002/j.1545-7249.2009.tb00227.x.

Webb, S., **Newton, J.**, & **Chang, A.** (2013). Incidental learning of collocation. *Language Learning*, **63**(1), 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x.

West, R. F., **Stanovich, K. E.**, & **Mitchell, H. R.** (1993). Reading in the real world and its correlates. *Reading Research Quarterly*, **28**(1), 35–50. https://doi.org/10.2307/747815.

Wetzel, M., **Zufferey, S.**, & **Gygax, P.** (2020). Second language acquisition and the mastery of discourse connectives: Assessing the factors that hinder L2-learners from mastering French connectives. *Language*, **5**(3), 35. https://doi.org/10.3390/languages5030035.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Wray, A. (2006). Formulaic language. In *Encyclopedia of language & linguistics* (pp. 590–597). Elsevier. https://doi.org/10.1016/B0-08-044854-2/04777-5

Zufferey, S., & **Gygax, P.** (2020). "Roger broke his tooth. *However*, he went to the dentist": Why some readers struggle to evaluate wrong (and right) uses of connectives. *Discourse Processes*, **57**(2), 184–200. https://doi.org/10.1080/0163853X.2019.1607446.

Zufferey, S., & **Gygax, P. M.** (2017). Processing connectives with a complex form-function mapping in L2: The case of French "en effet. *Frontiers in Psychology*, **8**(1198), 1–11. https://doi.org/10.3389/fpsyg.2017.01198.

Zufferey, S., **Mak, W.**, **Degand, L.**, & **Sanders, T.** (2015). Advanced learners' comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Second Language Research*, **31**(3), 389–411. https://doi.org/10.1177/0267658315573349.

https://doi.org/10.1017/S136672892510045X Published online by Cambridge University Press