Applied Psycholinguistics (2022), **43**, 889–919
doi:10.1017/S0142716422000182

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Morphological knowledge in English learner university students is sensitive to language statistics: A longitudinal study

Daniel Schmidtke*, Sadaf Rahmanian and Anna L. Moro

Department of Linguistics and Languages, McMaster University, MELD Office, L.R. Wilson Hall, Hamilton, Ontario, Canada L8N 1E9
*Corresponding author. Email: schmiddf@mcmaster.ca

(Received 5 April 2021; revised 23 March 2022; accepted 1 June 2022; first published online 19 July 2022)

Abstract
Exposure to statistical patterns of language use affects language production and comprehension. In this longitudinal study of English language learner (ELL) university students, we examined the interplay between language experience and language statistics as a window into the formation and stability of morphological representations in memory. We hypothesized that within-participant change in sensitivity to distributional properties of complex words on written production would reflect changes in morphological knowledge. At two timepoints, separated by 8 months of language exposure, a sample of ELLs ($n = 196$) completed a written suffix completion task. The largest gains in production accuracy were observed for derived words ending in less productive suffixes. In addition, across both timepoints we found a consistent effect of derivational family entropy, such that derived words belonging to morphological families with equally dominant members were less accurately produced. Both effects indicate that ELLs exploit distributional cues to morphological structure and shed light on two aspects of morphological knowledge in ELLs. First, knowledge of suffixes becomes more entrenched in memory, independently of knowledge of the full forms of derived words. Second, ELLs draw upon interlexical connections between morphological family members during written word production.

Keywords: derived words; morphology; ELLs; word production; spelling; language development

Introduction

Though decades of psycholinguistic research have provided valuable insight into the mental storage and processing of morphologically complex words, it remains unclear how knowledge of morphological structure develops within a growing mental lexicon. A common experimental strategy that is used to probe the mental representation of morphological structure is to examine the impact of distributional, frequency-based properties of complex words on lexical processing outcomes (see reviews in Amenta &

Crepaldi, 2012 and Marelli et al., 2020). For example, the effects of the frequencies of morphemes (e.g., *play* and *-er* in *player*) and the size of the morphological families of complex words (e.g., *playful*, *playlist*, *playoff*) are influential in the processing of complex words. Since these effects are used as indices of the mental storage and processing of morphological structure, by implication, observing changes in sensitivity to these frequency-based properties as a function of language experience ought to shed light on the development of morphological knowledge in a dynamic mental lexicon. The present study tests this hypothesis in a novel way by examining how within-individual gains in language experience and the distributional morpho-lexical properties of derived words influence the development of morphological knowledge in English language learner (ELL) university students.

Lexical characteristics as indices of morphological knowledge

A common factor uniting most accounts of complex word processing is that complex word processing in a first (L1) and second language (L2) is sensitive to a wealth of quantitative properties pertaining to the full forms of complex words and their morphological constituents (for L1 accounts, see Baayen & Schreuder, 1999; Caramazza et al., 1988; Kuperman et al., 2009; Libben, 2005, 2014; Moscoso del Prado Martín, 2007; for L2 accounts, see Diependaele et al., 2011; Feldman & Kroll, 2020; Libben et al., 2017). In particular, the effects of frequency-based properties of complex words and their constituents on measures of cognitive effort during lexical processing are often used as indices of access to the mental representations of complex words and their morphological structure (for behavioral studies, see e.g., Andrews, 1986; Baayen et al., 2007; Juhasz et al., 2003; Kuperman & Van Dyke, 2011; Taft, 2004; for neurophysiological studies, see e.g., Fruchter & Marantz, 2015; Leminen et al., 2013; Solomyak & Marantz, 2010). This study focuses on three commonly used measures: the frequencies of complex words, their constituents, and their morphological family members.

The presence of the derived word frequency effect (often referred to as surface or whole-word frequency effect), independently of effects of morphological properties, has been interpreted as a signature of access to the whole complex word representation (e.g., Beauvillain, 1996; Fruchter & Marantz, 2015; Niswander et al., 2000; Taft, 1979; for an excellent review see Lõo et al., 2018a). Derived word frequency predicts acoustic durations and naming latencies of words in speech production (e.g., Caselli et al., 2016; Lõo et al., 2018a), and reading times in visual word recognition (Baayen et al., 2007; Kuperman et al., 2009). Turning to morphological effects, a common diagnostic measure of access to morphological structure is the effect of stem frequency, e.g., the standalone frequency of *humor* as a stem in the complex word *humorous*. Another diagnostic of access to morphological structure is affix productivity, defined as the number of word types that share the same suffix as the target word, e.g., the number of words that end in *-ous* during the processing of *humorous* (Baayen et al., 2007; Bertram et al., 1999; Bertram et al., 2000; Kuperman et al., 2010; Laudanna & Burani, 1995; Laudanna et al., 1994). In summary, the effect of derived word frequency signals access to the whole complex word representation, and the stem frequency and affix productivity on complex word processing typically serve as evidence in support of the engagement of morphological units during word recognition.

The effect of morphological family size (i.e., the number of derived and compound word types sharing a stem, e.g., *joyful*, *joyous*, *joyride*) are also used as diagnostics of the processing of morphological structure. Complex words that have larger morphological families tend to be processed faster than those that have smaller families (e.g., Bertram et al., 2000; De Jong et al., 2000; Lõo et al., 2018b). Family size effects are also present in L2 lexical processing (Dijkstra et al., 2005; Mulder et al., 2014; Mulder et al., 2015), demonstrating that effects of morphological relations between words occur once a morphological paradigm is acquired in a non-native language. Research on L1 morphological processing also shows that the frequency distribution of words within an inflectional and derivational paradigm predicts complex word processing (for an excellent overview see Milin and Blevins, 2020). An important finding in this regard is that complex words with morphological families that have fewer dominant members, that is, those with a small number of highly frequent members existing in the language, are recognized faster than those with morphological families that have many dominant members, that is, those with many relatively equally frequent members (Fruchter & Marantz, 2015; Moscoso del Prado Martín et al., 2004; Schmidtke et al., 2017). Such family-based effects are argued to arise as a consequence of spreading lexical co-activation throughout a morphological family (De Jong et al., 2000).

If the previously described effects signal access to complex words and their morphemic units during complex word recognition, then what can they tell us about how morphological information is formed in a continuously developing mental lexicon? The present study addresses this question by drawing inspiration from an approach that examines the interplay between individual language experience and language statistics on morphological knowledge (Falkauskas & Kuperman, 2015; Ulicheva et al., 2020). The premise of this approach is that, while statistical patterns of language use affect language processing, individual differences in the *amount of exposure* to these patterns lead to individual differences in language comprehension and production (MacDonald, 2013). Here, we build on prior work in a novel way by examining whether *within-participant* change in language experience modulates the effects of the distributional properties of complex words in written production. We reason that the within-participant changes in sensitivity to frequency-based morphological effects, should they emerge, would constitute evidence of change in morphological knowledge as a result of experience with language. In what follows, we summarize empirical research that examines the interactions between language experience and morpho-lexical knowledge, before outlining our hypotheses.

Language experience and morpho-lexical knowledge

It is widely accepted that language experience is instrumental in forming, refining, and maintaining mental representations of linguistic structure (MacDonald, 2013; Perfetti, 2007; Seidenberg & MacDonald, 1999; Stanovich & West, 1989). A key principle of these accounts is that internalized word knowledge is shaped by the distributional properties of the input, including frequency statistics and co-occurrence patterns of words and phrases. Results from experimental morphological research on the language comprehension-production interface are broadly consistent with this idea. We highlight three strands of research in this regard, paying particular attention to the comprehension-production interface in the written modality.

First, there is substantial evidence from spelling research that monolingual children track graphotactic regularities (i.e., statistical patterns of the positioning of letters in words) from printed language input, including those specific to morphological structure, and take advantage of such regularities when developing spelling ability (Deacon et al., 2008). Deacon et al. reviewed several studies supporting this notion, including a study by Pacton et al. (2005), that demonstrated that children's spelling of French-derived words is influenced by graphotactic regularities even in conditions in which it would have been possible to rely on abstract morphological rules. Deacon et al. (2008) propose that spelling regularities are therefore acquired by attending to "the co-occurrence of letters, sounds, and meaning within words and the frequencies of pairings of words within sentences" (p. 123). This line of research suggests that the written production of complex word forms is partly a reflection of the learning of graphotactic patterns via exposure to printed language.

A second source of evidence in support of the effects of the morpho-distributional properties of complex words on written production comes from typing tasks: these tasks show that keystroke latencies to morphologically complex words are influenced by stem and whole-word frequency (e.g., Feldman et al., 2019; Sahel et al., 2008). Since the frequencies of constituent morphemes and whole words have each been shown to affect production latencies, it is argued that written production draws upon knowledge of morpheme units and holistic representations. This line of research aligns with the larger body of written production work that shows that morphological structure (e.g., the presence of a morphological constituent boundary) plays a role in typing latencies (e.g., Badecker et al., 1990; Bertram et al., 2015; Gagné & Spalding, 2016; Orliaguet & Boë, 1993; Pynte et al., 1991; Weingarten et al., 2004).

Third, cross-sectional research has perhaps more directly addressed the relationship between language experience and the distributional properties of complex words. A word definition study comparing third- and sixth-grade Finnish school children (Bertram et al., 2000) showed that, for both groups, Finnish words with highly productive suffixes (e.g., *laula* + *ja*, meaning "singer") elicited more accurate definitions than those with less productive suffixes (e.g., *taru* + *sto*, meaning "mythology"). However, for sixth-graders, the difference between less and more productive derivational morphemes was not as dramatic as for third-graders. In addition, an eye-movement study by Falkauskas and Kuperman (2015) found that the facilitative effect of spacing bias in compound word recognition (e.g., *ringtone* vs. *ring tone*) was stronger among readers with greater language experience. This finding indicates that exposure to written materials serves to heighten sensitivity to the statistical regularities of complex word spelling (see also Häikiö et al., 2011, for a separate eye-movement study showing the effects of grade level on sensitivity to compound spelling format). More recently, Ulicheva et al. (2020) conducted a series of nonword classification tasks in which participants judged whether an orthographic string resembled a noun or an adjective. They reported that suffix spellings that provide stronger cues to a lexical category were more likely to be judged as members of that category, and that participants with greater reading

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

experience, vocabulary knowledge, and spelling ability were more sensitive to category-specific cues.

Taken together, the research summarized above strongly suggests that readers are sensitive to the distributional and semantic characteristics of morphologically complex words, and that this sensitivity is more acute among readers with greater experience with written language materials. Given these findings, it ought to be possible to demonstrate an empirical association between change in exposure to written language and change in sensitivity to that distributional information *within the same group of individuals*. In what follows, we hypothesize two ways in which language experience may shape morphological knowledge and outline our predictions about how these changes may affect the development of the written production of complex word forms in ELLs.

Hypothesis 1. Morpheme entrenchment

As noted earlier, the effects of stem frequency and suffix productivity are used as diagnostics of sensitivity to morphological structure during lexical processing. We therefore selected these variables in the present study to investigate experience-related changes in the strength of morphological representations during learning. The refinement of and initial reliance on morphological knowledge in language and literacy development has been demonstrated in prior work. It has been established that developing L1 readers begin to form mental representations of complex words by detecting and segmenting words into morphemes (e.g., *perform* and *-ance* in *performance*). This idea is also supported by the results of studies of developing readers of English (Beyersmann et al., 2012), French (Beyersmann et al., 2015), German (Hasenäcker et al., 2020), Hebrew (Schiff et al., 2012), and Italian (Burani et al., 2002). Morphological knowledge is also a predictor of literacy outcomes in developing monolingual readers (see review in Goodwin & Ahn, 2013) and Chinese-speaking ELLs (Jiang & Kuo, 2019; Lam et al., 2012; Zhang & Koda, 2012; 2013). Based on these results, we therefore expect ELLs to rely on knowledge of morphemes: we expect derived words with more frequent stems and more productive suffixes to be more accurately produced than derived words composed of less frequent stems and less productive suffixes. Our specific interest lies, however, in the interactions between exposure to language and stem frequency and exposure to language and suffix productivity.

How might the interaction between language exposure and frequency inform complex word learning? The word frequency effect reflects the stability and entrenchment of lexical representations: more frequent words are encountered more often, which results in stronger lexical representations (Brysbaert et al., 2018). Critically, both L1 and L2 word processing studies show that the word frequency effect becomes weaker as language experience increases, such that less experienced readers, and readers with smaller vocabularies, are especially slow at processing low-frequency words (Brysbaert et al., 2017; Cop et al., 2015; Diependaele et al., 2013; Duyck et al., 2007; Mainz et al., 2017; Monaghan et al., 2017; Whitford & Joanisse, 2018; Whitford & Titone, 2012). In other words, when contrasting low and high proficiency readers, the greatest differences in word processing efficiency are found for low-frequency words. The lexical entrenchment

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

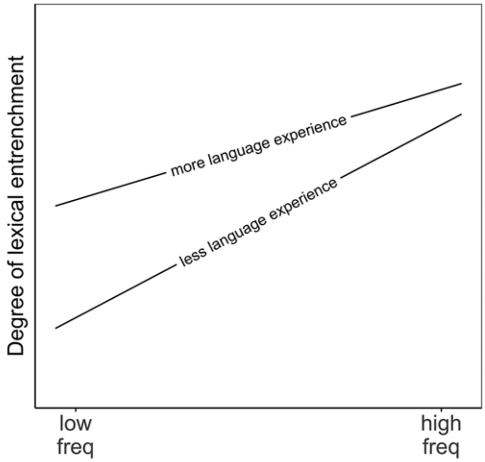

Figure 1. Predicted pattern of language experience on lexical entrenchment based on prior research.

hypothesis was proposed to account for this interaction: less experienced readers "require more energy" to process low-frequency words because these individuals have had fewer opportunities to strengthen their memory representations of these lexical items (Diependaele et al., 2013, p. 846; for a similar proposal see also Kuperman & Van Dyke, 2013). The pattern of the lexical entrenchment effect, based on prior research, is depicted in Figure 1, and shows that greater language experience is associated with a flatter positive slope for the frequency effect.

We investigate the entrenchment effect at the lexical *and* morpheme level. Indeed, Bertram et al. (2000) reported an entrenchment pattern for suffix productivity in a cross-sectional study of Finnish monolinguals. However, to our knowledge such a pattern has not yet been observed in ELLs, nor has the trend been observed in a within-participants longitudinal study design. Based on the predictions of the lexical entrenchment hypothesis, it is expected that continued exposure to language will lead to the greatest observable changes in the production of complex words composed of low-frequency stems or less productive affixes. Evidence for this hypothesis would be supported by observing an interaction between timepoint and suffix productivity or stem frequency, resembling the entrenchment pattern presented in Figure 1. We reason that observing the entrenchment pattern for stem frequency or suffix productivity would indicate that experience with language provides more opportunities to solidify representations of stems and suffixes embedded within derived words, with gains being especially pronounced for derived words with stems and suffixes that occur less frequently in the language.

Hypothesis 2. Morphological connectivity

Prior work has addressed the concept of *relational knowledge* (Tyler & Nagy, 1989), a dimension of morphological knowledge that involves recognizing that multiple words share a common morpheme, that is, the knowledge that *humanity*, *humanistic*, *humanize* and *humankind* all belong to the same derivational family. While a skilled native speaker of English may know many morphologically complex words and have greater implicit awareness of the connections among common members of a morphological family, it is not clear how this relational knowledge develops in language learners. In the present study of ELLs, we examine interlexical connectivity among members of a morphological family, that is, *morphological connectivity*, and track the development of this knowledge over time.

We selected derivational family entropy (Moscoso del Prado Martín et al., 2004) as an index of morphological connectivity. Because derivational family entropy takes into account the relative frequencies of members of morphological families, the presence of its effect on written production would entail that ELLs tap into information within the network of a morphological family in addition to the target-derived word itself. Entropy is lower, that is, there is less uncertainty in a morphological paradigm, when the frequency of occurrence of one family member dramatically exceeds the frequencies of all other members. On the other hand, entropy is greater when the frequencies of family members are more uniformly distributed. Based on the pattern of effects of prior L1 visual word processing studies outlined earlier (e.g., Fruchter & Marantz, 2015; Moscoso del Prado Martín et al., 2004; Schmidtke et al., 2017), we expect target forms with greater derivational family entropy to elicit more production errors as a result of a relative lack of precision in morphological knowledge for these items. We predict that it will be more difficult for learners to pinpoint the correct derived word when that word belongs to a more diffuse morphological family, leading to a higher likelihood of written production error. Any effect of derivational family entropy would be evidence of sensitivity to implicit knowledge of relations among morphological families, that is, morphological connectivity.

Critically, we expect to find a weaker negative effect of derivational family entropy earlier in language development compared to later in language development. We predict the effect of entropy to be more acute later in language development as a result of continued exposure to morphologically complex words in the language. Our reasoning is based on the results of lexical processing reviewed earlier which point to an increased sensitivity to the statistical properties of morphologically complex words as a function of language experience (e.g., Falkauskas & Kuperman, 2015; Ulicheva et al., 2020). During language learning, many newly acquired complex words will belong to the same morphological family as complex words that already exist in lexical memory, thereby expanding the size of a morphological family. We predict that as morphological families expand in the lexicon, through exposure to language, so does sensitivity to relative probabilities of members within those morphological families, that is, an increase in morphological connectivity. We expect that this increase in sensitivity will be expressed as a stronger negative effect of entropy on written word production. The stronger negative effect later in language learning might be visible as a larger penalty in accuracy for high entropy words or as a larger gain in accuracy for low entropy words.

Although our hypothesis advocates a causal link between exposure to language and precision in producing the correct derived form within a morphological family, we cannot rule out an opposite effect arising from an increased ability to discriminate between members of a morphological family (Baayen et al., 2011)[1].

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

For instance, instead of observing a stronger effect of entropy as a consequence of language exposure, it is possible that the effect becomes weaker over time as learners are able to better handle diffusion in a high entropy morphological family. This would entail that the greatest gains in written production would be found for words belonging to high entropy families. It is also possible for both effects to occur simultaneously, whereby the effect of an increase in discrimination skill masks the gradual strengthening of the entropy effect as a result of language experience.

The present study

In the present study, we examined the morphological entrenchment and morphological connectivity hypotheses. These hypotheses were tested in a within-participant longitudinal study of the written production of derived words in English. The participant cohort was a sample of adult ELLs enrolled in an 8-month university-level English bridging program at a Canadian university. Participants were tested at two timepoints: at the beginning (t_1) and end (t_2) of the bridging program. The bridging program affords a unique opportunity to examine intra-individual change in morphological knowledge for two reasons. First, the duration of the program is fixed, which provides a controlled window of time across which we are able to examine change in sensitivity to morphological knowledge. Second, we are able to examine change in a homogeneous sample of language learners. The sample represents a population of Chinese adult ELLs who are aged 18–24 years, have L1 Cantonese or Mandarin, and are all at approximately the same level of English language proficiency.

Method

Design

The design was longitudinal with two testing timepoints (t_1 and t_2). Data for t_1 were collected during the first month of the bridging program and data for t_2 were collected during the final month of the program.

Participants

A total of 196 consenting participants (78 female, 113 male, 5 undisclosed) took part in the experiment at t_1 and t_2[2]. The average age of participants at t_1 was 19.88 ($SD = 2.71$). Participants were recruited from the 2019–2020 ($n = 167$) and 2020–2021 ($n = 29$) academic sessions of the McMaster English Language Development program, an 8-month full-time intensive bridging program. The bridging program provides two semesters (over 21 hr per week) of English instruction designed for international students who meet the academic requirements for an undergraduate program but whose overall scores on the Academic version of the International English Language Testing System (IELTS) test do not meet the university's English language proficiency threshold of 6.5. All participants were living in Canada for the duration of the study. Prospective students must obtain a minimum overall IELTS score of 5 to be admitted to the bridging program. IELTS is one of the officially recognized English language proficiency qualifications for Canadian Higher Education institutions and is assessed on a 9-band scale, ranging from non-user (band score 1) to expert (band score 9), with a band score of 6 equivalent to a competent user and a band score of 7 equivalent to a good user[3]. In IELTS terms, participants would be described as "limited" to "modest" users of English upon entry to the program (median overall IELTS score = 5.5), and as "modest" to "competent" language users at program completion (i.e., approximately an IELTS score of 6.5).

Language instruction

The program includes both exposures to advanced academic texts and some explicit morphological instruction. Students are exposed to advanced (university-level) readings spanning multiple academic disciplines (e.g., economics, engineering, health sciences) over the entire program, and across all courses in the program (ten), through core language teaching materials that include academic content and related activities, chapters of undergraduate textbooks not intended for a second language audience, and through assignments that require the reading of journal articles (e.g., a short literature review). Early in the program, in a course (one of five courses in first term) that focuses on the development of effective reading strategies, there is some explicit morphological instruction. Students are taught to analyze multimorphemic words into constituent roots and affixes (including Latinate units), e.g., *dis* + *mis* + *ive*. Although students practise combining roots and affixes (e.g., by identifying additional words containing a particular morpheme, such as *act* in *transaction* and *react*), the focus is on morphological decomposition as a strategy for understanding unfamiliar words encountered when reading. Importantly, the selection of complex words used in teaching materials was conducted independently and without knowledge of the present experiment and its hypotheses. None of the materials in teaching were selected based on the frequency-based properties examined in the current study.

Test of morphological structure

The test of morphological structure was an adapted version of the 28-item derivation task from Carlisle (2000). For each trial of the task, participants were given a base word (e.g., *teach*) and an incomplete sentence (e.g., *He was a good* ______.). Participants were instructed to modify the base word to appropriately complete the sentence (e.g., *He was a good teacher.*). The position of the target word was the last word of the sentence for every item. Target form suffixes included *-th* (e.g., *warmth*); *-ance* (e.g., *appearance*); *-er* (e.g., *washer*); *-ity* (e.g., *activity*); *-ion* (e.g., *expression*); *-ous* (e.g., *adventurous*); and *-able* (e.g., *profitable*). Further details on the construction of the test are available in Carlisle (2000). Since the task was administered at two time-points in the present study, we split the test into two lists of 14 items each (List A and List B). When possible, suffixes were split evenly across both lists. Tables A.1 and A.2 in Appendix A contain the test items for lists A and B, respectively.

Procedure

Participants were administered each test individually in English. For the 2019–2020 cohort, at t_1 participants produced answers using a pen and paper. At t_2 participants

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

were given a digital version of the test where responses were typed into a computer[4]. All 2020–2021 participants completed the task online at both timepoints. All test items were presented on the same page/screen at both timepoints. At t_1, half of the participants were administered List A and the remaining half were administered List B. Lists were counterbalanced across testing timepoints such that those who were administered List A at t_1 were administered List B at t_2, and vice-versa. Trials for the pen and paper version of the task at t_1 were digitized and cross-checked by two research assistants. Participation in the study was voluntary and participants received course credit for their participation (course credit was awarded for completion of the test and was not based on performance on the task). The study was approved by the McMaster University Research Ethics Board Ethics Board (protocol 2018-0239).

Scoring morphological production accuracy

We used the POMplexity spelling error classification system (Bahr et al., 2020; Bahr et al., 2020; Benson-Goldberg, 2014; Quick & Erickson, 2018) as a framework for categorizing the accuracy of target word spellings. The POMplexity system codes spelling errors across three different linguistic categories: Phonological (P), Orthographic (O), and Morphological (M). We used the morphological POMplexity criteria described in Quick and Erickson (2018) to evaluate the spelling accuracy of the responses. The responses were annotated in such a way that a spelling that was identical to the target-derived word was classed as “correct”, e.g., *remarkable* for the target form *remarkable*. A “plausible” spelling contained errors that indicate awareness of the target morpheme, but contain a grapheme transposition (e.g., *reasonalbe* for *reasonable*), omission (e.g., *swimer* for *swimmer*), addition (e.g., *humanitty* for *humanity*), or other obvious typing errors (e.g., *sborption* for *absorption*). Another type of spelling error included “homophone substitution” (e.g., *appearence* for *appearance*, *gloryous* for *glorious* and *mysteries* for *mysterious*). Spelling errors for which an incorrect suffix (correctly spelled or not) was combined with the target stem were annotated as “morpheme substitution” errors (e.g., *gloryful* for *glorious* and *humanable* for *humanity*). Spelling errors wherein an illegal orthographic sequence was produced for a suffix or a stem were classified as “nonmorpheme substitution” errors (e.g., *humanalixl* for *humanity*). Finally, responses that did not show an attempt at providing a suffix (e.g., *do not know*), did not include a suffix (e.g., *swim* for *swimmer*) or contained the wrong stem (e.g., *master* for *mystery*), were annotated as “omission errors”. Our research questions pertain to L2 knowledge of English derivational morphology and not knowledge of English spelling conventions. Therefore, “correct” forms, and “plausible” and “homophone substitution” errors were categorized as correct responses (1), and the remaining responses were categorized as incorrect (0). Appendix B provides a plotted tabulation of all spelling types across timepoints.

Reliability

We assessed parallel forms reliability and the internal consistency of the test of morphological structure. Parallel forms reliability was examined by comparing the differences in means in accuracy for List A and List B at each timepoint

(Gulliksen, 1950). We also examined the impact of the switch from pen-and-paper to computer input partway through the testing of the 2019–2020 cohort (see Procedure); it is possible that the switch to typing artificially inflated results. If this was the case then growth between timepoints would be disproportionately larger for the 2019–2020 cohort, that is, the cohort that switched instruments between timepoints, compared to growth for the 2020–2021 cohort, that is, the cohort that typed results at both timepoints. A generalized linear mixed-effects model with an underlying binomial distribution was fitted to accuracy scores ($0 =$ incorrect; $1 =$ correct). List (List A vs. List B), Timepoint (t1 vs. t2), and Cohort (2019–2020 vs. 2020–2021) were included as fixed effects, including the interactions between Timepoint and List and Timepoint and Cohort. Random intercepts were included for participants and items. The Timepoint $\times$ List interaction term was not statistically significant [$\hat{\beta} = 0.02$; $SE = 0.28$; $z = 0.07$; $p = .94$], indicating that the difference in average accuracy between lists was equivalent at t_1 (List A; $M = 57.43\%$, List B; $M = 62.81\%$) and at t_2 (List A; $M = 64.58\%$, List B; $M = 71.65\%$). A main effect of List was also not significant in a model that did not include a Timepoint $\times$ List interaction [$\hat{\beta} = 0.41$; $SE = 0.45$; $z = 0.9$; $p = .37$]. We therefore concluded that List A and List B produced similar results at each timepoint. In addition, the Timepoint $\times$ Cohort interaction was not statistically significant [$\hat{\beta} = -0.09$; $SE = 0.27$; $z = -0.33$; $p = .74$]. This effect indicates that the amount of between-timepoint change was the same for the 2019–2020 cohort ($\Delta = 7.88\%$) and 2020–2021 cohort ($\Delta = 8.08\%$), even when there was a switch to typing from pen-and-paper partway between 2019–2020 testing sessions. We therefore concluded that the switch to typing did not artificially inflate responses. Cronbach's alpha reliability coefficient was 0.76 for Form A and 0.66 for Form B.

Independent variables

Suffix productivity and stem frequency

We obtained estimates of frequency of occurrence for stem words from the 51 million-token SUBTLEX-US corpus of US film and media (Brysbaert & New, 2009). We refer to the frequency of the stem word as Stem frequency.

Suffix productivity is formally defined as the number of word types that share a suffix with the target word. For example, the measure takes into account the number of times *-ity* appears as a suffix across all unique derived word forms, such as *activity*, *density*, *majority*, in English. We retrieved this information from the English Lexicon Project (ELP: Balota et al., 2007), which provides morphological information for 79,672 English words. We ensured that we counted all suffixes in complex words using the morphological parse provided in the MorphSp column of the ELP database. We did not include derived words containing bound stems (e.g., *populous*) in our suffix productivity counts. We also discounted plural and/or possessive forms of derived words, therefore ensuring that words such as *swimmers* and *swimmer's* were not included as an extra count for *swimmer*. Furthermore, we manually coded a small number of derived words with the *-th* suffix that did not appear in the database as morphologically complex.

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

Derivational family entropy

We computed derivational family entropy for the target-derived words by estimating frequencies of words that also share the target's stem form (e.g., *humanistic* and *humanize* for the target derived word *humanity*). Derivational entropy was defined as Shannon entropy calculated over the probability distribution p of the derived word's morphological family (obtained by dividing the frequencies of family members by the cumulative family frequency): $H = -\Sigma log_2(p) * p$.

Additional lexical characteristics

As control variables, we included the frequency of the target word form taken from SUBTLEX-US, which we refer to as Derived word frequency. We also considered word length (measured in number of characters) and orthographic neighbourhood density. The orthographic neighborhood of a word was estimated using the average Levenshtein distance (OLD20), which is defined as the average orthographic distance from the 20 nearest orthographic neighbors. OLD20 was estimated for each target word in our stimulus list using the library vwr (Keuleers, 2013) in the R statistical computing software program (Version 4.1.2; R Core Team, 2021). We used all unique words in the SUBTLEX-US corpus as the source lexicon with which to estimate orthographic neighborhood density. A series of independent two-sample Mann–Whitney U-tests were conducted to examine the differences in each lexical characteristic (critical variables and controls) across stimulus Lists A and B. All *p*s were $> .67$, indicating that stimuli lists were reasonably matched for each lexical variable. We also considered the effect of words that do not undergo a phonological shift when a suffix is added (e.g., *teacher*) versus those derived words whose forms do undergo such a shift (e.g., *discussion*). We do not discuss this measure further because it neither reached statistical significance nor improved model fit.

The distributional characteristics of all lexical predictor variables are reported in Table 1. Correlations between lexical statistics are presented in Table 2. The full list of stimuli with lexical statistics is provided in Appendix C. The materials are also hosted on the Open Science Framework at the following link: https://osf.io/8egz7/[5].

Prior L2 experience

We also included age of initial English language instruction as a measure of prior of English language experience. The average age of initial English language instruction was 8.64 years ($SD = 4.58$).

Statistical considerations

Model fitting procedure

We applied an explanatory item response analysis to the longitudinal data set (De Boeck & Wilson, 2004). We fitted a generalized linear mixed-effects multiple regression model (Baayen et al., 2008; Jaeger, 2008; Pinheiro & Bates, 2000) to morphological production accuracy, which has a binomial underlying distribution (0 = incorrect; 1 = correct). We used restricted maximum likelihood estimations.

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

Table 1. Descriptive statistics for the independent variables. Reported are the range, mean, and standard deviations of the original and transformed variables after selection and trimming procedures

Independent variable	Original			Transformed		
	Range	*M*	*SD*	Range	*M*	*SD*
Stem frequency	44:34433	3,474.11	6,800.52	−2.25:2.26	0	1
Derived frequency	18:2842	591.64	611.43	−2.33:1.72	0	1
Suffix productivity	45:2163	651.68	631.26	−1.67:1.42	0	1
Derivational family entropy	0:2.68	1.07	0.67	−1.58:2.4	0	1
Length	5:11	8.57	1.73	−2.06:1.4	0	1
OLD20	1.35:3.35	2.39	0.51	−2.05:1.89	0	1

Table 2. Correlation matrix of lexical variables

	1.	2.	3.	4.	5.
1. Stem frequency (log)					
2. Derived word frequency (log)	0.51**				
3. Suffix productivity (log)	−0.35	−0.17			
4. Derivational family entropy	−0.02	−0.17	−0.18		
5. Word length	−0.48**	0.05	0.21	−0.07	
6. Orthographic neighbourhood	−0.35	0.03	−0.20	0.04	0.82***

Lower triangle provides Pearson correlation coefficients.
*Correlation is significant at the .05 level.

The fixed effects structure of the model included interactions between Timepoint and each of the critical lexical variables of interest: Stem frequency, Suffix productivity, and Derivational family entropy. Any interactions that were not statistically significant were removed from the model. The removal of nonsignificant interactions did not change the direction of any of the remaining effects, nor did it change the significance of any effect at the .05 alpha level. All frequency-based measures (including suffix productivity) were log-transformed before they were entered into models. Timepoint was included in models as a factor using contrast coding ($t_1 = -1$; $t_2 = 1$). We included Derived word frequency, length (in characters) and OLD20 of the target word as control variables. All continuous independent variables were z-transformed in order to ensure the interpretable comparison of effects across variables with different scales. We also controlled for possible list counterbalancing artefacts by including a fixed effect of list order (two levels: AB and BA).

We also included random intercepts for Participant with random slopes for Timepoint. By-participant random slopes for Timepoint account for individual differences in longitudinal change. A property of the structure of the data is that target words are nested under suffix type (the same suffixes are used with multiple stems). We therefore fitted a model with random intercepts for Item nested under

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

Suffix. Using AIC values as a diagnostic of model fit (AIC values closer to negative infinity indicate greater goodness of fit), we compared the nested random effects structure to a model without random intercepts for Suffix. Chi-square goodness-of-fit tests compared AIC values across models and confirmed that the addition of random intercepts for Suffix as a nested effect with Item did not provide any gains in goodness-of-fit[6]. Therefore, the final model included (i) random intercepts for Item and (ii) random intercepts for Participant with random slopes for Timepoint.

Software

All data analysis was performed in the R statistical environment (Version 4.1.2; R Core Team, 2021). The generalized linear mixed-effects multiple regression model was fitted using the lme4 package (using the BOBYQA algorithm for optimization, Bates et al., 2015). We obtained p-values for lmer model fits using the lmerTest package (Kuznetsova et al., 2017), which uses Satterthwaite's degrees of freedom method. Contrast coding was implemented using the car package (Fox & Weisberg, 2019). The significance of the fixed effects from the generalized linear-mixed effects model was estimated with Type-II Wald tests using the Anova() function also provided by the car package. Effect estimates were extracted from the model using the effects package (Fox & Weisberg, 2019). All plots were generated using the ggplot2 package (Wickham, 2016). For the critical variables of interest, where significant interaction effects were observed with Timepoint, we examined the significance of the simple slopes of t_1 and t_2 using the emtrends function in the emmeans package (Lenth, 2021). A simple slope was considered significant if the estimated 95% confidence interval contained zero.

Model criticism

We assessed the level of harmful collinearity in our final model in two ways. First, in the spirit of Tomaschek et al. (2018), we estimated the variance inflation factor (VIF) for each fixed effect in R using the vif() function in the car package. A VIF score of > 10 for a fixed effect indicates that its model coefficient is poorly estimated due to overfitting (Kutner et al., 2005). VIF scores for all effects were $<$ 6.5; critical effects of interest were all < 1.5. Second, we computed the condition number (κ) using the languageR package in R (Baayen & Shafaei-Bajestan, 2019). $\kappa = 5$, indicating that the level of multicollinearity between all predictors is below the harmful level of 30 (Belsley et al., 2005). We are therefore satisfied that overfitting due to multicollinearity was not an issue in our model. We refitted the model after removing outliers, which were defined as data points with absolute standardized residuals exceeding 2.5 standard deviations (Baayen & Milin, 2010).

Statistical power

A statistical power analysis was performed for sample size estimation (see Brysbaert, 2021 for an overview of statistical power issues in bilingualism research). We conducted a power simulation study based on the fixed and random effects of the generalized mixed-effects model fitted to data from the initial study cohort (Brysbaert & Stevens, 2018; Green & MacLeod, 2016). We used the simr R package (Green & MacLeod, 2016) to estimate the sample size required to detect effect sizes

Table 3. Mean morphological production accuracy (%) to words (SEs in parentheses) across tertiles of each critical lexical variable

Lexical variable	Timepoint	Lower tertile	Middle tertile	Upper tertile
Stem frequency	t_1	54.55 (1.65)	60.15 (1.62)	65.54 (1.57)
	t_2	58.23 (1.63)	73.75 (1.45)	72.5 (1.48)
Suffix productivity	t_1	53.63 (1.64)	55.56 (1.65)	71.16 (1.5)
	t_2	65.82 (1.58)	66.63 (1.56)	71.97 (1.48)
Derived word frequency	t_1	51.25 (1.65)	53.36 (1.66)	75.54 (1.42)
	t_2	60.79 (1.62)	66.05 (1.56)	77.67 (1.38)
Derivational family entropy	t_1	73.22 (1.46)	60.51 (1.62)	46.63 (1.65)
	t_2	80.87 (1.3)	65.54 (1.57)	57.98 (1.64)

of $d = .4, .2, .1$ and .05 with power set to 80% and higher. We used the extend() and powerCurve() functions to perform Monte Carlo simulations for each effect size. We obtained power estimates for each main effect and interaction effect of interest for sample sizes ranging from 100 to 500 in increments of 25. We retained all the default settings of the powerCurve() function and set the analysis to conduct 100 iterations per simulation. The results indicated that a minimum of 125 participants are required to detect the smallest effect size, $d = .05$, with > 80% power (for the interaction between timepoint and any of the critical lexical variables). To ensure we were able to remain sufficiently overpowered, we collected additional data from a subsequent cohort of participants, setting the target sample size at 190 participants.

Results

The raw data set included 5,488 trials ($t_1 = 2{,}744$; $t_2 = 2{,}744$), where a trial is defined as a participant's response to an individual target item on the test of morphological structure. Table 3 provides descriptive statistics of the results, divided into tertiles for each lexical measure per each timepoint. Table 4 provides a summary of an analysis of variance for the final model. The fixed and random effects of the final model are provided in Appendix D.

Entrenchment effects

There was a statistically significant main effect for Timepoint [$\hat{\beta} = 0.249$; $SE = 0.048$; $z = 5.141$; $p < .001$], confirming that there was a reliable average within-participant increase in morphological production accuracy, $\Delta = 8\%$ ($Mt_1 = 60\%$; $Mt_2 = 68\%$). There was a statistically significant interaction between Timepoint and Suffix productivity [$\hat{\beta} = -0.128$; $SE = 0.037$; $z = -3.445$; $p = .003$]. There was also significant interaction between Timepoint and Derived word frequency [$\hat{\beta} = -0.073$; $SE = 0.037$; $z = -1.983$; $p = .047$]. The interaction between Timepoint and Stem frequency was not significant [$\hat{\beta} = -0.012$; $SE = 0.04$; $z = -0.334$; $p = .738$]. This interaction was removed from the final model. There was also a nonsignificant main effect of Stem frequency [$\hat{\beta} = 0.319$; $SE = 0.214$;

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

Table 4. Analysis of variance table for the fixed effects of the generalized linear mixed-effects model fitted to production accuracy

Fixed effects	$\chi^2(df)$	p
Timepoint	31.263 (1)	.000
Suffix productivity (log)	3.796 (1)	.051
Derived word frequency (log)	7.544 (1)	.006
Derivational family entropy	13.372 (1)	.000
Stem frequency (log)	2.229 (1)	.135
Orthographic neighbourhood density	0.255 (1)	.613
Word length	0.181 (1)	.670
Age of initial English instruction	0.007 (1)	.932
List order	0.122 (1)	.727
Timepoint × Suffix productivity (log)	11.867 (1)	.001
Timepoint × Derived word frequency (log)	3.932 (1)	.047

$z = 1.493$; $p = .135$], which indicates that the frequency of the stem did not influence the morphological production accuracy of derived words.

The significant Timepoint × Suffix productivity interaction effect is visualized in Figure 2 (panel a). Tests of the simple slopes indicated that the effect of Suffix productivity was significant at t_1 [$\hat{\beta} = 0.550$; $SE = 0.221$; 95% CI = 0.117 to 0.984; $p < .05$] but not at t_2 [$\hat{\beta} = 0.293$; $SE = 0.222$; 95% CI = −0.141 to 0.117; $p > .05$]. The qualitative pattern of the significant interaction indicates that the greatest exposure-related gains in morphological production accuracy were for target words that contained relatively less productive suffixes. For example, target-derived words consisting of the least productive suffix, -*th*, underwent a predicted 23% change in morphological production accuracy, whereas targets comprising the most productive suffix, -*er*, saw a 2% gain in morphological production accuracy (see Table 3 for a breakdown of raw accuracy for each tertile of Suffix productivity at each timepoint). In short, the overall trend of this interaction aligns with the predicted pattern of the lexical entrenchment account (Diependaele et al., 2013, see Figure 1).

The significant Timepoint × Derived word frequency interaction effect is visualized in Figure 2 (panel b). Tests of the simple slopes indicated that the effect of Derived word frequency was significant at t_1 [$\hat{\beta} = 0.598$; $SE = 0.195$; 95% CI = 0.217 to 0.979; $p < .05$] and at t_2 [$\hat{\beta} = 0.452$; $SE = 0.194$; 95% CI = 0.071 to 0.833; $p < .05$]. As indicated by the simple slopes, the effect of derived word frequency was weaker at t_2, and the qualitative pattern of the interaction converged with the expectations of the lexical entrenchment account. There were larger gains in production accuracy for low frequency derived words compared to high-frequency derived words: there was a 20% gain in production accuracy for the least frequent derived words, and a smaller 3% gain in production accuracy for the most frequently occurring derived words. Interestingly, the size of the entrenchment

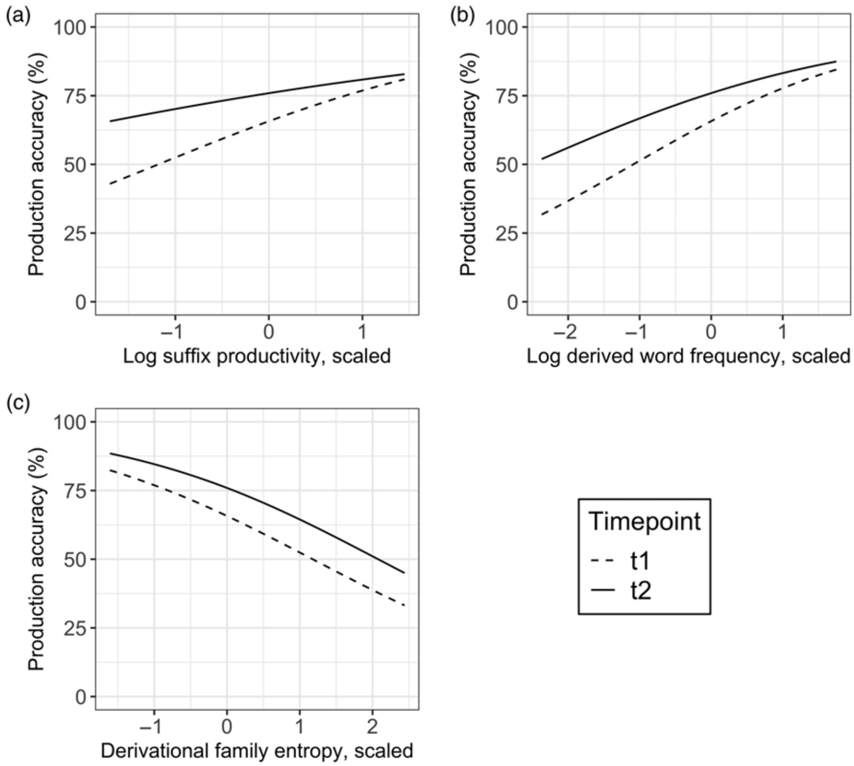

Figure 2. Production accuracy effects. Panel a: partial interaction effects of timepoint by suffix productivity; panel b: partial interaction effects of timepoint by derived word frequency; panel c: partial main effects of derivational family entropy and timepoint.

effect was smaller for derived word frequency than it was for suffix productivity. Intraindividual gains in morphological production accuracy were 12 times larger for derived words containing the least productive suffixes relative to derived words containing the most productive suffixes. This change was less dramatic for derived word frequency: intraindividual gains were seven times larger for the least frequent words relative to the most frequent words. We discussed the difference between the magnitudes of the suffix productivity and derived word frequency entrenchment effects in the General discussion.

Connectivity effects

There was a significant main effect of Derivational family entropy [$\hat{\beta} = -0.554$; $SE = 0.151$; $z = -3.657$; $p < .001$], but we did not find a significant interaction between Timepoint and Derivational family entropy [$\hat{\beta} = -0.056$; $SE = 0.035$; $z = -1.585$; $p = .113$]. The direction of the main effect of Derivational family entropy was negative, indicating that, at both timepoints, morphological production was less accurate for target words that belong to derivational paradigms with many relatively equally frequent members (Figure 2, panel c). The effect size for the main effect of derivational family entropy (estimated for the minimum and maximum

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

entropy values) indicates that morphological production accuracy for target-derived words with low derivational family entropy was 85%, compared to 39% for target-derived words with the highest family entropy. Therefore, the results support the idea that participants were sensitive to the relative frequencies of members within a derivational family, but sensitivity to this source of stored morphological information did not change as a result of the language exposure captured in this study. In sum, this result supports the morphological connectivity hypothesis, indicating that ELLs draw upon distributional information about members of the target word's morphological family during the written production of English derived words.

In addition, we conducted a *post hoc* analysis to examine whether differences in the amount of prior L2 exposure may have been responsible for failing to detect changes in the effect of derivational family entropy over time. A three-way interaction between Age of initial English language instruction (as a proxy of quantity of prior L2 exposure), Derivational family entropy, and Timepoint was conducted and did not produce a significant effect [$\chi^2 = 0.622$; $p = .43$]. This result indicates that the effect of morphological connectivity remained stable regardless of the amount of prior L2 exposure[7].

General discussion

The current study sought to address an outstanding question about the relationship between language experience and the development of morphological knowledge: how does an individual's exposure to the distributional patterns of language contribute to *that same individual's* change in complex word knowledge? This question is critical for understanding whether exposure to language influences a learner's ability to efficiently exploit statistically driven cues to morphological structure during production. We addressed this question by evaluating the extent of exposure-related changes in L2 learners' sensitivity to the distributional properties of complex words as a diagnostic of the development of morphological knowledge.

Morpheme entrenchment

Critically, our study found that the extent of *exposure-related change* in the written production of derived words in ELLs is modulated by suffix productivity. The largest gains in complex word knowledge were found for English words with the least productive suffixes, which is consistent with the predictions of the lexical entrenchment account (Diependaele et al., 2013). In observing that the entrenchment effect operates upon suffix productivity, we argue that exposure to language specifically reinforces knowledge of the morphological elements of words. Taken together, this result builds upon past research (e.g., Bertram et al. 2000; Cop et al., 2015; Whitford & Titone, 2012) by providing within-participant developmental evidence that the mechanism of lexical entrenchment occurs at the morpheme level, that is, *morpheme entrenchment*.

The results indicated that the morpheme entrenchment effect was specific only to suffixes and not stems. The interaction between stem frequency and timepoint, and the main effect of stem frequency, were not statistically significant. It may be possible to conclude from this finding that continued exposure to language leads

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

only to changes in the production of suffixes, and that gradual entrenchment of stem units in the lexicon is absent. It is also possible, though, that the absence of any effect of stem frequency could be an artifact of the morphological production task itself. Participants were already provided with the stem and were asked to modify it to complete the sentence, which leads us to speculate that this task relies more heavily on compositional processing associated with suffix representations and not stems. To summarize, it is possible that the representations of low-frequency stems were reinforced over time, but that participants did not tap into this information given the demands of the task.

Derived word entrenchment

We found a significant interaction between derived word frequency and timepoint. The pattern of the interaction effect was consistent with the predictions of the lexical entrenchment account. For the present sample of ELLs, full forms of complex words, particularly low frequency words, were more accurately produced by the end of the bridging program. Thus, language exposure was associated with the reinforcement of the full forms of low frequency derived words in ELLs. Based on the negative and weak correlation between suffix productivity and derived word frequency ($r = -.17$, $p > .05$; see Table 2), it is not likely that the interaction between suffix productivity and timepoint could be explained by the interaction between derived word frequency and timepoint, and vice versa (see also low VIF values for these effects reported in Statistical considerations).

The test of individual slopes indicated that derived word frequency was a significant predictor of production accuracy at t_1 and t_2, but that suffix productivity was only a reliable predictor of accuracy at t_1 (see Results). The nonsignificant effect of suffix productivity at t_2 indicate that as suffix forms become more entrenched (see discussion above), derived words with suffixes that are less productive become easier to correctly produce, so much so that these derived words are no longer a problem for written word production. However, whole word frequency remains a critical predictor: low-frequency whole word forms may still negatively impact written word production accuracy relative to high frequency derived words. What might be a reason for the difference in magnitude of the entrenchment effect for suffix productivity compared to derived word frequency? Recall that the central claim of the entrenchment account is that exposure to words drives the consolidation of words in memory, and that language users have fewer opportunities to strengthen the lexical representations of low-frequency items because they are less prevalent in the language. Thus, it is possible that participants did not encounter the lowest frequency derived words as often as the least productive suffixes within the duration of the study, leading to a smaller entrenchment effect for low frequency-derived words compared to the least productive suffixes. The frequency data of our materials support this view (see Table 1): derived words, on average, occur less often than isolated suffix forms.

Morphological connectivity

ELLs in this study were affected by the distributional properties of morphological families but did not show evidence of exposure-driven change in the strength of this

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

effect. The effect of derivational family entropy was negative and remained stable at both timepoints. That is, at both t_1 and t_2, morphological production accuracy was highest for morphological families with lower derivational entropy, that is, derived words that belong to families with fewer equally dominant members. Since the measure of derivational family entropy relies on the relative frequencies of words within a given morphological paradigm, the presence of an entropy effect necessarily implies that ELLs in this study exploited lexical knowledge about morphological families during derived word production. We interpret this as a stable effect of *morphological connectivity*, that is, sensitivity to inter-lexical connectivity between the members of a morphological family. Previous L2 research (Mulder et al., 2014) has shown that morphological family size influences language processing. Here, we show that L2 lexical knowledge comprises both morphological family size *and* the relative frequencies of members of a morphological family.

The main effects of timepoint and derivational family entropy indicate that instead of observing a bias toward longitudinal gains for either low or high entropy words, there was an overall improvement across the entropy continuum. As noted in Introduction, it is also a possibility that greater morphological connectivity was masked by an increase in individual discrimination ability. That is, a steeper negative effect (stemming from larger gains for low entropy words or a higher penalty for high entropy derived words) was counteracted upon by an opposite effect: an improvement in the ability to select the correct form within a derivational family. Future research should be devoted to more clearly disentangling both possibilities in ELLs.

Insights from a recent study may also shed light on why the effect of morphological connectivity remained stable across the program. In an investigation of English morphological awareness among adult ELLs registered in US universities whose L1 was either Turkish or Chinese, Wu and Juffs (2021) uncovered a persistent effect of L1 morphological experience on the development of morphological awareness in English. Despite advanced proficiency in English for both groups, L1 Turkish ELLs outperformed L1 Chinese ELLs in tasks tapping into derivation and morphological relatedness, suggesting that the effects of an L1 experience with an agglutinative language such as Turkish facilitated morphological awareness in English. Moreover, the study suggests that ELLs with an L1 rich in inflection and derivation can develop more acute English morphological awareness even with an overall lower English proficiency than ELLs whose L1 is an isolating language such as Chinese. For our participants, who are also L1 Chinese ELLs, and whose English proficiency is lower than that of Wu and Juffs' (2021) learners, the L2 experience afforded by the program regarding morphological connectivity is likely insufficient to act upon the persistent influence of L1 experience with an isolating morphological system.

Taken together, the morpheme entrenchment and connectivity findings accord with prior work on L2 morphological processing (Diependaele et al., 2011; Mulder et al., 2015) by showing that both L1 and L2 lexical knowledge comprises the same types of distributional information about morphological structure. Furthermore, as noted in the Introduction, the hypotheses put forward in the present study were motivated by the idea that the mental structuring of language reflects the distributional properties of the input acquired through experience (MacDonald, 2013;

Stanovich & West, 1989). Our results align with this premise: even at the outset of the longitudinal study (t_1) ELLs were sensitive to the distributional properties of words as encoded in the orthography.

Relationship to prior research

Much of the existing research on L2 morphological awareness is correlational and has been conducted on bilingual children who learn English as a second language (see the meta-analysis in Ke et al., 2021); few studies have examined morphological awareness in adult L2 learners of English (Zhang & Koda, 2012). The goal of these studies is to examine whether individual variability in morphological awareness translates into interindividual differences in reading comprehension and vocabulary knowledge. The present study contributes to this body of research in the following three ways. First, our analysis focuses on the properties of the items in the morphological awareness task. We examined sources of variability of within-participant change in morphological knowledge that stem from the statistical-distributional properties of words and the morphological families they belong to. In this regard, our findings supplement existing longitudinal ELL research by Jiang and Kuo (2019) which shows how within-participant change in morphological production varies across complex words with different phonological properties and parts of speech. Second, our focus on established frequency-based markers of morphological knowledge, coupled with a longitudinal design, allowed us to link research on the development of L2 morphological knowledge to existing psycholinguistic theories of word learning and morphological processing. Third, the focus on university-level Chinese ELLs enabled us to contribute to research on bilingualism and the development of morphological knowledge from an important dimension that is less prominent in the literature: the lens of the adult learner in a learning environment (post-secondary) who needs to engage in a sophisticated manner with an additional language that is typologically distant from their L1.

Practical implications

According to a report by the Institute of International Education (Mason, 2021), based on UNESCO data, China was the leading place of origin for international students for the 11^{th} consecutive year in 2020: almost one million students (993,367) from China were registered in higher education institutions outside China in 2020, the vast majority in the USA, Australia, the UK, and Canada. Our findings therefore have implications for the educational practice of Chinese ELLs, especially those enrolled in bridging programs. Bridging programs are established with the aim of "bridging" lower proficiency students into post-secondary programs. However, at present, the number of studies that quantify the extent of linguistic gains of students enrolled in bridging programs is limited. Based on our results, it would be worthwhile for developers of advanced English as a second language materials and texts to take into account the need for learners to repeatedly encounter a rich variety of low-frequency affixes. Because derived words with less productive suffixes (e.g., *-th*, *-ence*/*-ance*, and *-ous*) are challenging to correctly produce early on, students should be exposed to materials containing an overrepresentation of derived words with these morpheme characteristics. Moreover, it is worth highlighting that, as with derived word frequency, derivational family entropy was a reliable predictor of production accuracy at t_1. This indicates that interlexical connections between members of a morphological family are present for ELLs even at the outset of the bridging program. In our population, the median overall incoming IELTS score was 5.5 (range = 5 to 7), suggesting that knowledge of derivational families is already established among ELLs with "limited to modest" knowledge of English (to use IELTS terminology). Therefore, ESL curriculum developers may wish to use this information when designing materials aimed at ELLs who fall within a similar IELTS proficiency band. For example, special attention could be directed at the creation of texts and exercises containing derived words belonging to morphological families with multiple similarly frequent members.

Additional work is needed to examine change in morphological knowledge as a consequence of language exposure. The current task was designed to examine morphological knowledge using written word production as an outcome measure. This task provides an indication of the accuracy of morphological knowledge, by assessing whether a correct suffix was applied to a stem. This method is limited, however, because it cannot capture cognitive processing effort during the task. Having established the viability of assessing the development of morphological knowledge in a relatively simple-to-administer production task, we can begin to investigate how language exposure determines speed of processing during written production. For example, future research could employ the same task and longitudinal design yet examine keystroke latencies as an additional measure (e.g., Feldman et al., 2019; Sahel et al., 2008). Furthermore, we wish to note that during the course of the present study, all ELLs were provided with explicit morphological instruction in addition to exposure to advanced academic texts. Although we are confident that responsiveness to word frequency information would stem from exposure to language, it is also possible that part of this knowledge results from explicit instruction. A clear distinction would be achieved in a controlled randomized intervention study with two treatment groups: explicit morphological instruction versus no explicit morphological instruction.

Conclusions

Our study shows that it is possible to capture change in L2 learners' sensitivity to morphological structure with increased language exposure across a fixed window of time. Through a written task, we discovered that written L2-derived word production is driven by (i) gains in knowledge of the distribution of suffixes; and (ii) knowledge of connections between words within a derivational morphological family. Derived word production among ELLs can be explained as a function of their learning and exploitation of the distributional properties of the morphological structure of words.

Notes

1. We thank an anonymous reviewer for this suggestion.
2. 198 consenting participants took part at t_1 but 2 participants did not return at t_2.
3. Most English-medium Canadian universities, including McMaster University, require a minimum overall IELTS score of 6.5 (with no IELTS sub-test below 6) for direct entry to undergraduate degree programs.

4. The medium of the experiment was different across timepoints in 2019–2020 because the university switched to online delivery during the COVID-19 pandemic.
5. When using variables that were obtained from other sources, please cite the original source.
6. The change of the random effect structure from the baseline model did not change the direction of any effect, nor did it make a *p*-value of any critical effect change its significance at the .05 level.
7. For completeness, we also report the non-significant three-way interactions for Age of initial English instruction, Suffix productivity and Timepoint [$\chi^2 = 0.044$; $p = 0.83$], and Age of initial English instruction, Derived word frequency and Timepoint [$\chi^2 = 0.005$; $p = 0.94$].

References

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, **3**, 232.
Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects? *Journal of Memory and Language*, **25**(6), 726–740.
Baayen, H., & Schreuder, R. (1999). War and peace: Morphemes and full forms in a noninteractive activation parallel dual-route model. *Brain and Language*, **68**(1–2), 27–32.
Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, **59**(4), 390–412.
Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, **3**(2), 12–28.
Baayen, R. H., Milin, P., Durđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naïve discriminative learning. *Psychological Review*, **118**(3), 438–481.
Baayen, R. H., & Shafaei-Bajestan, E. (2019). languager: Analyzing linguistic data: A practical introduction to statistics [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=languageR (R package version 1.5.0).
Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, **2**(3), 419–463.
Badecker, W., Hillis, A., & Caramazza, A. (1990). Lexical morphology and its role in the writing process: Evidence from a case of acquired dysgraphia. *Cognition*, **35**(3), 205–243.
Bahr, R. H., Lebby, S., & Wilkinson, L. C. (2020). Spelling error analysis of written summaries in an academic register by students with specific learning disabilities: phonological, orthographic, and morphological influences. *Reading and Writing*, **33**(1), 121–142.
Bahr, R. H., Silliman, E. R., & Berninger, V. W. (2020). Derivational morphology bridges phonology and orthography: Insights into the development of word-specific spellings by superior, average, and poor spellers. *Language, Speech, and Hearing Services in Schools*, **51** (3), 640–654.
Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, **39**(3), 445–459.
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects modelsusing lme4. *Journal of Statistical Software*, **67**(1), 1–48.
Beauvillain, C. (1996). The integration of morphological and whole-word form information during eye fixations on prefixed and suffixed words. *Journal of Memory and Language*, **35**(6), 801–820.
Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity* (Vol. **571**). John Wiley & Sons.
Benson-Goldberg, S. (2014). *Spelling of derivationally complex words: The role of phonological, orthographic, and morphological features (Unpublished master's thesis)*. University of South Florida.
Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, **42**(3), 390–405.
Bertram, R., Laine, M., & Karvinen, K. (1999). The interplay of word formation type, affixal homonymy, and productivity in lexical processing: Evidence from a morphologically rich language. *Journal of Psycholinguistic Research*, **28**(3), 213–226.
Bertram, R., Laine, M., & Virkkala, M. M. (2000). The role of derivational morphology in vocabulary acquisition: Get by with a little help from my morpheme friends. *Scandinavian Journal of Psychology*, **41**(4), 287–296.

Bertram, R., Schreuder, R., & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**(2), 489–511.

Bertram, R., Tonnessen, F. E., Stromqvist, S., Hyona, J., & Niemi, P. (2015). Cascaded processing in written compound word production. *Frontiers in Human Neuroscience*, **9**, 207.

Beyersmann, E., Castles, A., & Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *Quarterly Journal of Experimental Psychology*, **65**(7), 1306–1326.

Beyersmann, E., Grainger, J., Casalis, S., & Ziegler, J. C. (2015). Effects of reading proficiency on embedded stem priming in primary school children. *Journal of Experimental Child Psychology*, **139**, 115–126.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? a tutorial of power analysis with reference tables. *Journal of Cognition*, **2**(1), 16. https://doi.org/10.5334/joc.72

Brysbaert, M. (2021). Power considerations in bilingualism research: Time to step up our game. *Bilingualism: Language and Cognition*, **24**(5), 813–818.

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, **3**, 530–548.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, **27**(1), 45–50.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, **41**(4), 977–990.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, **1**(1), 9. https://doi.org/10.5334/joc.10

Burani, C., & Caramazza, A. (1987). Representation and processing of derived words. *Language and Cognitive Processes*, **2**(3–4), 217–227.

Burani, C., Marcolini, S., & Stella, G. (2002). How early does morpholexical reading develop in readers of a shallow orthography? *Brain and Language*, **81** (1–3), 568–586.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, **28**(3), 297–332.

Carlisle, J. F. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders. *Applied Psycholinguistics*, **9**(3), 247–266.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, **12**(3), 169–190.

Caselli, N. K., Caselli, M. K., & Cohen-Goldberg, A. M. (2016). Inflected words in production: Evidence for a morphologically rich lexicon. *Quarterly Journal of Experimental Psychology*, **69**(3), 432–454.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, **22**(5), 1216–1234.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer Science & Business Media.

De Jong, N. H., Schreuder, R., & Harald Baayen, R. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, **15**(4–5), 329–365.

Deacon, S. H., Conrad, N., & Pacton, S. (2008). A statistical learning perspective on children's learning about graphotactic and morphological regularities in spelling. *Canadian Psychology/Psychologie Canadienne*, **49**(2), 118–124.

Diependaele, K., Dunabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, **64**(4), 344–358.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first-and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, **66**(5), 843–863.

Dijkstra, T., Moscoso del Prado Martín, F., Schulpen, B., Schreuder, R., & Harald Baayen, R. (2005). A roommate in cream: Morphological family size effects on interlingual homograph recognition. *Language and Cognitive Processes*, **20** (1–2), 7–41.

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **33**(4), 663–679.

Falkauskas, K., & Kuperman, V. (2015). When experience meets language statistics: Individual variability in processing English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **41**(6), 1607–1627.

Feldman, L. B., Dale, R., & van Rij, J. (2019). Lexical and frequency effects on keystroke timing: Challenges to a lexical search account from a type-to-copy task. *Frontiers in Communication*, **4,** 17.

Feldman, L. B., & Kroll, J. F. (2020). Learning and using morphology and morphosyntax in a second language. *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.604

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.

Fruchter, J., & Marantz, A. (2015). Decomposition, lookup, and recombination: MEG evidence for the full decomposition model of complex visual word recognition. *Brain and Language*, **143,** 81–96.

Gagné, C. L., & Spalding, T. L. (2016). Effects of morphology and semantic transparency on typing latencies in English compound and pseudocompound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **42**(9), 1489–1495.

Goodwin, A. P., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of reading*, **17**(4), 257–285.

Grainger, J., & Ziegler, J. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, **2,** 54.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, **7**(4), 493–498. Retrieved from https://CRAN.R-project.org/package=simr. doi: 10.1111/2041-210X.12504

Gulliksen, H. (1950). The reliability of speeded tests. *Psychometrika*, **15**(3), 259–269.

Häikiö, T., Bertram, R., & Hyönä, J. (2011). The development of whole-word representations in compound word processing: Evidence from eye fixation patterns of elementary school children. *Applied Psycholinguistics*, **32**(3), 533–551.

Hasenacker, J., Schroter, P., & Schroeder, S. (2017). Investigating developmental trajectories of morphemes as reading units in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **43**(7), 1093–1108.

Hasenäcker, J., Beyersmann, E., & Schroeder, S. (2020). Morphological priming in children: Disentangling the effects of school-grade and reading skill. *Scientific Studies of Reading*, **24**(6), 484–499.

Henry, R., & Kuperman, V. (2013). Semantic growth of morphological families in English. *Psihologija*, **46**(4), 479–495.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, **59**(4), 434–446.

Jiang, Y.-L., & Kuo, L.-J. (2019). The development of vocabulary and morphological awareness: A longitudinal study with college EFL students. *Applied Psycholinguistics*, **40**(4), 877–903.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, **68,** 601–625.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, **94**(2), 223–244.

Ke, S., Miller, R. T., Zhang, D., & Koda, K. (2021). Crosslinguistic sharing of morphological awareness in biliteracy development: A systematic review and meta-analysis of correlation coefficients. *Language Learning*, **71**(1), 8–54.

Keuleers, E. (2013). VWR: Useful functions for visual word recognition research [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=vwr (R package version 0.3.0).

Kuperman, V., Bertram, R., & Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language*, **62**(2), 83–97.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, **35**(3), 876–895.

Kuperman, V., & Van Dyke, J. (2011). Individual differences in visual comprehension of morphological complexity. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 33, p. 1643–1648).

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology. Human Perception and Performance*, **39**(3), 802–823.

Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W., et al. (2005). Applied linear statistical models (Vol. 5). McGraw-Hill Irwin New York.

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. doi: 10.18637/jss.v082.i13

Lam, K., Chen, X., Geva, E., Luo, Y. C., & Li, H. (2012). The role of morphological awareness in reading achievement among young Chinese-speaking English language learners: A longitudinal study. *Reading and Writing*, **25**(8), 1847–1872.

Laudanna, A., & Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 345–364). Erlbaum.

Laudanna, A., Burani, C., & Cermele, A. (1994). Prefixes as processing units. *Language and Cognitive Processes*, **9**(3), 295–316.

Lehtonen, M., Vorobyev, V. A., Hugdahl, K., Tuokkola, T., & Laine, M. (2006). Neural correlates of morphological decomposition in a morphologically rich language: An fMRI study. *Brain and Language*, **98**(2), 182–193.

Leminen, A., Leminen, M., Kujala, T., & Shtyrov, Y. (2013). Neural dynamics of inflectional and derivational morphology processing in the human brain. *Cortex*, **49**(10), 2758–2771.

Lenth, R. V. (2021). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=emmeans (R package version 1.7.0).

Libben, G. (2005). Everything is psycholinguistics: Material and methodological considerations in the study of compound processing. *The Canadian Journal of Linguistics/La revue canadienne de linguistique*, **50**(1), 267–283.

Libben, G. (2014). The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology*, **31**(1–2), 8–25.

Libben, G., Goral, M., & Baayen, H. (2017). Morphological integration and the bilingual lexicon. In M. Libben, M. Goral, & G. Libben (Eds.), *Bilingualism: A framework for understanding the mental lexicon* (pp. 197–216). John Benjamins.

Lõo, K., Jarvikivi, J., & Baayen, R. H. (2018b). Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing. *Cognition*, **175,** 20–25.

Lõo, K., Jarvikivi, J., Tomaschek, F., Tucker, B. V., & Baayen, R. H. (2018a). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology*, **28**(1), 71–97.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, **4,** 226.

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology*, **8,** 1164.

Marelli, M., Traficante, D., & Burani, C. (2020). Reading morphologically complex words: Experimental evidence and learning models. *Word Knowledge and Word Usage*, 553–592.

Mason, L. (2021). International student mobility flows and COVID19 realities. Institute of International Education. https://www.iie.org/Research-and-Insights/Publications/IntlStudent-Mobility-Flows-and-C19-Realities

Milin, P., & Blevins, J. P. (2020). *Paradigms in morphology*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.551

Monaghan, P., Chang, Y.-N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, **93,** 1–21.

Moscoso del Prado Martín, F. (2007). Co-occurrence and the effect of inflectional paradigms. *Lingue e linguaggio*, **6**(2), 247–263.

Moscoso del Prado Martín, F., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, **94**(1), 1–18.

Mulder, K., Dijkstra, T., & Baayen, R. H. (2015). Cross-language activation of morphological relatives in cognates: The role of orthographic overlap and task-related processing. *Frontiers in Human Neuroscience*, **9,** 16.

Mulder, K., Dijkstra, T., Schreuder, R., & Baayen, H. R. (2014). Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language*, **72,** 59–84.

Niswander, E., Pollatsek, A., & Rayner, K. (2000). The processing of derived and inflected suffixed words during reading. *Language and Cognitive Processes*, **15** (4–5), 389–420.

Orliaguet, J.-P., & Boë, L.-J. (1993). The role of linguistics in the speed of handwriting movements: Effects of spelling uncertainty. *Acta Psychologica*, **82**(1–3), 103–113.

Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development*, **76**(2), 324–339.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, **11**(4), 357–383.

Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-Effects Models in S and S-Plus*, 3–56.

Pynte, J., Courrieu, P., & Frenck, C. (1991). Evidence of repeated access to immediate verbal memory during handwriting. *European Journal of Psychology of Education*, **6**(2), 121–125.

Quick, N., & Erickson, K. (2018). A multilinguistic approach to evaluating student spelling in writing samples. *Language, Speech, and Hearing Services in Schools*, **49**(3), 509–523.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brothers brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, **11**(6), 1090–1098.

R Core Team. (2021). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. https://www.R-project.org/

Sahel, S., Nottbusch, G., Grimm, A., & Weingarten, R. (2008). Written production of German compounds: Effects of lexical frequency and semantic transparency. *Written Language & Literacy*, **11**(2), 211–227.

Schiff, R., Raveh, M., & Fighel, A. (2012). The development of the Hebrew mental lexicon: When morphological representations become devoid of their meaning. *Scientific Studies of Reading*, **16**(5), 383–403.

Schmidtke, D., Matsuki, K., & Kuperman, V. (2017). Surviving blind decomposition: A distributional analysis of the time-course of complex word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **43**(11), 1793–1820.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, **23**(4), 569–588.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, **55**(2), 151–218.

Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, **22**(9), 2042–2057.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, **24**(4), 402–433.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, **7**(4), 263–272.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology Section A*, **57**(4), 745–765.

Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, **71,** 249–267.

Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of memory and language*, **28**(6), 649–667.

Ulicheva, A., Marelli, M., & Rastle, K. (2020). Sensitivity to meaningful regularities acquired through experience. *Morphology*, 1–22.

Weingarten, R., Nottbusch, G., & Will, U. (2004). Morphemes, syllables, and graphemes in written word production. *Trends in linguistics studies and monographs*, **157,** 529–572.

Whitford, V., & Joanisse, M. F. (2018). Do eye movements reveal differences between monolingual and bilingual childrens first-language and second-language reading? A focus on word frequency effects. *Journal of Experimental Child Psychology*, **173,** 318–337.

Whitford, V., & Titone, D. (2012). Second-language experience modulates first-and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, **19**(1), 73–80.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wu, Z., & Juffs, A. (2021). Effects of L1 morphological type on L2 morphological awareness. *Second Language Research.* https://doi.org/10.1177/0267658321996417
Zhang, D., & Koda, K. (2012). Contribution of morphological awareness and lexical inferencing ability to L2 vocabulary knowledge and reading comprehension among advanced EFL learners: Testing direct and indirect effects. *Reading and Writing*, **25**(5), 1195–1216.
Zhang, D., & Koda, K. (2013). Morphological awareness and reading comprehension in a foreign language: A study of young Chinese EFL learners. *System*, **41**(4), 901–913.
Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, **131**(1), 3–29.

Appendix A. Test of morphological structure

Table A.1. Test of morphological structure, List A

Base Word	Sentence	Target Word
1. warm	He chose the jacket for its ____________________.	[warmth]
2. permit	The father refused to give ____________________.	[permission]
3. profit	Selling lemonade in the summer is ____________________.	[profitable]
4. protect	She wore glasses for ____________________.	[protection]
5. perform	Tonight is the last ____________________.	[performance]
6. revise	This paper is his second ____________________.	[revision]
7. reason	Her argument was quite ____________________.	[reasonable]
8. major	He won the vote by a ____________________.	[majority]
9. deep	The lake was dangerous because of its ____________________.	[depth]
10. adventure	The trip sounded ____________________.	[adventurous]
11. active	He tired after so much ____________________.	[activity]
12. swim	She was a strong ____________________.	[swimmer]
13. humor	The story was quite ____________________.	[humorous]
14. glory	The view from the hilltop was ____________________.	[glorious]

Table A.2. Test of morphological structure, List B

Base Word	Sentence	Target Word
1. teach	He was a very good ___________________.	[teacher]
2. appear	He cared about his ___________________.	[appearance]
3. express	"OK" is a common ___________________.	[expression]
4. four	The cyclist came in ___________________.	[fourth]
5. remark	The speed of the car was ___________________.	[remarkable]
6. expand	The company planned an ___________________.	[expansion]
7. equal	Boys and girls are treated with ___________________.	[equality]
8. long	They measured the ladder's ___________________.	[length]
9. absorb	She chose the sponge for its ___________________.	[absorption]
10. human	The kind man was known for his ___________________.	[humanity]
11. wash	Put the laundry in the ___________________.	[washer]
12. assist	The teacher will give you ___________________.	[assistance]
13. mystery	The dark glasses made the man look ___________________.	[mysterious]
14. produce	The play was a grand ___________________.	[production]

Appendix B. Response type distribution

Figure B.1 provides the counts of the different spelling types broken down by timepoint. The response types are based on the POMplexity framework (Bahr et al., 2020) outlined in Methods. A chi-square test of association confirmed that there was a relationship between response type and Timepoint, $\chi^2(5, N = 5{,}488) = 110.11$, $p < .001$. This analysis indicates that there was a significant imbalance in the number of response

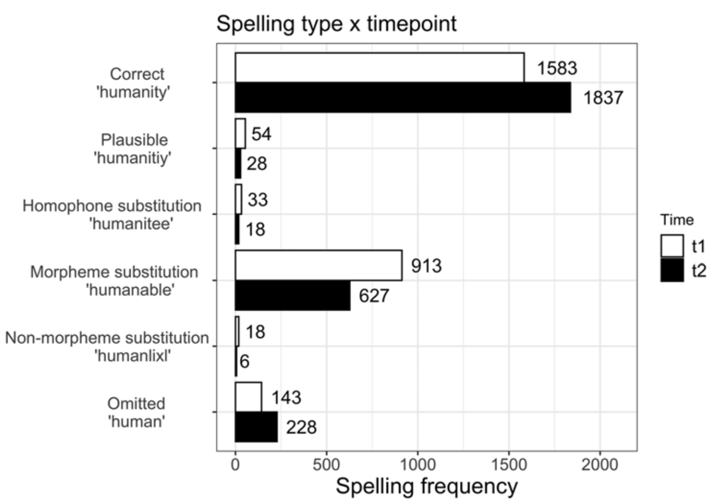

Figure B.1. Frequency of spellings broken down by type and timepoint.

https://doi.org/10.1017/S0142716422000182 Published online by Cambridge University Press

types across timepoints. The figure shows that there were more fully correct spellings, e.g., *humanity*, at t_2 compared to t_1. In addition, there were fewer "plausible" and "homophone" substitution errors at t_2 relative to t_1, e.g., *humanitiy*. Finally, the largest change in error type across timepoints was found for morpheme substitution errors: cases for which an incorrect suffix was applied to the stem form, e.g., *humanable*.

Appendix C. List of stimuli with distributional lexical properties

Table C.1. List of stimuli with distributional lexical properties

Target	Suffix	Derived freq.	Stem freq.	Suffix prod.	Deriv. entropy	OLD20
absorption	ion	18	101	954	1.104	2.90
activity	ity	657	584	453	2.684	2.20
adventurous	ous	62	685	253	1.569	2.90
appearance	nce	591	1,192	75	1.384	3.35
assistance	nce	558	400	75	0.983	2.90
depth	th	421	3,896	45	1.817	1.85
equality	ity	97	682	453	1.590	2.20
expansion	ion	88	178	954	0.646	2.55
expression	ion	651	914	954	0.896	2.35
fourth	th	1,349	13,045	45	0.163	1.85
glorious	ous	474	1,096	253	1.138	2.50
humanity	ity	495	6,363	453	2.510	2.50
humorous	ous	75	849	253	0.491	2.30
length	th	358	34,433	45	1.279	1.85
majority	ity	261	5,343	453	−0.000	2.65
mysterious	ous	765	1,171	253	0.668	3.05
performance	nce	1,113	821	75	0.284	3.15
permission	ion	1,587	617	954	0.356	2.70
production	ion	645	564	954	1.239	2.40
profitable	able	125	559	617	1.651	2.85
protection	ion	1,197	3,583	954	0.912	2.20
reasonable	able	946	9,858	617	1.226	2.50
remarkable	able	641	224	617	1.107	2.80
revision	ion	29	44	954	0.911	1.85
swimmer	er	190	1,622	2,163	0.344	1.75
teacher	er	2,842	3,715	2,163	0.041	1.55
warmth	th	227	2,659	45	1.796	2.00
washer	er	104	2,077	2,163	1.041	1.35

Appendix D. Fixed and random effects of the generalized linear mixed-effects model fitted to production accuracy

Table D.1. Fixed effects of the generalized linear mixed-effects model fitted to production accuracy. Conditional $R^2 = 0.46$. *SD* of the residual = 0.96. *N* = 5,488. *N* after trimming = 5,478

	Estimate	*SE*	*z*	*p*
Intercept	0.928	0.175	5.288	.000
Timepoint	0.249	0.048	5.141	.000
Suffix productivity (log)	0.422	0.218	1.932	.053
Derived word frequency (log)	0.525	0.191	2.749	.006
Derivational family entropy	−0.554	0.151	−3.657	.000
Stem frequency (log)	0.319	0.214	1.493	.135
Orthographic neighbourhood density	−0.179	0.355	−0.505	.613
Word length	0.157	0.369	0.426	.670
Age of initial English instruction	−0.006	0.072	−0.085	.932
List order	−0.051	0.146	−0.349	.727
Timepoint × Suffix productivity (log)	−0.128	0.037	−3.445	.001
Timepoint × Derived word frequency (log)	−0.073	0.037	−1.983	.047

Table D.2. Random effects of the generalized linear mixed-effects model fitted to production accuracy

Random effect	*SD*	Correlations between by-participant slopes and intercepts
Item	0.74	
Participant	0.9	
Timepoint by participant	0.22	0.21

Cite this article: Schmidtke, D., Rahmanian, S., and Moro, AL. (2022). Morphological knowledge in English learner university students is sensitive to language statistics: A longitudinal study. *Applied Psycholinguistics* **43**, 889–919. https://doi.org/10.1017/S0142716422000182