Journal of Memory and Language 113 (2020) 104127

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

A story about statistical learning in a story: Regularities impact eye movements during book reading

Joshua Snell*, Jan Theeuwes

Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Netherlands
Institute of Brain and Behavior Amsterdam, Netherlands

ARTICLE INFO

Keywords:
Reading
Statistical learning
Oculomotor control
Syntactic processing
Book reading

ABSTRACT

A wealth of research attests to the key role of statistical learning in the acquisition and execution of skilled reading. Little is known, however, about how regularities impact the way readers navigate through their linguistic environment. While previous studies have mostly gauged the recognition of single words, oculomotor processes are likely influenced by multiple words at once. With these premises in mind, we performed analyses on the GECO book reading corpus to determine whether repeatedly encountering a given sentence structure improves oculomotor control. In the reading materials we labeled structures on the basis of both low- and high-level properties: respectively word length combinations (e.g., a 2-letter word followed by a 6-letter word followed by a 4-letter word) and syntactic structures (e.g., an article followed by a noun followed by a verb). Our analyses show that repeatedly encountering a structure leads to fewer and shorter fixations, and fewer corrective saccades. Critically, learning curves are steeper for structures that have a higher overall frequency, hence evidencing true statistical learning over and above readers' general tendency to accelerate as they progress through the book. Further, data from Dutch-English bilingual readers suggest that these types of learning occur across languages and at various levels of proficiency. We surmise that the reading system is tuned to statistical regularities pertaining not just to single words but also combinations of words. These regularities impact both linguistic processing and oculomotor control.

Introduction

It is evident that any language, both oral and written, is rife with statistical regularities. Perhaps unsurprisingly, it has been suggested that language acquisition itself largely depends on one's ability to detect these (e.g., Chater & Manning, 2006; Kuhl, 2004; see Siegelman, Bogaerts, Christiansen, & Frost, 2017, for an extensive review). Statistical learning bears prominence not only during language acquisition, but also in the way highly proficient language users process linguistic input. The most straightforward example is that regular words are recognized faster than irregular words (e.g., White, Warrington, McGowan, & Paterson, 2015). Additionally, predictable words—i.e., words that more regularly succeed a given word or clause—are recognized faster than unpredictable words (e.g., Ehrlich & Rayner, 1981; Rayner & Well, 1996; Kretzschmar, Schlesewsky, & Staub, 2015); and learned associations between a given context and its constituent words may similarly facilitate recognition (McDonald & Shillcock, 2001).

While the above examples pertain to linguistic processing, language processing is not a purely linguistic matter. Our ability to read, for instance, depends not only on linguistic processing but also on attentional- and oculomotor control. One might reasonably claim that spoken language can be 'passively' processed; but this does not hold for written language, where the receiving party—the reader—has to pro-actively move the eyes from word to word to garner the message being conveyed. It is precisely this pro-active component of (written) language processing for which we do not yet know to what extent statistical learning plays a role. Simply put, although there are accounts of how statistical regularities influence word recognition, much less is known about how statistical regularities influence *the way we navigate through our linguistic environment*. This is the starting point of the present work.

This paper addresses the following question: Do statistical regularities affect oculomotor control during text reading? It should be noted that the developing reader might be explicitly instructed about certain regularities, such as which letters belong to which sounds (and when and why there are exceptions); but a myriad of other regularities are likely implicitly learned along the course of encountering large amounts of text. For example, readers will gradually learn on which syllables to

* Corresponding author.
E-mail address: J.J.Snell@VU.nl (J. Snell).

https://doi.org/10.1016/j.jml.2020.104127
Received 19 August 2019; Received in revised form 28 April 2020; Accepted 29 April 2020
Available online 12 May 2020

put lexical stress, and that (English) words may often end with double letters (e.g. '*boss will miss*') but will never begin with double letters (Arciuli & Simpson, 2011). Concerning our question of interest, readers are not explicitly instructed on how to move their eyes from word to word; but they nevertheless do learn to optimize word targeting strategies, evidenced by the fact that average fixation landing positions gradually shift towards the word center during the first years of learning to read (Ducrot, Pynte, Ghio, & Lété, 2013). Furthermore, saccade amplitudes are strongly influenced by word length (e.g., McConkie, Kerr, Reddix, & Zola, 1988; Rayner, 1998).

Despite these strategies, saccades are error-prone even in the most experienced readers. Readers quite frequently execute a regressive saccade very quickly after the fixation onset, indicative of a correction after initially overshooting the target location (Radach, Heller, & Inhoff, 1999; Rayner, 1998; Vitu & McConkie, 1998). Additionally, the distribution of initial landing positions (ILPs) is generally quite diffuse, and especially so in poor readers (e.g., Gagl, Hawelka, & Hutzler, 2014). In short, readers do not have perfect oculomotor control—and the issue investigated here is whether certain statistical regularities help improve this. Below, we will discuss, respectively, the potential impact of low-level visual regularities (i.e. combinations of word lengths) and high-level linguistic regularities (i.e., syntactic structures).

Word lengths

Given that the calculation of saccade amplitude is impacted not just by the length of the fixated word but also by the length of parafoveal upcoming words (e.g., Kliegl, Grabner, Rolfs, & Engbert, 2004; McConkie, Kerr, Reddix, & Zola, 1988; Schotter, Angele, & Rayner, 2012; Snell, van Leipsig, Grainger, & Meeter, 2018b), our first question was whether highly regular combinations of word lengths (e.g., a 3-letter word followed by a 6-letter word followed by a 4-letter word) would prompt cleaner oculomotor behavior than irregular word length combinations. Here, cleaner oculomotor behavior would be reflected in *narrower ILP distributions* and *fewer corrective saccades*, indicative of more accurate saccade targeting.

In line with the idea that word length regularities may guide oculomotor behavior, it has long been assumed that the reader's past experiences are an important factor in saccade programming (e.g., leading to a generally preferred amplitude of 7 character spaces in English; see McConkie et al., 1988). On the other hand, recent work by Cutter, Drieghe and Liversedge (2017, 2018) suggests that such oculomotor processes can rapidly adapt to novel textual properties. Specifically, they found that switching between sentences comprising variable versus uniform word lengths impacted eye movements on a trial-by-trial basis (2018). Note that, on the basis of these results alone, one could still hypothesize a role for long-term statistical learning. Although uniform word length sequences were intermixed with variable word length sequences, all sentence types (specifically, sentences comprising 3-, 4- and 5-letter words, respectively) did occur many times throughout the experiment. It could be argued that this may, during the course of the experiment, have led the reader to develop a motor plan for when encountering 3-letter word sequences, a motor plan for when encountering 4-letter word sequences, et cetera. Yet, such a scenario generates the prediction that effects should be stronger in the second half of the experiment than in the first half; and as it happens, an (unpublished) analysis by Cutter et al. suggested that this was not the case. A potential cause for the absence of long-term learning in their study may have been the fact that the sentences were not interrelated (e.g., the sentences did not form a story or overarching context).

Syntax

Whereas an impact of low-level visual regularities on oculomotor behavior would easily tie in with prior literature (e.g. Schotter et al., 2012; Rayner, 1998), less is known about the potential impact of high-level linguistic regularities. Therefore, our second question was whether highly regular syntactic structures lead to cleaner oculomotor behavior than irregular structures.

It should be noted that some evidence for implicit pattern recognition in syntactic comprehension already exists. Firstly, priming studies have shown that a sentence may be processed faster if the preceding sentence had the same grammatical structure (e.g., Ledoux, Traxler, & Swaab, 2007; Giavazzi et al., 2018; Tooley & Traxler, 2010, for a review). However, such effects are somewhat elusive and, given the trial-to-trial nature of the paradigms involved, may be short-lived. On the other hand, Bock and Griffin (2000) have argued for a key role of long-term memory and implicit learning, at least within the realm of syntactic priming in speech production. Additionally, in the auditory domain, Fine, Jaeger, Farmer, and Qian (2013) observed that the processing disadvantage when encountering (syntactically ambiguous) garden-path sentences can be reduced by context-driven expectations, such as expectations prompted by the speaker or by the genre. Such associations between context and syntax may well be attained through statistical learning. Further strengthening this conception is the work of Saffran and Wilson (2003), who found that 12-month old infants are sensitive to regularities not only at sub- but also supra-lexical levels.

What has remained unclear, is whether these syntactic effects—be they the consequence of short-term priming or long-term statistical learning—truly rely on syntactic representations (e.g., a representation coding for the structure *article noun verb adverb*), or rather on word-to-word transitional probabilities. In case of the latter, for instance, recognition of the word '*The*' might pre-activate a set of candidate follow-up words that happen to predominantly fall in the *noun* category (see e.g. Thompson & Newport, 2007). By this rationale, aforementioned effects would possibly be semantic rather than syntactic in nature. As shall be seen in due course, in order to reveal true syntactic statistical learning, in the present study we have controlled for the potential contribution of semantics.

If certain regularities are only implicitly learned along the course of processing large amounts of text, then it is conceivable that their impact may exclusively be revealed through large corpus analyses (see also Arciuli & Simpson, 2011). This conception prompted us to invoke the work of Cop, Dirix, Drieghe, and Duyck (2017), who published a corpus containing the eye-tracking data of 33 participants reading a whole book. This incredibly rich corpus has already proven its worth in previous investigations (Snell, Grainger, & Declerck, 2018a) and contains all the ingredients necessary for answering our questions, such as the eye position coordinates and number of fixations on all words, as well as the grammatical category of all words occurring in the book. Going beyond general word frequency effects, the strategy employed here allowed us to determine whether oculomotor behavior evolves as a function of statistical regularities spanning combinations of words, picked up within the scope of a story. Such findings would show that oculomotor processes at least in part rely on parameters that flexibly tune to broader contexts, such as whole books.

The present study may potentially aid the ongoing development of computational models of reading. Both well-established models (e.g., *E-Z Reader* by Reichle, Rayner, and Pollatsek (2003); *SWIFT* by Engbert, Nuthmann, Richter, and Kliegl (2005); *Glenmore* by Reilly and Radach (2006)) and more recent models (e.g., *OB1-reader* by Snell et al., 2018b) as of yet lack mechanisms by which regularities may impact the recognition process. Although these models accommodate word frequency effects, frequencies are usually implemented as fixed values per word rather than, for example, being adjusted by recent recurrences. Following the present investigation, we shall return to this issue in Section 4.2 of the Discussion.

Lastly, the GECO corpus allowed us to obtain cross-lingual evidence. Among the participants were Dutch-English bilinguals who read one half of the book in Dutch and the other half in English (counter-balanced). We were therefore able to inspect whether certain types of statistical learning may be contingent on proficiency. For instance, one might hypothesize that the influence of low-level visual regularities (i.e., word length combinations) should be observable in Dutch (L1) and English (L2) alike, while the influence of high-level regularities (i.e., syntactic structures) may be more fleeting in L2. On the other hand, there is some evidence that syntactic representations are shared between languages; both in terms of language production (Hartsuiker, Pickering, & Veltkamp, 2004) and language comprehension (Vaughan-Evans, Kuipers, Thierry, & Jones, 2014; Declerck, Wen, Snell, Meade, & Grainger, 2020). In such a case, effects observed in L2 English reading by the Dutch-English bilinguals may be comparable to L1 English reading by the English monolinguals.

Methods

Cop et al. (2017) let 18 Dutch-English bilingual subjects[1] (F = 16, age range = 18–24) read the novel *The mysterious affair at Styles*, by Agatha Christie. As noted above, subjects read one half of the book in Dutch (their L1) and the other half in English (L2), in counter-balanced order. The subjects' second language profiency was of an upper-intermediate (60–80%) level. Cop et al. (2017) additionally collected data from 14 English monolinguals (F = 8, age range = 18–36) who read the whole book in English. For our main questions—whether low- and high-level regularities impact oculomotor processes at all—we relied on data from the latter group, as this allowed us to inspect the whole book without having to take language switches into account. Data from the Dutch-English bilinguals, in turn, was used firstly to replicate results established in the English monolinguals, and subsequently to assess the impact of proficiency (L2 English versus L1 English reading).

Cop et al. (2017) collected data for 59,716 words in the Dutch version of the novel (per two subjects, as Dutch subjects read half of the book in English), and for 54,364 words in the English version. The story was presented in paragraphs (max. 145 words per display) with black 14-point Courier New font on a light grey background and with triple line spacing. Subjects could press a button to move from one paragraph to the next. After each of the 18 book chapters, subjects were presented multiple-choice questions to ensure that they had paid attention throughout the chapter. Reading of the entire novel took approximately four hours and was carried out in four 1-hour sessions. Eye movements were tracked with the EyeLink 1000 system (SR Research, Canada) with a sampling rate of 1 kHz.

Our datasets

For our analyses we looked at the first four words of each sentence with a length of at least four words.[2] All four-word combinations were labeled by word length (in number of characters; e.g., '*He left the house*' would be labeled as *2-4-3-5*) and syntactic structure (e.g., '*He left the house*' would be labeled as *pronoun-verb-article-noun*). Subsequently, we determined for each structure label how often it occurred in the book. From the total of 5301 sentences in the English materials, and 5190 sentences in the Dutch materials, only a relatively small portion comprised a structure that occurred more than a handful of times throughout the book. The 300 most frequent length structures and syntax structures for the English version of the book are plotted in Fig. 1 (note that the distributions of Dutch structure frequencies were comparable). In order not to muddy our statistical waters with a wealth of single-occurrence structures, our analyses were confined to these 300 most frequent combinations. As such we included ~2500 (~47%) of the book's sentences in the syntax analyses, and ~1800 (~34%) of the book's sentences in the word length analyses.

Variables of interest

We gauged four dependent variables: *Viewing time*, *Number of fixations*, *Corrective saccade probability* (averaged across the four words) and *Initial Landing Position* (*ILP*) *distributions*.

Viewing time and number of fixations were calculated as respectively the sum of fixation durations and the sum of fixations across all four words. Given that these measures include refixations (even after the reader made a regression back onto the word), they could be argued to reflect the overall ease of reading. To gauge initial commitments—i.e., the fixation behavior when foveating words for the first time, which might more directly reflect oculomotor processes without influences of context comprehension—we additionally report analyses of the fixations and fixation durations after excluding all refixations.

Corrective saccades were assumed to have occurred if two consecutive fixations were registered on the same word with the first fixation having a duration below 100 ms (for more on corrective saccades, see e.g. Bowers & Poletti, 2017; Rayner, 1998).

With respect to ILP distributions, it should be noted that these are classically presented as a histogram of landing positions whereby the left and right end of the horizontal axis represent the left and right word boundary, respectively. In the present study, however, ILP distributions were assessed as the *spread* of the three fixation locations observed at the final three encounters of each structure (per subject). For instance, if a given structure occurred 20 times in the book, for the ILP spread we would only inspect the fixation locations observed in the 18th, 19th and 20th encounter. This allowed us to compare all structure frequencies-i.e., including the least frequent structures which only occurred three times throughout the book. As such, differences between low- and high-frequency structures could be attributed to the number of previous encounters. ILP spreads were collected and analysed for each of the four word positions separately, with the expectation that effects increase in strength with each successive word as a consequence of increasing certainty about the type of structure being read.

Analyses

We employed Linear Mixed-effects Models (LMMs) in the *R* statistical computing environment to analyze the data on a sentence-by-sentence basis. All models-save for those analyzing ILP distributions-included the factors Structure Frequency (as reflected in Fig. 1A and B) and Encounter (i.e., how often the structure has been encountered thus far), as well as their interaction. The crucial aspect was to not just establish a main effect of Encounter (as such an effect may simply be driven by readers' general tendency to accelerate as they progress through the book), but also an interaction between Encounter and Structure Frequency, as this would provide evidence that more regular structures facilitate oculomotor processes over and above general task practice.[3]

To ensure that effects were indeed driven by our regularities of interest (i.e., word length combinations and syntactic structures) rather than by repetitions of specific word identities or phrases, we also entered the number of previous encounters for single words, word pairs, triplets and quadruplets as covariates in the models.[4] In addition to

[1] Cop et al. (2017) report data from a 19th subject who read half of the book, but that subject is excluded from our analyses.

[2] The reason why we did not inspect whole sentence structures, is that virtually all structures were unique (i.e., having only a single occurrence in the whole book) at the level of whole sentences.

[3] One may be aware that the critical test of an Encounter × Struct. Freq. interaction relies on the implicit assumption that all structures were evenly distributed throughout the book. Our hypotheses could also be tested by assessing the interaction between Struct. Freq. and the structure's location in the book. Importantly, we verified that statistical patterns with the use of this alternative variable were virtually the same.

[4] Note that upon each encounter, a structure yielded multiple values for the number of times single words, word pairs and word triplets were previously encountered. Those values were averaged for each datapoint. For instance, to

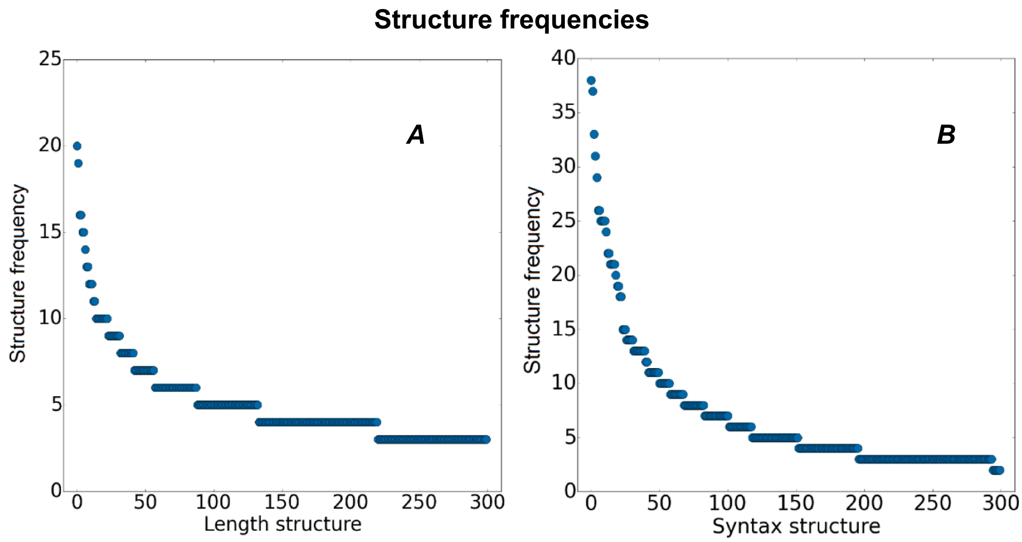

Fig. 1. Ordered arrangements of the 300 most frequent four-word length combinations (A) and syntax combinations (B) occurring in the English version of the book. Each dot represents a structure type.

these factors, Subjects were added as random effect. The models included the by-Subject random intercept as well as by-Subject random slopes for both experimental factors as well as their interaction. Note that Items were not included as random effect because items could not occur across all conditions (e.g., a given sentence would always have the same Structure Frequency value).[5] We report *b*-values, Standard Errors (SEs) and *t*-values, with $t > |2|$ deemed significant.

For the ILP spreads, models did not include the Encounter factor. This is because ILP spreads were defined as the standard deviation of the ILPs of the final three encounters per structure per subject. As noted in Section 2.2, the reason why we looked only at the final three encounters, is so that we were able to compare high frequency structures to low frequency structures, the latter of which occurred no more than three times in the whole book. Further, whereas Structure Frequency was treated as a continuous variable in the models analyzing viewing time, number of fixations and rate of corrective saccades, for the ILP spreads Structure Frequency was binned into two levels by the median split. The latter was done to ensure sufficient statistical power, given that each structure yielded but a single datapoint per participant in the ILP analyses (as opposed to all other analyses, where structures yielded a datapoint per encounter per participant).

To assess the risk that effects attributed to one type of regularity were confounded with the other, we determined the number of sentences in the book that, upon being reached by the reader, had matching values for Encounter/Structure Frequency between syntactic regularities and word length regularities. This turned out to be the case for 22 sentences, equalling 308 datapoints (0.9% of data in the syntax analyses, 1.2% of data in the word length analyses). Given this low number, we deemed the risk of confounding regularity effects negligible.

(*footnote continued*)
calculate the number of previous word pair encounters when the participant saw the structure "*The man reads now*", we averaged the number of times "*The man*" was seen before, the number of times "*man reads*" was seen before, and the number of times "*reads now*" was seen before.

[5] The basic structure of our Linear Mixed-effect Models as denoted in *R* syntax:*Dependent variable* ~ StructFreq * Encounter + (1+ StructFreq*Encounter | Subject)

Results

Below, we report our assessment of length effects and syntax effects separately in respectively Sections 3.1 and 3.2. The impact of having previously encountered specific word identities or phrases is reported in Section 3.3 (but note that these were also included as covariates in the models reported in Sections 3.1 and 3.2). Prior to all analyses, we excluded encounters where the viewing time was beyond 2.5 SD from the grand mean. On average this led to the exclusion of ~2% of datapoints. As noted in Section 2.3, for establishing true statistical learning, in our analyses (with the exception of ILP distributions) we rely not on a main effect of the number of times a structure has been encountered, nor on a main effect of the overall frequency of the structure, but on the interaction between these two variables.

In the interpretation of results reported in Sections 3.1 and 3.2, we must highlight one particular aspect of the data. In general, as shall be seen below, we obtained strongly significant effects evidencing statistical learning. These effects of repeatedly encountering length and syntax structures were established while controlling for the impact of repeatedly encountering specific word identities or phrases (Section 3.3). However, from the graphs presented in Figs. 2 and 4, it can be seen that effects were consistent up until ~7 encounters, after which the data become very noisy.[6] As shall be elaborated upon in the Discussion, this may be caused not only by the fact that there are fewer datapoints for more frequent structures (see Fig. 1), but also by the fact that statistical learning effects may generally reach an asymptote fairly quickly (e.g., Verstynen & Sabes, 2011; Wang & Theeuwes, 2018).

Word length regularities

The impact of word length regularities on viewing time, number of fixations and corrective saccade probability is plotted in Fig. 2. Word length effects on ILP spreads per word are plotted in Fig. 3.

Viewing time

Repeatedly encountering a given combination of word lengths significantly decreased the viewing time ($b = -9.92$, SE = 2.58,

[6] Because of the non-linear aspect of the data beyond the ~7th encounter, we also ran our LMMs (which assume linearity) on the subset of data between 0 and 7 encounters. Importantly, all main effects and interactions as reported in Sections 3.1 and 3.2 persisted.

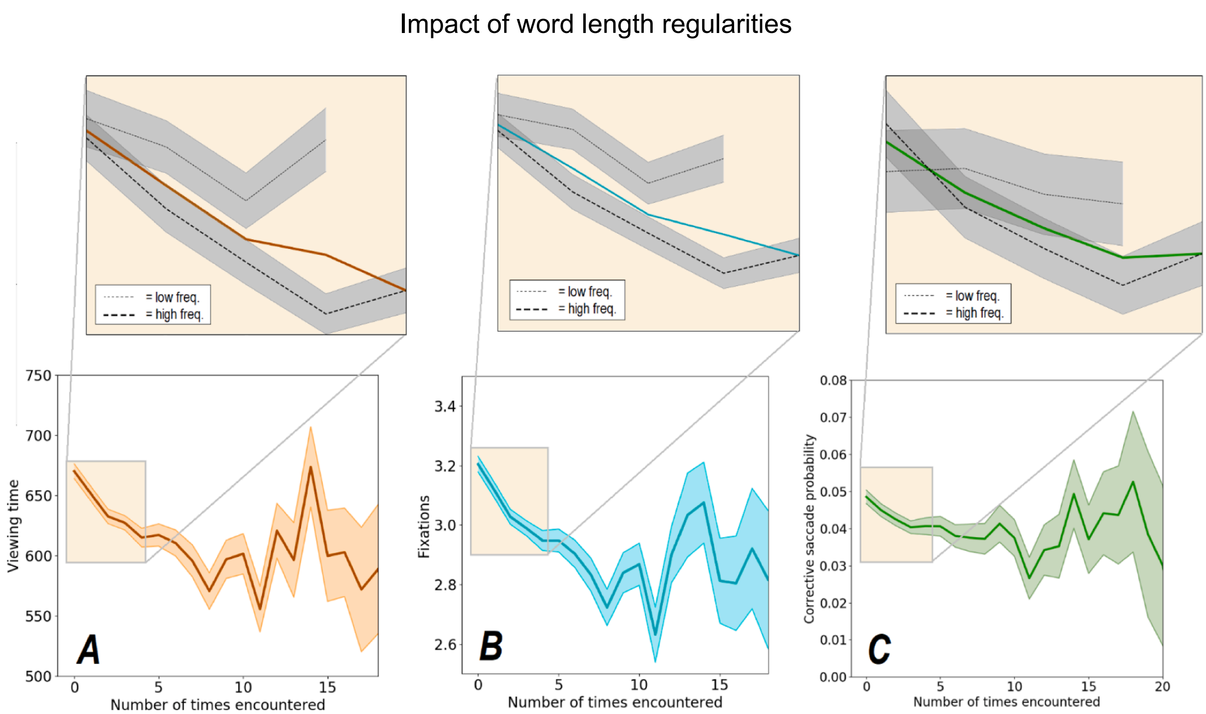

Fig. 2. The impact of repeatedly encountering a given combination of word lengths on viewing time (A), number of fixations (B), and corrective saccade probability (C). Zooming in on the first four encounters, the top panels show separate effects of repeatedly encountering the structure when it is above (thick dashed line) or below (thin dotted line) the frequency median. Note that low-frequency structures occurred no more than three times in the book. Shaded areas around the curves depict standard errors.

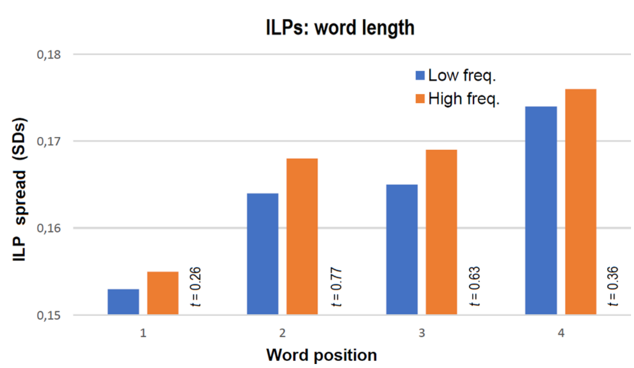

Fig. 3. ILP spreads (in standard deviations) per word position, for word length combination frequencies above (orange) and below (blue) the frequency median. Significance values for individual word positions are listed on the right of each pair of columns.

$t = -3.84$). Frequent structures were read faster ($b = -5.82$, SE $= 1.02$, $t = -5.73$), and we also observed an interaction, with learning curves being steeper for more frequent structures ($b = 1.04$, SE $= 0.19$, $t = 5.49$), hence evidencing true statistical learning over and above general task practice. Isolating the durations of the first fixations on all words (i.e., excluding refixations) we observed a marginally significant effect of Encounter ($b = -3.40$, SE $= 1.86$, $t = -1.82$), a significant effect of Frequency ($b = -2.97$, SE $= 0.73$, $t = -4.06$) and again a significant interaction ($b = 0.46$, SE $= 0.14$, $t = 3.35$).

Number of fixations

Similar patterns were observed in the number of fixations, with main effects of Encounter ($b = -4.82*10^{-2}$, SE $= 1.20*10^{-2}$, $t = -4.02$) and Frequency ($b = -2.75*10^{-2}$, SE $= 4.71*10^{-3}$, $t = -5.84$), as well as an interaction between these factors ($b = 4.72*10^{-3}$, SE $= 8.83*10^{-4}$, $t = 5.35$). The same patterns were observed when excluding all refixations, with effects of Encounter ($b = -0.03$, SE $= 0.01$, $t = -3.14$), a significant effect of Frequency ($b = -0.02$, SE $= 3.45*10^{-3}$, $t = -5.78$) and again a significant interaction ($b = 3.03*10^{-3}$ SE $= 6.47*10^{-4}$, $t = 4.68$).

Corrective saccades

The corrective saccade probability decreased with each successive encounter ($b = -3.06*10^{-3}$, SE $= 9.29*10^{-4}$, $t = -3.29$) and was generally lower for more frequent length combinations ($b = -9.42*10^{-4}$, SE $= 3.66*10^{-4}$, $t = -2.58$). Here again we observed an interaction ($b = 2.25*10^{-4}$, SE $= 6.85*10^{-5}$, $t = 3.28$), indicating that learning curves in saccade targeting were steeper for more regular structures.

ILP spreads

Our hypothesis that regularities should cause narrower ILP distributions was not confirmed. Although word length regularities facilitated saccade targeting accuracy (reflected in the reduced corrective saccade probability), this did not lead to more consistent ILPs (overall, $b = 0.004$, SE $= 0.003$, $t = 1.49$; see Fig. 3).

Syntactic regularities

Syntactic learning effects on viewing time, number of fixations and corrective saccade probability are plotted in Fig. 4. Syntax effects on ILP spreads per word are plotted in Fig. 5.

Viewing time

Repeatedly encountering a given syntactic structure significantly decreased the time spent viewing it ($b = -2.30$, SE $= 0.62$, $t = -3.69$). Frequent structures were read faster (b = -1.22, SE $= 0.22$, $t = -5.53$), and we also observed an interaction, with the effect of repeatedly encountering a structure being stronger for more frequent structures ($b = 0.08$, SE $= 0.02$, $t = 4.60$). Only counting the first fixation on each word we observed the same patterns, with a

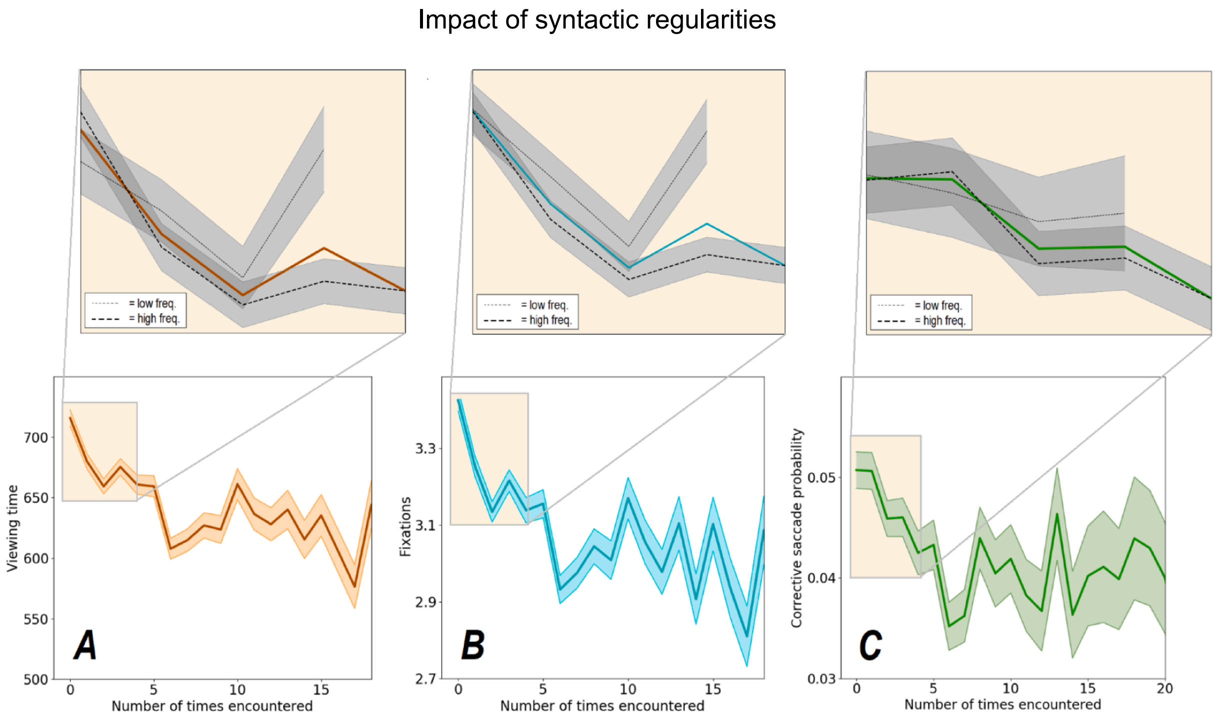

Fig. 4. The impact of repeatedly encountering a given syntactic structure on viewing time (A), number of fixations (B), and corrective saccade probability (C). Zooming in on the first four encounters, the top panels show separate effects of repeatedly encountering the structure when it is above (thick dashed line) or below (thin dotted line) the frequency median. Note that low-frequency structures occurred no more than three times in the book. Shaded areas around the curves depict standard errors.

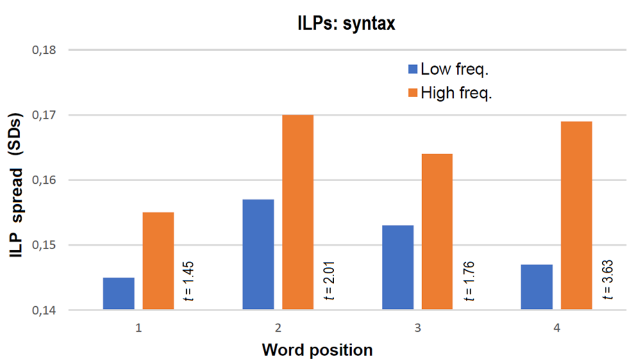

Fig. 5. ILP spreads (in standard deviations) per word position, for syntactic structure frequencies above (orange) and below (blue) the frequency median. Significance values for individual word positions are listed on the right of each pair of columns. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

significant effect of Encounter (b = −1.15, SE = 0.43, t = −2.66), Frequency (b = −0.84, SE = 0.15, t = −5.47) and an interaction (b = 0.04, SE = 0.01, t = 3.47).

Number of fixations

Similar patterns were observed in the number of fixations, with main effects of Encounter (b = −0.01, SE = 0.003, t = −3.23) and Structure Frequency (b = −0.01, SE = 0.001, t = −6.78), as well as an interaction between the two factors ($b = 3.96*10^{-4}$, SE = $7.83*10^{-5}$, t = 5.06). Only counting the first fixation on each word we again observed the same patterns, with a marginally significant effect of Encounter ($b = -4.09*10^{-3}$, SE = $2.13*10^{-3}$, t = −1.92), Frequency ($b = -6.16*10^{-3}$, SE = $7.56*10^{-4}$, t = −8.14) and an interaction ($b = 2.66*10^{-4}$, SE = $5.73*10^{-5}$, t = 4.64).

Corrective saccades

The number of corrective saccades decreased with each successive encounter ($b = -6.23*10^{-4}$, SE = $2.20*10^{-4}$, t = −2.84) while Structure Frequency had no significant influence on this measure ($b = -8.52*10^{-5}$, SE = $7.80*10^{-5}$, t = −1.09). We did however again observe an interaction ($b = 1.54*10^{-5}$, SE = $5.92*10^{-6}$, t = 2.60), indicating that learning curves in saccade targeting were steeper for more regular structures.

ILP spreads

Although syntactic regularities led to improved saccade targeting accuracy, this did not lead to more consistent ILPs for any of the word positions (Fig. 5). On the contrary, landing positions were more diverse for more frequent syntax structures (overall, b = 0.01, SE = 0.006, t = 2.03). Combined with the reduced number of corrective saccades, this may indicate that readers had a better idea of where to focus for optimal information uptake, leading to more variance in the case of structures that had been encountered more often (see Discussion).[7]

Impact of repeatedly encountering specific word identities or phrases

As explained in Section 2.3, we determined the number of item repetitions at the level of single words as well as combinations of words (word pairs, triplets and quadruplets). These were found to have an impact to a varying degree.

The viewing time was decreased by repeatedly encountering specific single word identities (b = −1.40, SE = 0.13, t = −11.19) and by repeatedly encountering specific word triplets (b = −17.80, SE = 6.44, t = −2.77); however, no effects were observed for word pairs (b = −2.64, SE = 1.62, t = −1.63) and word quadruplets

[7] Note that we only considered the final three encounters of each structure; so the larger spread of landing positions for more frequent structures is not the result of a larger number of datapoints.

($b = -1.71$, SE = 9.57, $t = -0.18$).

Similarly, repeatedly encountering specific single word identities or word triplets led to fewer fixations (respectively, $b = -0.01$, SE = 0.001, $t = -12.70$ and $b = -0.06$, SE = 0.03, $t = -2.00$). Again however, no such effects were observed for word pairs ($b = -0.01$, SE = 0.01, $t = -0.79$) or quadruplets ($b = -0.02$, SE = 0.04, $t = -0.37$).

No effects in the corrective saccade probability were observed for previously encountering single words ($b = -4.84*10^{-5}$, SE = $4.51*10^{-5}$, $t = -1.07$), word pairs ($b = -1.05*10^{-3}$, SE = $5.85*10^{-4}$, $t = -1.79$), triplets ($b = 1.45*10^{-3}$, SE = $2.32*10^{-3}$, $t = 0.63$) or quadruplets ($b = -2.27*10^{-3}$, SE = $3.44*10^{-3}$, $t = -0.66$).

Replications in Dutch and L2 English

Effects observed in the English monolinguals were also established in the Dutch bilinguals (based on data from the half of the book that was read in Dutch). Regarding word length regularities, we found significant interactions between Encounter and Structure Frequency on the viewing time ($b = 0.25$, SE = 0.06, $t = 4.30$) and number of fixations ($b = 0.002$, SE = $4.62*10^{-4}$, $t = 4.58$), but only a marginally significant interaction in the corrective saccade probability ($b = 6.64*10^{-5}$, SE = $3.55*10^{-5}$, $t = 1.87$). Again, no effect of repeated word length structure encounters on the ILP spread were found ($b = -0.001$, SE = 0.005, $t = -0.21$).

Similar effects were established for syntactic regularities, with interactions between Encounter and Structure Frequency in viewing time, $b = 0.03$, SE = 0.01, $t = 3.75$ (after excluding refixations, $b = 0.02$, SE = 0.005, $t = 3.62$); in number of fixations, $b = 1.37*10^{-4}$, SE = $3.31*10^{-5}$, $t = 4.13$ (after excluding refixations, $b = 1.33*10^{-4}$, SE = $2.42*10^{-5}$, $t = 5.48$); in the rate of corrective saccades, $3.58*10^{-6}$, SE = $1.34*10^{-6}$, $t = 2.68$. No effect of syntax was established in the ILP distribution (overall, $b = 2.92*10^{-4}$, SE = 0.006, $t = -0.05$).

Next, we assessed whether the Dutch readers exhibited statistical learning during their reading of the English half of the book. For word length, we observed interactions between Encounter and Structure Frequency in the viewing time ($b = 0.87$, SE = 40, $t = 2.18$; however, when excluding refixations this particular effect disappeared: $b = 0.14$, SE = 0.26, $t = 0.56$), number of fixations ($b = 0.004$, SE = 0.001, $t = 2.48$; when excluding refixations, $b = 0.002$, SE = 0.001, $t = 2.47$), and in the corrective saccade probability ($2.94*10^{-4}$, SE = $1.47*10^{-4}$, $t = 2.00$). No effect was observed in the ILP spread ($b = -0.006$, SE = 0.006, $t = -0.90$).

For syntax, we observed interactions in L2 English reading between Encounter and Structure Frequency in the viewing time ($b = 0.20$, SE = 0.06, $t = 3.49$; when excluding refixations, $b = 0.01$, SE = 0.004, $t = 2.65$) and number of fixations ($b = 0.001$, SE = $2.60*10^{-4}$, $t = 3.26$; when excluding refixations, $b = 5.60*10^{-4}$, SE = $1.72*10^{-4}$, $t = 3.25$) but not in the corrective saccade probability ($b = 5.98*10^{-6}$, SE = $1.48*10^{-5}$, $t = 0.41$). Yet, replicating L1 English reading, during their reading of the English half of the book Dutch readers showed wider ILP spreads for repeatedly encountered syntax structures ($b = 0.02$, SE = 0.006, $t = 2.14$).

Discussion

Little is known about the extent to which statistical regularities impact the way we move our eyes through the oceans of linguistic information typically encountered during reading. Here we have reported extensive analyses of the GECO book reading corpus (Cop et al., 2017) with the aim to cover considerable ground in this regard. Specifically, we have gauged effects of repeatedly encountering sentence structures both in terms of their low-level visual properties (word length combinations) and in terms of high-level linguistic properties (syntactic structures).

Our results provide first evidence for the notion that readers are sensitive to statistical regularities spanning combinations of words as these unfold within a book. Repeatedly encountering a given combination of word lengths or syntactic categories leads to fewer and shorter fixations. Implicitly learned patterns further prompt more accurate saccade targeting, evidenced by a reduced number of corrective saccades. Critically, we have established that these effects are not simply the result of general task practice, because steeper learning curves were found for structures with a higher overall frequency. Hence, the present results compel us to claim that statistical regularities picked up within the scope of a story affect linguistic processing as well as oculomotor control.

Findings are not entirely in line with prior expectations, however. We expected that regularities would prompt cleaner oculomotor behavior, reflected in a narrower spread of initial landing positions (ILPs). Opposite to this hypothesis, ILP spreads in English readers widened upon an increased number of repeated encounters-an effect merely reflected in a numerical trend for word length regularities, but a significant effect for syntactic regularities. These patterns were replicated in the Dutch-English bilingual readers for the portion of the book read in English. Taken together with the decreased number of corrective saccades, a possible explanation here would be that enhanced syntactic processing spurred enhanced lexical processing (e.g., Snell & Grainger, 2017, for evidence that syntactic representations provide feedback to the lexical level). This in turn caused readers to have a better idea of where to focus for optimal information uptake. It is conceivable that an increased influence from the lexical level would prompt more variance in landing positions, as the identity of one word may be mostly constrained by a different set of constituent letters than that of another word, (e.g., the first few letters provide the most constraint in '*water*', whereas the final few letters provide the most constraint in '*window*'); and such knowledge could be used to a greater extent when lexical candidates are more strongly activated. Caution is in order, however, as this pattern was not observed for the portion of the book read in Dutch by the Dutch-English bilinguals. Whether the scenario outlined above is indeed contingent on the language at hand will have to be tested more directly in future research.

While the above account of an increased ILP spread upon an increased number of previous encounters is admittedly speculative, another factor that more directly connects to the data deserves consideration here. Frequently encountered structures received fewer fixations, which, by mathematical consequence, necessitated longer saccade amplitudes. As longer saccade amplitudes lead to increased saccade targeting error (Cutter, Drieghe and Liversedge, 2017, 2018; McConkie et al., 1988), this may have induced a larger spread of ILPs. Yet, opposite to the former account, this alternative scenario does not accommodate the reduced number of corrective saccades for frequently encountered structures.

We must also reflect on the fact that although we established statistically clear-cut effects, the data as presented in Figs. 2 and 4 are very noisy beyond 5–10 encounters. Two factors are at play here. Firstly, previous studies show that the impact of some statistical regularities reaches an asymptote relatively quickly (e.g., Wang & Theeuwes, 2018; Verstynen & Sabes, 2011). Secondly, the materials used in the present work afforded but a relatively small number of datapoints for structures with more than 5–10 occurrences. This may have precluded the observation of a neat flat line for highly frequent structures.

Given that most effects observed in the English monolingual readers were also depicted by the Dutch-English bilingual readers both in L1 Dutch reading as well as L2 English reading, one pending investigation concerns the question to what extent the impact of these regularities may be transferred across languages. For instance, as oculomotor parameters have mostly been assumed to depend on low-level visual factors (e.g., Schotter et al., 2012), one might reason that the impact of word length regularities should effectively persist when switching from one language to the other. As concerns syntax, given that there is a body of evidence suggesting that syntactic representations are shared between languages (e.g., Declerck et al., 2020; Hartsuiker, Pickering, & Veltkamp, 2004; Vaughan-Evans, Kuipers, Thierry, & Jones, 2014), one might expect that the long-term impact of syntactic regularities on oculomotor control should similarly persist when switching languages—granted that the two languages have at least some syntactic rules in common.

Several factors prevented us from effectively discerning the impact of a language switch within the scope of the present paper. One could argue that if learned regularities are indeed transferred across languages (upon switching halfway through the book), then our Dutch-English bilinguals should have depicted similar patterns in the second half of the book as our English monolinguals. However, this does not take into account the difference in (English) proficiency between these two groups. Thus, to effectively answer this particular question, one would need to compare reading performance in the second half of the book between two balanced groups of bilingual readers: one of which does, and one of which does not make the language switch. We reckon that such tests of language-nonspecificity in statistical learning are an interesting avenue for future research.

Local versus global statistics

Is statistical learning for sentence structures context-specific? The top panels of Figs. 2 and 4 seem to suggest so, as low- and high-frequency structures were initially read with equal ease. If the structures' frequency values in *The Mysterious Affair at Styles* were to correlate with the general frequency of these structures outside the book (we assume that they are, although at this point we have no good means to verify this), and if knowledge of structure probability is amassed globally throughout one's lifetime, then we should have expected an initial difference in the ease of reading low- versus high-frequency structures. The absence of this initial difference suggests that readers may dive into new contexts—in this case, a book—with a relatively clean slate—at least as it pertains to regularities spanning combinations of words.

Why would these types of regularity impact locally but not globally? Indeed, this is in discord with, for instance, the impact of single word frequency, which undoubtedly operates globally and across contexts. From an entropy point of view, it is worth considering that humans use a finite number of words to generate a virtually infinite number of sentences. As a result it must be more difficult to expect a given sentence structure than it is to expect a given word (indeed, probabilities must by mathematical consequence be lower); and upon encountering a sentence structure, it may more quickly fade from memory. The difference between sentence structures and single words in this regard is further illustrated by Fig. 1: here it can be seen that the most frequent structures occurred somewhere in the range of 10 to 40 times in the whole book. The most frequent single word ('*the'*), meanwhile, occurred 3693 times in the whole book. Statistical learning for sentence structures may thus simply be too weak to survive outside the local context.

There is nonetheless the possibility that readers exerted some 'conscious' control here. Knowledge of these subtle statistics may help the reader to adapt to a particular writing style, which may especially benefit the processing of contexts as large as a book. One way to verify this scenario is to let participants read two books in parallel. If structure frequency effects in one book arise, to any extent, independently from frequency values in the other book, this would possibly indicate that top-down factors are at play here.

Implications for models of reading

In general, one may deem it fairly surprising that syntactic regularities were found to impact oculomotor processes at all, considering that prior literature mostly emphasizes a key role for low-level visual factors (e.g., Schotter et al., 2012). As of yet, no single model of reading could accommodate this phenomenon, as saccade error margins generally have been implemented as hardcoded, fixed parameters (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Rayner, & Pollatsek, 2003; Reilly & Radach, 2006; Snell et al., 2018b). Further, as indicated in the Introduction, none of the well-established models are 'learning' models: the ease of lexical activation is assumed to depend on general word frequency values, but those values are not updated when words are repeatedly encountered within a given session or context. On top of that one would need feedback connections from the level of syntactic representations to the lexical level. Thus far, the only model to assume this is the recent OB1-reader model (Snell & Grainger, 2017, 2019; Snell et al., 2018b).

As OB1-reader appears to offer potential for accommodating the present results, we are compelled to outline more concrete modeling perspectives here. OB1 would need to be augmented with two ingredients. Firstly, the ease of activating syntactic representations should depend on adjustable structure frequency values. Secondly, OB1′s saccade target calculations need to be expanded with optimal processing principles. As of yet, saccade target calculations in OB1 depend both on the lengths of upcoming words, as well as the lexical activation of words associated with upcoming positions. This culminates in a preferred target word, and thereafter the ultimate location decision is simplified in the sense that OB1 simply targets the preferred word's center. This final step would need to be substituted for a process where OB1 determines the optimal location for resolving ambuigity. For instance, if OB1 has two lexical candidates associated with the upcoming word position, '*bean*' and '*bear*', these should lead OB1 to target the end rather than the beginning of the word. Meanwhile, the word beginning should be targeted in the case of candidates '*bean*' and '*dean*'.

Such optimal processing principles, combined with aforementioned learning mechanisms, provide OB1 all the key ingredients for accounting for the present effects: frequent syntactic structures are more strongly activated, and consequently provide stronger feedback (i.e., more constraint) to the level of lexical representations. This likely results in a more strongly confined set of more strongly activated words, such as '*bean*' and '*bear*'; and hereafter the optimal processing location is easily determined. Weaker feedback, on the other hand, would imply less constraint and therefore a larget set of weakly activated candidates, e.g., '*hear*', '*bean*', '*bear*' and '*boar*'. This time around the optimal location for resolving ambiguity is not clear; and as a result, OB1 may find itself ending up at a location from where a corrective saccade is warranted. Future modeling endeavors will reveal the plausibility of this account of our effects.

Conclusion

Concluding our story about statistical learning in a story, here we have reported cross-linguistic evidence that language- and oculomotor processing are impacted by both low- and high-level statistical regularities spanning combinations of words. We surmize that corpora such as GECO provide an excellent platform for investigating statistical learning in language comprehension.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

JS was funded by the European Commission, grant H2020-MSCA-IF-2018 833223. JT was funded by a European Research Council (ERC) advanced grant 833029 – [LEARNATTEND]. The authors have no competing interests to declare. The authors would like to thank Marc

Brysbaert, Simon Liversedge and one anonymous reviewer for their excellent comments and suggestions to an earlier version of the paper.

References

Arciuli, J., & Simpson, I. (2011). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*, 286–304.
Bock, K., & Griffin, Z. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General, 129*, 177–192.
Bowers, N., & Poletti, M. (2017). Microsaccades during reading. *PLoS ONE, 12*, e0185180.
Chater, N., & Manning, C. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Science, 10*, 335–344.
Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods, 49*, 602–615.
Cutter, M., Drieghe, D., & Liversedge, S. (2017). Reading sentences of uniform word length: Evidence for the adaptation of the preferred saccade length during reading. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 1895–1911.
Cutter, M., Drieghe, D., & Liversedge, S. (2018). Reading sentences of uniform word length — II: Very rapid adaptation of the preferred saccade length. *Psychonomic Bulletin & Review, 25*, 1435–1440.
Declerck, M., Wen, Y., Snell, J., Meade, G., & Grainger, J. (2020). Unified syntax in the bilingual mind. *Psychonomic Bulletin & Review, 27*, 149–154.
Ducrot, S., Pynte, J., Ghio, A., & Lété, B. (2013). Visual and linguistic determinants of the eyes' initial fixation position in reading development. *Acta Psychologica, 142*, 287–298.
Ehrlich, S., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 22*, 75–87.
Engbert, R., Nuthmann, A., Richter, M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777–813.
Fine, A., Jaeger, T., Farmer, T., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLOS ONE, 8*, e77661.
Gagl, B., Hawelka, S., & Hutzler, F. (2014). A similar correction mechanism in slow and fluent readers after suboptimal landing positions. *Frontiers in Human Neuroscience, 8*, 355.
Giavazzi, M., Sambin, S., de Diego-Balaguer, R., Le Stanc, L., Bachoud-Lévi, A., & Jacquemot, C. (2018). Structural priming in sentence comprehension: A single prime is enough. *PLOS ONE, 13*, e0194959.
Hartsuiker, R., Pickering, M., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science, 15*, 409–414.
Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16*, 262–284.
Kretzschmar, F., Schlesewsky, M., & Staub, A. (2015). Dissociating word frequency and predictability effects in reading: Evidence from coregistration of eye movements and EEG. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1648–1662.
Kuhl, P. (2004). Early language acquisition: Cracking the code. *Nature Reviews Neuroscience, 5*, 831–843.
Ledoux, K., Traxler, M., & Swaab, T. (2007). Syntactic priming in comprehension: Evidence from event-related potentials. *Psychological Science, 18*, 135–143.
McConkie, G., Kerr, P., Reddix, M., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research, 28*, 1107–1118.
McDonald, S., & Shillcock, R. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech, 44*, 295–323.
Radach, R., Heller, D., & Inhoff, A. (1999). Occurrence and function of very short fixation durations in reading. In W. Becker, H. Deubel, & T. Mergner (Eds.). *Current oculomotor research*. Boston MA: Springer.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.
Rayner, K., & Well, A. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3*, 504–509.
Reichle, E., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–526.
Reilly, R., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research, 7*, 34–55.
Saffran, J., & Wilson, D. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*, 273–284.
Schotter, E., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics, 74*, 5–35.
Siegelman, N., Bogaerts, L., Christiansen, M., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Phil. Trans. R. Soc. B. 372*, 20160059.
Snell, J., & Grainger, J. (2017). The sentence-superiority effect revisited. *Cognition, 168*, 217–221.
Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences, 23*, 537–546.
Snell, J., Grainger, J., & Declerck, M. (2018a). A word on words in words: How do embedded words affect reading? *Journal of Cognition, 1*, 40.
Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018b). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological Review, 125*, 969–984.
Thompson, S., & Newport, E. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*(1), 1–42.
Tooley, K., & Traxler, M. (2010). Syntactic priming effects in comprehension: A critical review. *Language & Linguistic Compass, 4*, 925–937.
Vaughan-Evans, A., Kuipers, J., Thierry, G., & Jones, M. (2014). Anomalous transfer of syntax between languages. *Journal of Neuroscience, 34*, 8333–8335.
Verstynen, T., & Sabes, P. (2011). How each movement changes the next: An experimental and theoretical study of fast adaptive priors in reaching. *Journal of Neuroscience, 31*, 10050–10059.
Vitu, F., & McConkie, G. (1998). On regressive saccades in reading. In G. Underwood (Ed.). *Eye guidance in reading and scene perception* (pp. 101–124). Oxford: Elsevier.
Wang, B., & Theeuwes, J. (2018). Statistical regularities modulate attentional capture. *Journal of Experimental Psychology: Human Perception & Performance, 44*, 13–17.
White, S., Warrington, K., McGowan, V., & Paterson, K. (2015). Eye movements during reading and topic scanning: Effects of word frequency. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 233–248.