Visual attention and cross-linguistic effects in reading: Simulations with BRAID-Acq, a probabilistic model of reading

Camille Charrier (camille.charrier@univ-grenoble-alpes.fr)
Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

Alexandra Steinhilber (alexandra.steinhilber@univ-grenoble-alpes.fr)
Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

Emilie Ginestet (emilie.ginestet@univ-grenoble-alpes.fr)
LaRAC, Université Grenoble Alpes, Grenoble, France

Sylviane Valdois (sylviane.valdois@univ-grenoble-alpes.fr)
Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

Julien Diard (julien.diard@univ-grenoble-alpes.fr)
Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

Abstract

Theories of reading are mostly based on English, that is rather atypical among alphabetic orthographies due to its inconsistent orthography-phonology mappings. Differences in length effects have been observed between languages. Psycholinguistic characteristics of the orthography, such as orthographic depth, seem to have an impact on reading strategies and could be correlated with different visual-attentional profiles. However, no computational model has yet demonstrated the impact of language-dependent visual-attentional mechanisms on reading. This study explores these effects using BRAID-Acq, a probabilistic reading model with a visual-attentional module. We simulated word and pseudoword reading in English, German, and French to examine how orthographic depth and visual attention shape processing. Our simulation results suggest an effect of the orthography on processing time. In particular, English requires a larger attentional quantity for efficient processing of words and pseudowords, offering a novel interpretation of difficulties in reading acquisition in English.

Keywords: Cross-linguistic effects; Visual attention; Probabilistic modeling; Grain size theory; Reading

Introduction

Regardless of language, readers face the challenge of analyzing the written form of words in order to infer their meaning and their corresponding spoken form. This complex activity has been the subject of numerous studies aimed at identifying the cognitive processes involved in expert reading and in learning to read. Although most studies were initially confined to the English language, they led to the proposal of theoretical models that were supposed to account for the cognitive system of reading in general (Frost, 2012). However, cross-linguistic research has shown that the English orthography is rather atypical among alphabetic systems. Further, numerous studies have reported differences in the manifestation of behavioral effects, such as the length effect, across languages.

The length effect in reading is the overall observation that it is easier or faster to process shorter words than longer ones (for an entry point, see New et al., 2006). It has been reported in many tasks, including lexical decision and reading aloud (Barton et al., 2014). These effects have been shown to impact both timing measurements, such as lexical decision reaction time or naming latencies, and eye movements, such as gaze duration and number of fixations. Furthermore, the length effect is known to interact with other classical behavioral effects, such as word familiarity. Specifically, it is more pronounced for pseudowords than for words (Marinelli et al., 2016; Perry & Ziegler, 2002; Rau et al., 2015).

Studies investigating the length effect in reading have been conducted in many alphabetic languages, ranging from the most consistent ones (languages with more consistent grapheme-phoneme mappings, such as Finnish, Spanish, and German), to the least consistent ones (languages with more inconsistent grapheme-phoneme mappings, such as French and English). In fact, the magnitude of the length effect appears to vary across languages. Cross-linguistic studies have indicated that this length effect is stronger in the most consistent orthographies, such as German, as compared to English, which is widely regarded as one of the least consistent of the alphabetic languages. For instance, Ziegler et al. (2001) compared naming latencies in German and English expert readers, and reported that length effects were more pronounced for pseudowords than for words, and more pronounced in German than in English.

These cross-linguistic findings are consistent with the Orthographic Depth Hypothesis (ODH; Frost et al., 1987; Katz & Frost, 1992), which suggests that reading performance is affected by orthographic depth. Readers of a shallow orthography would be more likely to rely on non-lexical processing, based on grapheme-phoneme correspondences, since the mapping from graphemes to phonemes is unambiguous and typically produces the correct word pronunciation. In contrast, due to the high prevalence of irregular grapheme-phoneme correspondences, the use of morphological and lexical processes would be necessary when reading an opaque orthography. This hypothesis is supported by the difficulties that readers of opaque orthographies, such as English, have

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

in learning to read, compared to readers of more shallow orthographies (Seymour et al., 2003). This also leads to differences in the observed length effects. For instance, Perry & Ziegler (2002) simulated in the DRC model (Coltheart et al., 2001) the difference in length effect between English and German by changing the speed balance between the lexical and non-lexical pathways. The observed word length effect in German and the greater length effect for pseudowords in German than in English were successfully simulated when sublexical processing was speeded up in the German model. The differential length effects observed in the two languages were thus interpreted as resulting from a greater reliance on the phonological pathway (based on grapheme-phoneme conversions) in German than in English, within the dual-route framework of reading.

These findings are also consistent with the Grain Size Theory (Ziegler & Goswami, 2005), which focuses instead on the difference in the size of sublexical units (both orthographic and phonological) according to language opacity. This theory postulates that readers of consistent languages would rely on smaller units for phonological decoding, while larger units would be needed in inconsistent languages. The reliance on smaller grain sizes in consistent orthographies would result in stronger length effects, whereas the use of larger grain sizes in inconsistent orthographies would favor whole-word processing, leading to smaller length effects. As proposed by Ziegler & Goswami (2005), the use of appropriate grain sizes would be influenced by several factors, including the advantage of small units that are orthographically less complex, the advantage of large units for fast processing, and the statistical regularities between orthographic and phonological units.

However, the classical version of the Grain Size Theory does not specify the cognitive processes that would select the appropriate grain size within words, depending on both language and reader expertise. An extension of this theory applied to bilingualism, the Grain Size Accommodation Theory (GSAT), proposed such a hypothesis (Lallier & Carreiras, 2018). According to this theory, learning to read in opaque orthographies would promote the use of larger visual grain sizes, compared to learning to read in shallow orthographies. Shallow orthographies typically consist of single-letter graphemes, whereas opaque orthographies tend to feature multiple-letter or contextual graphemes.

Assuming that multi-letter parallel processing is controlled by visual attention (Bogon et al., 2014; Lobier et al., 2013), the GSAT postulates that acquiring proficiency in reading inconsistent orthographies would necessitate a larger visual attention spread over the words. From a behavioral perspective, the extent of visual attention available for multi-letter processing is estimated by measuring the visual attention span (VAS; Frey & Bosse, 2018). Research on the role of visual attention in reading has been extensive, and a growing body of studies suggests that it may vary across languages. In the GSAT framework, visual attention dispersion could then be hypothesized to be the cognitive mechanism explaining crosslinguistic differences, such as the different length effects reported between shallow and opaque orthographies. As the word is made of multi-letter, often contextual, graphemes, reading in an opaque orthography requires the reader to spread its visual attention over words, resulting in a larger VAS and higher attentional demand of the task, compared to reading in a shallow orthography.

However, we lack direct evidence in support of this hypothesis. We also lack predictions from computational models of reading. In fact, dual-route computational models of reading aloud (Coltheart et al., 2001; Perry et al., 2007, 2010) do not include any flexible visual attentional component and can only account for cross-linguistic effects in the context of the ODH. In contrast, a sophisticated model of visual attention was implemented in the framework of the BRAID family of models (Ginestet et al., 2019, 2022; Steinhilber, 2023; Valdois et al., 2021; Phenix et al., 2025), but no simulations attempted to compare the model's predictions for languages that differed in orthographic depth.

The purpose of this paper is to assess the ability of the BRAID-Acq model to account for differences in length effects during reading in three languages varying in orthographic depth. Since attentional resources are directly linked to the VAS, we predict that opaque orthographies, especially English, will require more attentional resources than shallow orthographies, such as German, with French between English and German. This would lead to a larger effect of attentional resources on length effects for English than for French, and larger effects for French than for German. In the rest of this paper, we first describe the most salient features of the BRAID-Acq model for the current study, then we describe the simulations we performed, and the experimental results we obtained.

Method

BRAID-Acq model

The BRAID-Acq model (for "Bayesian word Recognition with Attention, Interference and Dynamics for reading Acquisition") is a single-route Bayesian model of reading aloud and novel word learning. It is the most recent extension of the BRAID family of models (Steinhilber, 2023). In this section, we focus on a brief conceptual description of the model. For more details, see Steinhilber (2023) and Phenix et al. (2025).

The model consists of two relatively symmetrical parts (see Figure 1): an orthographic branch and a phonological branch. The first branch is purely orthographic, in that it only features knowledge about words and the corresponding letters, and letter processing and perception. The second branch is purely phonological, with knowledge about words and the corresponding phonemes, and phoneme processing. These two branches are connected at the lexical level, which bridges the orthographic and phonological information of each known word. The model places emphasis on the role of attention in reading, which modulates information propagation in the model, and simulates visual-attentional exploration.

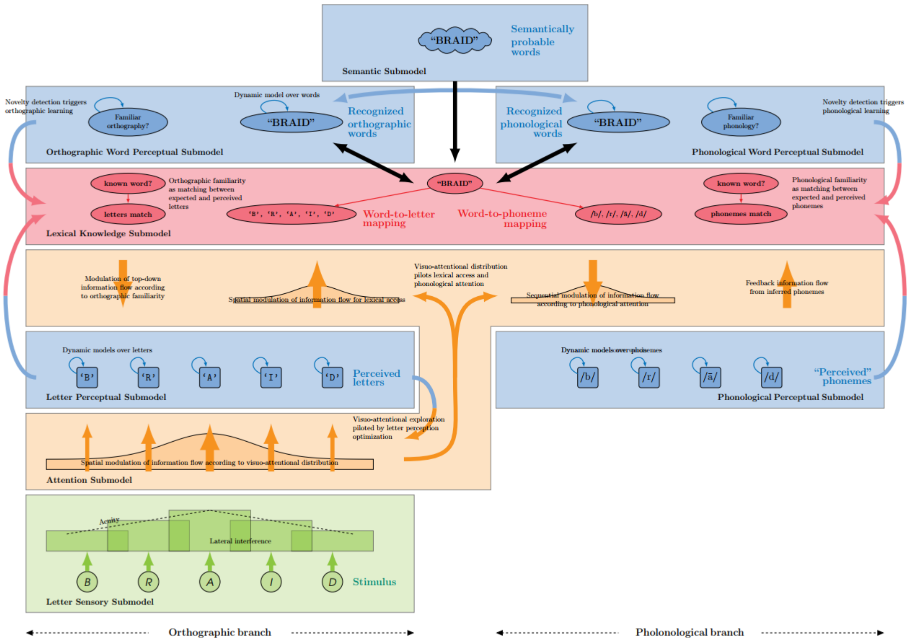

Figure 1: Schematic representation of the BRAID-Acq architecture. See text for details. Adapted from (Steinhilber, 2023).

The model architecture is organized into five submodels: sensory, perceptual, lexical, semantic and attentional.

The first submodel is the letter sensory submodel (in green, Figure 1). The input is visual and consists of a sequence of letters; sensory processing simulates properties of the visual system, such as acuity, lateral interference between letters, and letter visual similarity. The second submodels are the perceptual submodels of letters and phonemes (in blue, middle of Figure 1). These two perceptual submodels concern the dynamic accumulation of information over time about the identity of letters and phonemes associated with the input stimulus. They represent the model's knowledge about letters in the stimulus and of the sequence of phonemes corresponding to the perceived stimulus. In the absence of incoming information flow, the accumulated perceptual evidence degrades over time, modeling memory decay. The lexical knowledge submodel (in red, Figure 1) interfaces the letter and phoneme sequence spaces with the word space. It contains the lexical knowledge of the model, i.e., the model's orthographic and phonological lexicons. If a word is known to the model, the lexical submodel encodes the sequence of letters and phonemes of that word. Perceptual submodels of orthographic and phonological words (in blue, top of Figure 1) enable the lexical identification of the input stimulus. They function in the same way as evidence accumulators (as in the perceptual submodels of letters and phonemes) and accumulate information over time about the identity of the word being processed by comparing the identity of the stimulus's letters and phonemes with its lexical knowledge. The model also features a semantic submodel; however, since it is deactivated in the current simulations where we consider isolated word reading, this submodel is not described here.

The attentional submodel (in orange, Figure 1) can be broken down into three parts, driven simultaneously by the same attentional distribution. This distribution is spatial, and favors processing in attended positions to the detriment of the rest. Attention influences: 1) the accumulation of perceptual evidence about letters by modulating the spatial distribution of the *bottom-up* flow of information from the letter sensory submodel to the letter perceptual submodel; 2) lexical access by modulating the flow of information between the letter perceptual submodel and the lexical knowledge submodel; 3) phonological percept computation by modulating the flow of information between the lexical knowledge submodel and the phoneme perceptual submodel. The attention distribution is mathematically defined by a Gaussian probability distribution. Therefore, it is defined by two parameters: its mean μ_A, which corresponds to the position of the visual fixation on the word, and its variance σ_A^2, which corresponds to the dispersion of attention around μ_A. The model considers a third parameter, Q_A, which is a multiplicative factor that represents the total amount of available attentional resources (default value 1).

Controling BRAID-Acq's attentional distribution over time enables a visual-attentional exploration of the stimulus in several attentional fixations, with the algorithm that follows (Steinhilber et al., 2023). The first fixation is performed according to the default attentional parameters ($\sigma_A^2 = 1.75$). It lasts for a fixed number of iterations. At this point, the variance σ_A^2 of the attentional distribution is computed for the rest of the simulation: the information accumulation speed over the percept is compared to a "prototypical" speed, calculated on a reference set of words from the lexicon. If the model considers that the information gain is slower than the prototype speed, then the variance σ_A^2 is decreased for the rest of the simulation, modeling the shift to a more serial reading. On the contrary, if the information gain is faster than the prototype speed, then the variance σ_A^2 is increased, allowing the simultaneous processing of more letters.

The exploration algorithm then selects positions for attentional fixations, according to a criterion based on the information accumulated about letters in each position, measured by the entropy of the perceptual probability distributions. A stopping criterion, based on the entropy of probability distributions over letters, enables the exploration process to be terminated when the letters are, on average, well perceived.

Overall, one of the main factors affecting the speed of information acquisition, and thus the dispersion of visual attention in the model, is the amount of attentional resources Q_A. A high Q_A value generally induces fast processing and therefore a wide attentional distribution. Conversely, a low Q_A value typically induces an attentional distribution focused on a reduced number of letters. Thus, by modulating the quantity of attentional resources (Steinhilber et al., 2023), BRAID-Acq is able to simulate a more serial reading (with a small Q_A) or a more parallel reading (with a high Q_A). Therefore, the attentional quantity Q_A could be interpreted as reflecting reading expertise, and it could be expected to vary with the child's development during reading acquisition.

Materials

Model configuration In this study, we examined the reading performance of three BRAID-Acq models, each configured with a different language: English, German and French. The lexical knowledge of these models was taken from lexical databases available online: from the Celex database for English and German (Baayen et al., 1996) and from the Lexique database for French (Lexique383_fr; New et al., 2004). Due to differences in size between the three databases (approximately 41,000, 51,000 and 30,000 words, respectively) and to avoid any lexicon size effect on the model's performance, the original lexicons were evened out in size by random sampling 25,000 words, after removing acronyms and abbreviations.

Opacity measure To ensure that the sampled lexicons accurately reflected the original opacity of each language, we calculated opacity measures with Borgwaldt et al. (2004)'s methodology: we computed the entropy of the initial letter-phoneme (and the initial phoneme-letter) mappings for each word, and weighted this value by the frequency of the initial letter (or phoneme). We then calculated the averaged opacity value for all words in each language. The calculated values (Figure 2) are not exactly the same as those reported by Borgwaldt et al. (2004). However, the relative positions of languages on the opacity scale is similar: the English lexicon is more opaque than the French lexicon, which in turn is more opaque than the German lexicon.

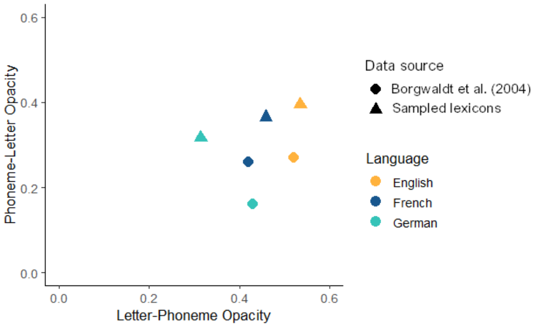

Figure 2: Opacity of English (yellow), French (blue) and German (teal) lexicons, from Borgwaldt et al. (2004) data (circles) and sampled lexicons used for this study (triangle). The axes are measures of the opacity of the phoneme to letter (*y*-axis) and letter to phoneme (*x*-axis) mappings.

Stimuli The simulations were performed on a subset of 1,400 words of 4 to 10 letters in each language. In each language, the words were selected by random sampling 200 words per length, in the frequency range between 2 and 1,000 occurrences per million.

Simulated task To investigate the effect of orthographic depth on reading within the BRAID-Acq model, we compared reading performance for words and pseudowords in all three languages (German, French and English). In the word condition and for each language, the orthographic and phonological forms of the words belonged to the model's lexical knowledge (i.e., the stimuli are known words to the model); in the pseudoword condition, we used the same words but their orthographic and phonological forms were removed from the model's lexical knowledge (i.e., the stimuli are pseudowords for the model). This allowed us to compare word and pseudoword reading for items that were otherwise strictly identical.

For the model, the reading task consists in exploring the input letter sequence through several visual-attentional fixations until either the model considers that the stimulus is perceived well enough (when the entropy of the orthographic percept is below a fixed threshold), or a maximum of 2,500 iterations was reached. The model's processing time for each word was measured in number of iterations, with a maximum of 2,500. Whatever the reason for processing termination, in the reading task, the output of the model is the sequence of phonemes with highest probability at each position.

For the two types of items (words and pseudowords), reading was considered successful if the simulated phonological output was identical to the pronunciation of the known word. Although this method of scoring is classical in empirical assessment of word reading, it differs from the scoring typically used for pseudowords. Indeed, behaviorally, the usual rule is to consider as correct any pronunciation that is plausible for the pseudoword, meaning that the pronunciation results from the use of an existing grapheme-phoneme mapping and in accordance to contextual rules. In opaque languages, it is thus typical that several pronunciations are scored as correct for the same pseudoword. Our scoring method was far more stringent, since the pseudoword pronunciation was scored as correct if and only if it was strictly identical to that of the corresponding word.

To investigate the role of visual attention on processing in each language, the simulations were repeated for a set of 4 values of the attention quantity $Q_A \in [0.5, 1, 1.5, 2]$. All other parameters of the model were identical for each simulation and all languages, and were set to their default values (see Steinhilber, 2023).

Results

Overall, all known words were read, either correctly or incorrectly, within the time limit of 2,500 iterations. However, this was not the case for pseudowords, which is indicative of difficulty in processing the stimulus. All Q_A values combined, the maximum duration of 2,500 iterations was attained for 18.5% (648) of the pseudowords in English, 7.5% (264) in French and 4.8% (167) in German. However, this occurred predominantly for the lowest Q_A values, particularly for $Q_A = 0.5$. Note that the results of the simulations reported below concern all items, including pseudowords that reached maximum duration. The model's global success rate for words was 83.0% for English, 82.8% for French and 87.9% for German. For pseudoword, the success rate was 47.7% for English, 63.3% for French and 57.5% for German.

The statistical analyses we conducted for this study are focused on processing time. We used a gamma generalized linear model with a log link (*glm* function in R; R Core Team, 2024), given the assumption that processing time is a continuous, positive, and asymmetric variable. The model incorporated the impact of language, word length (quantified by the number of letters), lexical status (differentiating between word and pseudoword) and attentional quantity Q_A. As this analysis does not include an in-depth study of errors, all items, whether erroneous or successful, were included in the subsequent analyses.

To start our analysis, we examined targeted main effects, to confirm that the model correctly processed the stimuli. Firstly, it was hypothesized that word length would have an effect on processing time, with longer words requiring more time to process. The results showed a significant main effect of length on processing times ($b = 0.121, z = 62.524, p < 0.001$), indicating that the model processed shorter words more rapidly, as predicted.

Secondly, we postulated that the increase of attentional capacities during child development could be mathematically represented by an increase in attentional resource Q_A. We thus predicted that an increase in Q_A would lead to faster processing. The analysis of processing time yielded to a significant main effect of Q_A ($b = -0.822, z = -96.850, p < 0.001$), indicating that the greater the attentional resource available to the model, the faster it processes.

Finally, given the predicted impact of language opacity on reading, we hypothesized a global language effect, whereby English would be processed more slowly than French, and French more slowly than German. Our findings revealed an overall main effect of language on processing time ($b = -0.376, z = -21.795, p < 0.001$). The contrast analysis confirmed that English was processed significantly more slowly than French ($b = 0.352, t(20976) = 90.768, p < 0.001$), and French in turn was processed significantly more slowly than German ($b = 0.084, t(20976) = 21.712, p < 0.001$).

We conducted further analyses, beyond these primary hypotheses. Recall that we reasoned that the length effect difference between opaque and shallow orthographies could be explained, at least partially, by the greater need for attentional dispersion for opaque languages, which would result in a higher demand for attentional resources during reading. Our simulation results showed a significant interaction between language, length and Q_A on processing times ($b = -0.008, z = -5.895, p < 0.001$). Figure 3 provides a visual representation of processing times for each language, as a function of length, item type and Q_A. Given the observation that small Q_A values induced a pronounced ceiling effect, related to the maximum duration simulation constraint, the analysis was repeated excluding $Q_A = 0.5$. This replicates the significant language, length and Q_A interaction ($b = 0.011, z = 6.357, p < 0.001$); the rest of our analysis relies on this second result.

As illustrated in Figure 3 (compare the slopes of processing times for different colored lines), increasing Q_A results in a reduction of the length effect. If increasing Q_A is interpreted as increasing reading expertise, this aligns with the observation that expert readers exhibit smaller length effects compared to beginner readers. However, the magnitude of this interaction varies with language. German and French benefited less from the increase of Q_A on the decrease of length effect on reading, in particular for reading pseudowords, than English. Indeed, the interaction between language, length, Q_A and lexicality was also significant ($b = 0.010, z = 3.591, p < 0.001$).

Our results also yield non-attentional crosslinguistic differences on length effects. Specifically, results revealed a significant interaction between language and length on processing time ($b = 0.023, z = 9.788, p < 0.001$). A contrast analysis confirmed that German produced a larger length effect than French ($b = -0.007, t(20976) = -3.747, p < 0.001$), and French produced in turn a larger length effect than English ($b = -0.014, t(20976) = -7.160, p < 0.001$). However, again, the ceiling effect for small Q_A values may modify these effects. A replication of this analysis, while excluding $Q_A = 0.5$, replicated the effect ($b = -0.017, z = -5.209, p < 0.001$) but shuffles the language order, with a larger length effect for English than French ($b = 0.005, t(16776) = 2.891, p = 0.011$), and in turn a larger length effect for German than for English ($b = -0.006, t(16776) = -2.962, p = 0.009$).

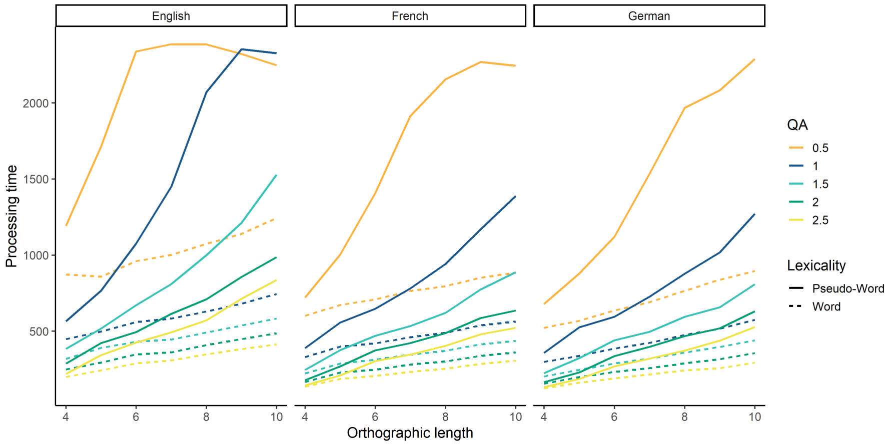

Figure 3: Processing times (y-axis) as a function of length (x-axis), for all languages (left panel: English; middle panel: French; right panel: German), for words (dashed lines) and pseudo-words (solid lines), as a function of the quantity of attention Q_A (line color).

Discussion

Our study was conducted to explore the ability of a computational model implementing a sophisticated visual-attentional module to replicate crosslinguistic modulations of length effect in reading aloud. The theoretical framework of the Grain Size Accomodation Theory (GSAT) proposed that visual attention may be directly associated with the visual grain size used in reading, with large grain size in opaque orthographies, and small grain size in shallow orthographies. However, so far, this theoretical proposal lacked evidence, either experimental or computational, to support it. We used the recent BRAID-Acq model to investigate this theoretical proposal.

Firstly, the model replicated the predicted main effects. It reproduced the length effect, as observed in the literature: processing time was larger for long words than for short words. Additionally, it also exhibits an overall language effect, with English yielding longer processing times. This is consistent with the widely accepted fact that, as one the most opaque languages, learning to read is particularly difficult for English readers. In the model, we found a specific difficulty for the model to process English, in particular for small attentional quantity, compared to French and German.

The present simulations are fully consistent with the Orthographic Depth Hypothesis (ODH). This hypothesis explains the behavioral differences between languages, and in particular in terms of length effects, by the orthographic depth difference between those languages. Authors have assumed that this would be due to a stronger reliance on sublexical processes for shallow orthographies compared to opaque orthographies, which would instead rely more on lexical processes. Relying on the relative properties of lexical and sublexical routes only makes sense in the dual-route models of reading. In contrast, our study is based on BRAID-Acq, a single-route model of reading, which nevertheless reproduces the expected overall effect of language. This finding calls into question the widely accepted dual-route interpretation of the ODH.

The most salient result of our study is the demonstration that visual attention can modulate the length effect in different ways across languages. Simulation results indicated that the modulation of attentional quantity on length effects differed across languages. Specifically, the benefit of an increase of Q_A was found to be more pronounced in English compared to French and German. Consequently, reduced attentional resources are more detrimental to English than to French and German. This evidence provides substantial support for the GSAT and suggests novel explanations for the pronounced difficulties experienced by young English-speaking children in acquiring literacy. If English-speaking children of the same age as German or French children, with the same attentional resources, have a linguistic pressure to disperse more their attentional resources during word and novel word reading, this could result in longer processing time and overall reading difficulties, thereby leading to excessive length effects. How limited attentional resources interact, during learning, with other linguistic characteristics that we have not considered here (e.g., morphological complexity, frequency, neighborhood density), across various languages, is an open question.

This study is a first step in investigating the involvement of attentional processes in cross-linguistic effects. However, it features a number of technical limitations, which could easily be overcome in future simulations. First, the maximum number of iterations (set to 2,500) yielded significant ceiling effects, especially for low Q_A values in English. Second, we used a visual-attentional exploration algorithm that induced a link between the amount of attention Q_A and its dispersion σ_A^2: decoupling them would allow investigating their specific roles on cross-linguistic effects. Finally, considering a more extensive array of languages would allow a more thorough study of the impact of psycholinguistic characteristics on cross-linguistic effects.

References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). *The celex lexical database.* University of Pennsylvania. (Web access: `http://celex.mpi.nl/`)

Barton, J. J. S., Hanif, H. M., Björnström, L. E., & Hills, C. (2014, August). The word-length effect in reading: A review. *Cognitive Neuropsychology*, *31*(5-6), 378–412. (Publisher: Routledge _eprint: https://doi.org/10.1080/02643294.2014.895314) doi: 10.1080/02643294.2014.895314

Bogon, J., Finke, K., Schulte-Körne, G., Müller, H. J., Schneider, W. X., & Stenneken, P. (2014). Parameter-based assessment of disturbed and intact components of visual attention in children with developmental dyslexia. *Developmental Science*, *17*(5), 697–713. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/desc.12150) doi: 10.1111/desc.12150

Borgwaldt, S. R., Hellwig, F. M., & de Groot, A. M. (2004). Word-initial entropy in five languages: Letter to sound, and sound to letter. *Written Language & Literacy*, *7*(2), 165–184. doi: 10.1075/wll.7.2.03bor

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). Drc: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. doi: 10.1037/0033-295x.108.1.204

Frey, A., & Bosse, M.-L. (2018). Perceptual span, visual span, and visual attention span: Three potential ways to quantify limits on visual processing during reading. *Visual Cognition*, *26*(6), 412–429.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, *35*(5), 263–279.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: a multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 104–115. doi: 10.1037/0096-1523.13.1.104

Ginestet, E., Phénix, T., Diard, J., & Valdois, S. (2019). Modeling the length effect for words in lexical decision: The role of visual attention. *Vision Research*, *159*, 10–20.

Ginestet, E., Valdois, S., & Diard, J. (2022). Probabilistic modeling of orthographic learning based on visuo-attentional dynamics. *Psychonomic Bulletin & Review*, 1–24. doi: 10.3758/s13423-021-02042-4

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In *Orthography, phonology, morphology, and meaning* (Vol. 94, p. 67-84). Advances in Psychology, North-Holland. doi: 10.1016/S0166-4115(08)62789-2

Lallier, M., & Carreiras, M. (2018). Cross-linguistic transfer in bilinguals reading in two alphabetic orthographies: The grain size accommodation hypothesis. *Psychonomic bulletin & review*, *25*(1), 386–401. doi: 10.3758/s13423-017-1273-0

Lobier, M., Dubois, M., & Valdois, S. (2013). The role of visual processing speed in reading speed development. *PloS one*, *8*(4), e58097.

Marinelli, C. V., Romani, C., Burani, C., McGowan, V. A., & Zoccolotti, P. (2016). Costs and benefits of orthographic inconsistency in reading: Evidence from a cross-linguistic comparison. *Plos One*, *11*(6), e0157457. doi: 10.1371/journal.pone.0157457

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*(1), 45–52.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new french lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516–524.

Perry, C., & Ziegler, J. C. (2002). Cross-language computational investigation of the length effect in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 990–1001. doi: 10.1037/0096-1523.28.4.990

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the cdp+ model of reading aloud. *Psychological Review*, *114*(2), 273–315. doi: 10.1037/0033-295X.114.2.273

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (cdp++) model. *Cognitive psychology*, *61*(2), 106–151.

Phenix, T., Ginestet, É., Valdois, S., & Diard, J. (2025). Visual attention matters during word recognition: A bayesian modeling approach. *Psychonomic Bulletin & Review*, 1–39.

R Core Team. (2024). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and english children and adults process the same words differently. *Journal of experimental child psychology*, *130*, 92–105. doi: 10.1016/j.jecp.2014.09.012

Seymour, P. H., Aro, M., Erskine, J. M., & with COST Action A8 Network, C. (2003). Foundation literacy acquisition in european orthographies. *British Journal of Psychology*, *94*(2), 143–174. doi: 10.1348/000712603321661859

Steinhilber, A. (2023). *Modélisation bayésienne de l'apprentissage de la lecture* (Unpublished doctoral dissertation). Université Grenoble Alpes (ComUE).

Steinhilber, A., Diard, J., Ginestet, E., & Valdois, S. (2023). Visual attention modulates the transition from fine-grained, serial processing to coarser-grained, more parallel processing: A computational modeling study. *Vision Research*, *207*, 108211.

Valdois, S., Phénix, T., Fort, M., & Diard, J. (2021). Atypical viewing position effect in developmental dyslexia: A behavioural and modelling investigation. *Cognitive Neuropsychology*, *38*(5), 319–335.

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. doi: 10.1037/0033-2909.131.1.3

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12*(5), 379–384. doi: 10.1111/1467-9280.00370