UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
The Role of Information in Visual Word Recognition:A Perceptually-Constrained Connectionist Account

Permalink
https://escholarship.org/uc/item/8g33n2b8

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors
Alhama, Raquel G.
Siegelman, Noam
Frost, Ram
et al.

Publication Date
2019

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

The Role of Information in Visual Word Recognition: A Perceptually-Constrained Connectionist Account

Raquel G. Alhama (rgalhama@mpi.nl)
Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

Noam Siegelman (noam.siegelman@yale.edu)
Haskins Laboratories, 300 George street, New Haven, CT, USA

Ram Frost (ram.frost@mail.huji.ac.il)
Department of Psychology, The Hebrew University, Jerusalem, Israel

Blair C. Armstrong (blair.armstrong@utoronto.ca)
Department of Psychology, University of Toronto, 1265 Military Trail, Toronto, ON, Canada

Abstract

Proficient readers typically fixate near the center of a word, with a slight bias towards word onset. We explore a novel account of this phenomenon based on combining information-theory with perceptual constraints in a connectionist model of visual word recognition. This account posits that the amount of information-content available for word identification varies across fixation locations and across languages. These differences contribute to the overall fixation location bias in different languages, make the novel prediction that certain words are more readily identified when fixating at an atypical fixation location, and predict specific cross-linguistic differences. We tested these predictions across several simulations in English and Hebrew, and in a behavioral experiment. The results confirmed that the bias to fixate closer to word onset aligns with reducing uncertainty in the visual signal, that some words are more readily identified at atypical fixation locations, and that these effects vary across languages.

Keywords: visual word recognition; computational modelling; connectionism; information theory; fixation location

Introduction

The fundamental aim of visual word recognition is to identify a word based on its constituent letters. Considerable computational and behavioral evidence from studying isolated visual word recognition, which typically involves seeing a single word presented at the center of visual fixation, suggests that a graded constraint satisfaction process selects a candidate that fits with the lower-level (visual/orthographic) and higher level representations (e.g., lexical information, McClelland & Rumelhart, 1981). In contrast to this methodology, in more naturalistic studies of reading via eye-tracking, considerable evidence suggests that readers tend to fixate more frequently near the middle of words, typically with a bias towards beginning of a word, with some variation across languages (see Figure 1, for example distributions from initial fixations during natural reading in English and Hebrew Siegelman et al., 2019).

A key question from considering this body of work, then, is how and why the visual system of a proficient reader tends to fixate at particular positions in a word, and on a related front, why these fixation distributions vary as a function of the language. Classic accounts focused on the low-level operations of the occulomotor system do not appear to offer a ready explanation of these effects, particularly in terms of cross-linguistic differences (see McConkie, Kerr, Reddix, Zola, & Jacobs, 1989 for a review of occulomotor theories; also Reichle, Rayner, & Pollatsek, 1999; Engbert, Nuthmann, Richter, & Kliegl, 2005). Accounts that hold more promise in this regard consider higher-level factors (e.g., morphology; Deutsch & Rayner, 1999).

Here, we explore an alternative more general account based on information theory in the visual signal and how it maps onto lexical representations. This work shares some conceptual similarity with prior work by Brysbaert and Nazir (2005), although the latter did not quantify information in the formal terms that we do, which may, as outlined in the discussion, explain some discrepancies between their results and ours. In our first study we examined the differences in information distributions as a function of fixation location in Hebrew and English, and found that these distributions shared key characteristics of the human fixation location distributions. In our second study, we instantiated a feed-forward connectionist model with a psychophysically-derived constraint on letter identification as a function of distance (eccentricity) from the target fixation. This model allowed us to examine how different amounts of information content can be extracted at different fixation locations in different languages during word recognition. If it succeeded in doing so, it could explain why there is a preferred fixation location in different languages due simply to how low-level constraints interact when identifying a word, in the absence of higher-level constraints (e.g., morphology, semantics). This model also served as a test-bed for probing whether words exist in different languages that, due strictly to the information content available at different fixation locations, are, perhaps counter-intuitively, more efficiently recognized by looking at fixation locations other than the overall preferred location in the language. These predictions were corroborated in a pilot behavioral experiment. Taken together, this research highlights how maximizing information in the visual signal could be a major driver of many behaviors observed within and between languages. It also offers specific predictions for broadening this account in future work, for instance, in maximizing information across words

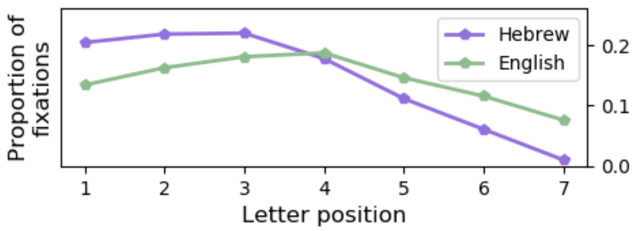

Figure 1: Distribution of fixation locations for 7-letter words in English and Hebrew. 1=start of word (left in English, right in Hebrew)

rather than within the processing of single words.[1]

Study 1: Information-content in the early visual-orthographic representation.

In our first study, we explored how much information-content was available for word recognition in the early visual-orthographic signal when fixating at different locations in the word. This was achieved computationally by passing the visual representation of a word at a particular fixation location through a visual filter that reflects how more visual information is extracted from the fixated location in a word and less information is extracted as a function of eccentricity (distance) from this location. We applied this procedure to samples of words from English and Hebrew, which belong to different language families, to gain insight into the language-specific versus language-general nature of the results.

Data We analyzed the 50,000 highest frequency words from the OpenSubtitles translated movie subtitle database (Tiedemann, 2012)[2]. We removed all words that contained foreign alphabet characters. For simplicity, we selected for our study only 7-letter words, because we predicted that strong effects of fixation location and information content would be more readily detected in longer words that could nevertheless be perceived with a single fixation. The resulting lists contains 5565 words in Hebrew, and 8145 in English.[3]

Architecture To simulate the constraints on visual perception imposed by the early visual perception system, we passed the representation of each word in each language through perceptual filters adapted from McConkie et al. (1989). In the original formulation of this model of perceptual filtering, the fixated letter was perceived with 100% accuracy, and the likelihood of successful perception fell off linearly as a function of eccentricity (see Figure 2, for examples from fixating letter 2 or letter 6 in a 7-letter word). The exact slope of this function, as exemplified by the $drop = 0.1$ and $drop = 0.25$ lines in the figure, leads to an initial linear change in the amount of extracted information, which eventually reaches floor.

In the original paper, the authors noted that the optimal value of the *drop* parameter remained to be determined. Thus, for this initial work, we opted to use a *drop* parameter of 0.25. This value was selected so as to capture most but not all the letters in a word when perceived from the start or end of the word, which we predicted would lead to relatively high, but below ceiling, recognition rates (confirmed and described in a later section) and substantial differences in information as a function of fixation location.

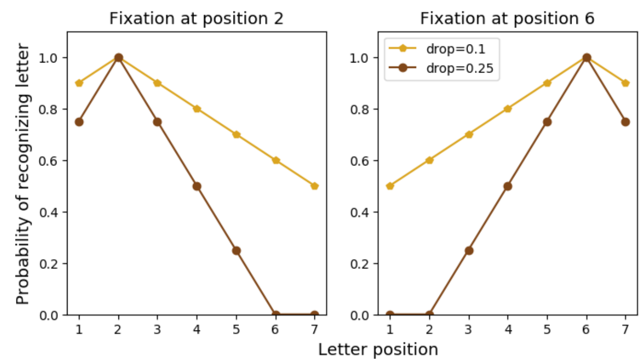

Figure 2: Probability of recognizing the constituent letters in a word when fixating letter position 2 (left) and letter position 6 (right) in a word according to the McConkie model.

Procedure We tested for how the fixation location could impact the information content extracted from the perceived word. For simplicity, and to test for strong modulations in word recognition due to the perceived information, here we focused on the information content extracted from a fixation near the beginning of a word (at the second letter position) and near the end of a word (at the sixth letter position). After passing each word through the McConkie filter, for each word, we calculated the remaining amount of uncertainty on the identity of the word *after* fixating at each of these fixation locations (a proxy of the information content in each location[4]). The measure of uncertainty we used was *entropy*, as proposed in Shannon (1948). Concretely, given the letters retrieved after fixating on a word, we computed the remaining entropy as $H = -\sum_{w=1}^{m} p_w log_2(p_w)$, where the words w belonged to the set of words m that have a perfect match with the identified letters, both in letter identity and letter position (e.g. the word 'therapy' would be in the set of matching words for the recognized letters 'ther - - -'). The probability of a matching word p_w was estimated as its relative frequency

[1]The code for our models and analyses is released at https://github.com/rgalhama/nnfixrec_cogsci2019.

[2]From https://github.com/hermitdave.

[3]We ruled out the possibility that vocabulary size drove any of our simulated behavioral effects by down-sampling the English corpus to be the same size as the Hebrew corpus in our simulations.

[4]Note that, throughout our paper, we use the term "information content" of a fixation location to quantify the contribution of the observed letters in minimizing the uncertainty *on the identity of the word*. This should not be confused with the *surprisal* conveyed by the letters in a fixation location. The former concerns a probabilistic models for words (based on word frequency and component letters), while the latter would be based on a probabilistic model of letter strings.

in the corpora.[5] To test whether information was distributed evenly across the beginning and end of all words in each language, we subtracted the entropy for each word at fixation location 2 from that at fixation location 6. If entropy were evenly distributed, these values should cluster around 0.

Additionally, from the distribution of entropy difference scores, we identified 100 words with the most extreme positive (50 words) and negative (50 words) values.[6] We refer to the words with more information content when fixating at position 2 (i.e. negative entropy differences) as the *maxIC*(2) words, and the words with more information content at position 6 (i.e,. positive entropy differences) as the *maxIC*(6) words. These words served as test items in Study 2.

Results

The difference in entropy values when a word was perceived at fixation location 2 versus fixation location 6 are plotted in Figure 3. It clearly shows that entropy information is not uniformly distributed across words, as in both English and Hebrew there are more negative scores. It also shows a relatively wide range of entropy values, with both languages containing words with entropy difference scores ranging from approximately -5 to 3. Further, the Hebrew scores tend to be more negative than the English scores.

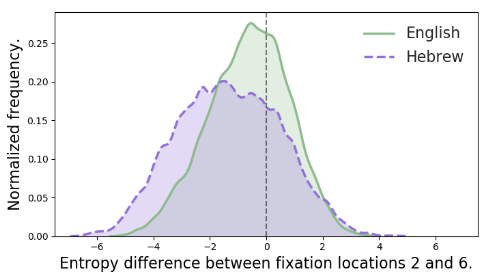

Figure 3: Distribution of entropy differences for 7-letter words.

To give a more concrete intuition into the magnitude of the difference scores and their relationship to successful visual word recognition, consider the case of the word "zooming", which has an entropy difference score of -5.1. If the fixated letter and the two letters on either side of this letter are preceived correctly, there is a 100% likelihood of successful recognition of this word when fixated at position 2 (i.e. when perceiving 'zoo-'). However, there is only a 3% success rate when fixated at position 6 (i.e. when perceiving '-ing').

Next, we selected 100 words per language with "extreme" entropy difference scores for use in Study 2. These items had mean difference scores, for *maxIC*(2), of -2.68 in English and -3.32 in Hebrew, and for *maxIC*(6), it was 2.19 in English and 2.58 in Hebrew.

In additional simulations, not reported in detail due to space constraints, we also confirmed that varying the exact shape of the McConkie function and the value of the *drop* parameter did not qualitatively alter these trends unless only the nearest items to the fixation location, or nearly all the words in the word, were perceived with 100% accuracy.

Discussion

The first simulation substantiated our predictions that different amounts of information content can be extracted by fixating at different locations in a word. Overall, there appears to be more information content present at the start of words in both languages, providing initial evidence for a language-general trend. Thus, the fixation distributions in different languages may at least be partially attributed to a system that attempts to minimize entropy in the visual signal in service of word recognition. This claim is further bolstered by the fact that the Hebrew distribution was even more shifted to contain more information when fixating at the beginning of a word, consistent with the stronger preference to fixate earlier in Hebrew words in behavioral data (see Figure 1). The broad distribution of values in each language also enabled us to select items with "extreme" entropy difference scores across fixation locations. This enabled us to test whether some words are more readily identified by fixating at a location other than the overall preferred fixation of the language (which is off-center, nearer to the beginning of the word).

Having thus established that the perceptual input to the word recognition system contains major differences in entropy based on fixation location, we next explored how these inputs could shape processing in a connectionist model of word recognition.

Study 2: A perceptually-constrained connectionist model of visual word recognition

The previous study focused on the distribution of information contained in the languages. In this study, we employed a connectionist model and a coordinated pilot behavioral experiment to investigate whether a learning model of word recognition is sensitive to these information patterns. This allowed us to develop new predictions about how performance in different fixation locations evolves in relation to reading proficiency: although it is beyond our goals to align model training (in epochs) with human reading experience —which is a non-trivial question–, our learning model provided us with insights into novel emergent processing dynamics that are not visible from an information-theoretic approach. In particular, we focused on whether the model and the human participants displayed an interaction between fixation location and the location of maximum information content in our "extreme" items selected in Study 1. Because some of the implementational decisions for the model were made to increase

[5] An alternative approach is to compute these values over word types rather than word tokens. Control simulations showed that both of these approaches were highly correlated in both languages, $r >$ .72, and that the correlations between entropy differences over types or tokens and word frequency were extremely small, $|r| < .04$

[6] We filtered some items to avoid the over-representation of letter combinations like "-ing" and to eliminate extremely low and high frequency items.

the similarity between the simulated task and the pilot behavioral task, we provide a brief overview of the behavioral task and findings before turning to the details of the simulation (for the complete report of this experiment, see Siegelman et al., 2019).

Overview of the behavioral task. A total of 23 native speakers of Hebrew (14 females, age range: 22-30, mean: 24.9) were presented with the a set of words including the 100 words with extreme information differences described previously, and an additional selection of 100 words (also of length 7)[7] with intermediate entropy differences, which we treat as fillers for the purpose of this paper.[8] In the task, participants first focused on a fixation cross for 1000 ms, and then were presented with a word for 100 ms. This brief presentation prevented multiple fixations and was expected to lower performance below ceiling. Critically, the word was presented in different locations, including cases wherein the letter at position 2 or position 6 appeared at the same location as the fixation cross (i.e., the word was left- or right-shifted from screen center where participants were fixating). Participants were then instructed to say or guess the word that had been presented. Responses were coded either as correct or incorrect. All words were presented in both fixation locations simulated in the model (i.e., fixation at position 2 vs. position 6). Words were presented in a random order.

The results are presented in Figure 4. The pattern of results indicated that accuracy was highest when a word was fixated at the location which contained the most information content, and lower at the location which contained less information content. Critically, this was true not only for words that had more information content early in the word, but also for words that contained more information content near the end of the word: a logistic mixed-effect model (with condition, fixation location, and their interaction as fixed effects, trial number and log-transformed frequency as control variables, by-subject and by-item random intercepts, and by-subject random slope for condition) revealed a significant interaction (B = 0.59, SE = 0.05, $p < 0.001$ and a significant main effect of fixation location (B = -0.23, SE = 0.05, $p < 0.001$). Thus, these findings do not simply reflect a preference to process words in the more frequently fixated location in a given language, which our prior study showed contains, on average, more information content. The presence of numerically larger differences across fixation locations for words

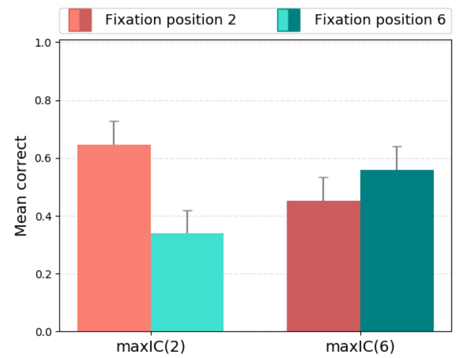

Figure 4: Mean correct responses of participants in the four conditions. Error bars = SEM.

with more information content near the start of a word relative to near the end of a word may suggest a more subtle interaction between information content and frequency of exposure to different locations, however. Additionally, averaging across the four experimental conditions, overall accuracy was significantly below ceiling.

Simulating the behavioral experiment

Model Architecture We implemented a feed-forward connectionist model that mapped perceptually-constrained distributed input of a word's constituent letters onto a localist representation of each word in the training vocabulary, as illustrated in Figure 5. There were 7 letter input slots, one for each position in a 7-letter word. Each of these slots had one unit for every letter in the alphabet and coded for the presence (1) or absence (0) of a given letter in that position (i.e., a binary one-hot coding). The distributed representation of the visual word was then input to a McConkie filter set to perceive the word at a particular fixation location (the procedure for specifying fixation locations is described later). Thus, the one-hot vectors would be down-scaled (using a drop parameter, d, of 0.25) to reduce the activity of having perceived a given letter as a function of eccentricity from the fixation location, as quantified in Equation 1:

$$x(i) = x(i) * max(0, 1 - eccentricity(i) * d) \tag{1}$$

Thus, the activity of the fixated letter remained unchanged, the activity of letters more than four slots distant from the fixated letter was set to 0, and activity in each letter-slot would decrease linearly between these two bounds.

To simulate the noisy nature of perceptual inputs in the human visual system, we next injected normally-distributed random noise (μ = 0.2, σ^2 = 0.05) into the unit activations (clipping activations to $[0, 1]$). We assumed that the activity after these processing steps was analogous to what would be available in an early visual-orthographic representation ("perceived input" in the Figure).

Next, we mapped the perceived input onto a one-hot log-softmax target output representation for each word in the vocabulary through a pool of 125 hidden units. The output of

[7] All the words we use are multisyllabic. This could create confounds in the modelling work if we mapped the input with phonological representations, but we intentionally focused only on orthographic factors, since our goal is to find out what structure exists in the orthographic signal alone in the absence of phonological considerations. Future work may investigate how these visual representations interact with phonological representations, similar to a classic "triangle" model (Seidenberg & McClelland, 1989).

[8] Although we did not test these items in the model, the behavioral results indicated that items with intermediate entropy difference scores were relatively unaffected by whether they were fixated near onset or offset, as predicted by the account.

the hidden units was determined first by computing the sigmoidal function of their net input, followed by the injection of uniform random (output) noise (mean = 0, range = 0.05). All the weights in the network were randomly initialized from a uniform distribution in the range $[\frac{-1}{\sqrt{fan\_out}}, \frac{1}{\sqrt{fan\_out}}]$, where fan_out was the number of units in the subsequent layer of the model.

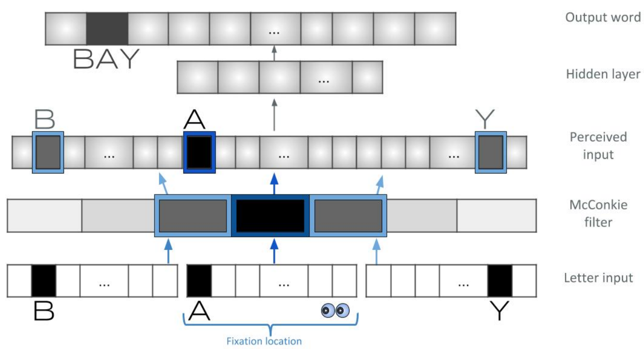

Figure 5: Model architecture. The implemented model perceived 7-letter words; here, we illustrate the model processing a 3-letter word (BAY) for simplicity. Full (1) and no (0) activity are shaded in black and white; intermediate values are shaded in grey.

Determining fixated locations in a word We evaluated several methods for selecting how often a word was perceived from a different fixation location, which we term "fixation location distribution schemes". One model sampled each fixation location equally (hereafter, the *uniform* fixation model). This provided an estimate of the impact of the entropy at different fixation locations that was unconfounded with how often humans typically fixate at each location, and how those distributions varied across the two languages under study. Another model employed the language-specific behaviorally-derived fixation distributions illustrated in Figure 1 (hereafter, the *behavioral* fixation model). A third model averaged these two fixation distribution schemes (hereafter, the *50/50* model). This "blended" model allowed us to interpolate between these two previously described schemes and simulated a case where a model was sensitive to frequency of exposure, but not necessarily to the raw values. The logic here was that a good model might standardize frequency information to ensure low-frequency information is also learned.

Training The model was trained by presenting a 7-letter word at a particular fixation location and computing the cross-entropy error between the output and the target representation. Error was accumulated in batches in which every seven-letter word in the target language was presented 20 times, with the likelihood of fixating at a particular location determined by the fixation distribution sampling scheme. Error was then backpropagated to adjust the weights between the perceived input and the output layer (learning rate = .005; weight decay = .0001) using stochastic gradient descent for the first 10 epochs, and the Adam algorithm thereafter (Kingma & Ba, 2014). The model was trained for 200 epochs (runs through each batch).

All models reached a stable high level of overall word recognition accuracy (near 80%) for approximately the last 50 epochs of training. The vast majority of the incorrect responses originate from words perceived at a suboptimal—and where applicable, less frequent—fixation location. Figure 6 provides representative data for the Hebrew words with extreme entropy values using the *uniform* model. (space constraints prevented the inclusion of plots from the other models, which were broadly similar). The presence of different effects during early training than at the end of training also makes novel predictions for future developmental studies.

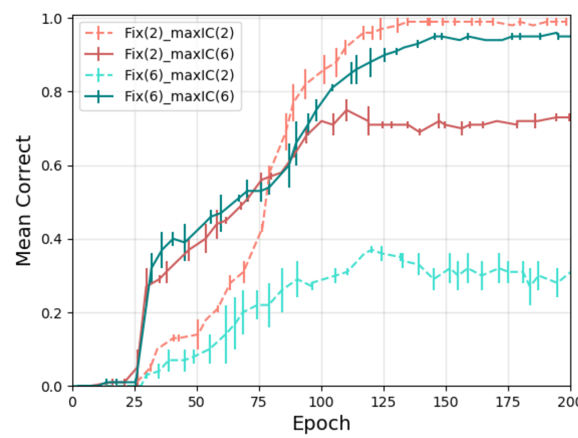

Figure 6: Accuracy for the *uniform* model trained on Hebrew for the words with extreme entropy difference scores. Error bars = SEM.

Testing We froze the weights on the trained models before testing them in a manner analogous to the behavioral experiment. In the test, we presented all the *maxIC*(2) and *maxIC*(6) words at both fixation location 2 and fixation location 6. We also tested several methods of bringing performance in the task below ceiling as in the behavioral experiment, including dimming model inputs (multiplying all input letter activations by a value less than 1), and increasing variance of the noise applied to the perceptual input (cf. Lambon Ralph, Lowe, & Rogers, 2007). These methods yielded similar overall results, so here we report only the results of dimming (dimming parameter = .35). We ran this simulation twice on models initialized with different random weights and report the average results.

Results

The results for the *uniform*, 50/50, and *behavioral* fixation models of English and Hebrew are presented in Figure 7. First, in contrast to the non-dimmed model at the end of training (see Figure 6), our testing procedure clearly succeeded in

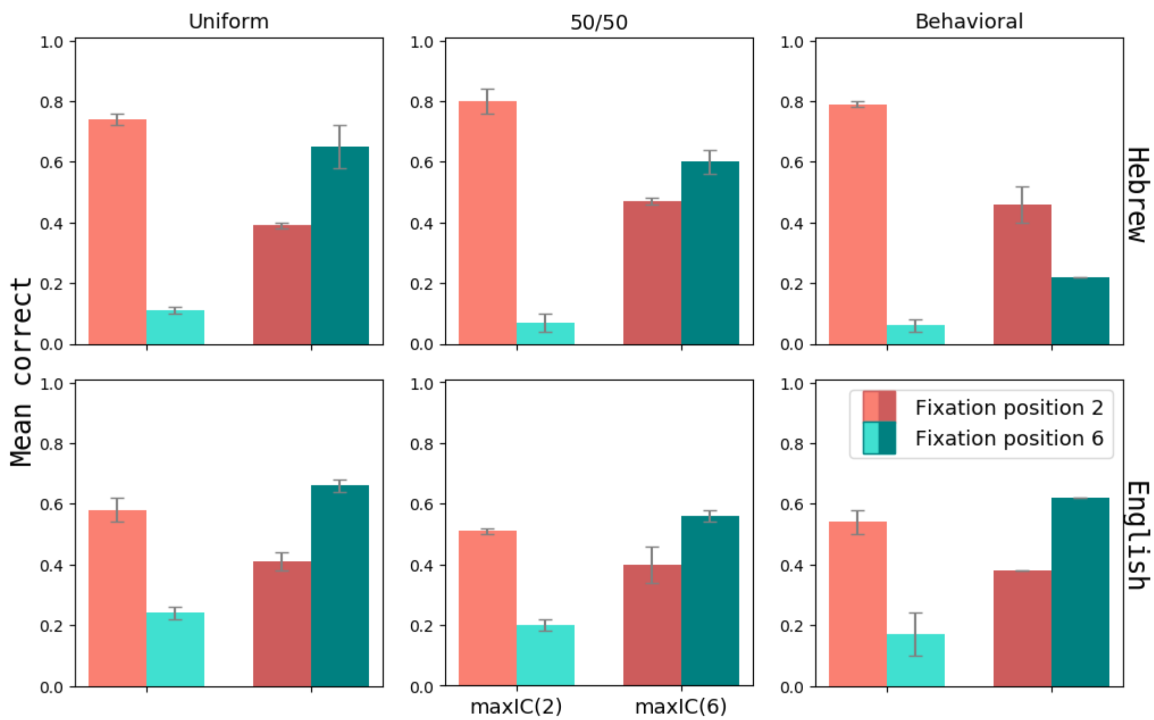

Figure 7: Results of testing the different fixation location distribution models in Hebrew [top] and English [bottom]. Error bars = SEM.

lowering overall accuracy to a level similar to that in the behavioral experiment.

The most critical finding was that, with the exception of the Hebrew *behavioral* fixation model, every one of these models produced a qualitatively similar interaction between fixation location and whether more information content was located at the beginning or end of a word, as in the behavioral results. The Hebrew simulations also show more pronounced effects for the $maxIC(2)$ items overall, as in the behavioral data. The reduced effects in English make for novel predictions for an experiment in that language.

Moving from the $uniform$ model through the $50/50$ model to the *behavioral* model, the effects of the behavioral fixation sampling scheme, which fixates words at position 2 more of then than position 6, is more apparent in Hebrew than in English. This is reflected by the fact that $max(IC2)$ words are perceived more accurately when fixating at position 2 in Hebrew and less accurately when fixating at position 6 when moving toward the behavioral fixation scheme.

The exceptional Hebrew *behavioral* fixation model appears to be an exaggerated extension of the effects of fixation location frequency outlined above. In the case of this model, even the $maxIC(6)$ words were responded to more accurately when fixating earlier in the word. The presence of this pattern only in Hebrew is at least partially explained by the more extreme differences in fixation location sampling distributions in Hebrew than in English. These results also suggest that the human visual recognition system may at least partially normalize the effects of fixation location frequency, given that the $50/50$ and $uniform$ fixation models produced qualitative results more similar to those in the behavioral experiment.

Discussion

The results of the second set of simulations largely paralleled those of the behavioral experiment, with both exhibiting an interaction between fixation location and the location with most information content. The simulations also showed the influence of the behavioral fixation location distributions in enhancing the perception of words at the most frequent fixation location, and suggest that the word recognition system normalizes the fixation location distribution to some degree. Further, although the qualitative findings were similar across languages, suggesting that a general principle is at play, at a quantitative level there were some differences between the two target languages. These differences align with the relatively higher information content at the beginning of Hebrew words and the greater likelihood of fixating at the beginning of words. Collectively, this work therefore indicates that the word recognition system is sensitive to the information content in different locations in a word, as constrained by the perceptual system.

These results are only in partial agreement with past work (Brysbaert & Nazir, 2005). In that work, participants were presented with partial word information for 5-letter words and asked to "guess" the word. The distribution of "guesses" relative to the correct response was then taken as their measure of uncertainty. Their results showed similar effects as in our study at word onset, but no effects at word offset. These findings were interpreted as suggesting that the effects of information content were only present at the preferred fixation location. A combination of factors likely explain the discrepancy between their claims and ours, including a more adequate formal quantification of information content and the use of longer words that may be more sensitive to perceptually-constrained information content effects.

The success of this work at the individual word level also points to important directions for future work. One major question raised by this work is how these principles could generalize to the multi-word level. Can a preceding word provide top-down context and reduce the uncertainty (i.e. the average information content) on the set of upcoming words so as to not only facilitate processing, but also alter the location of an upcoming fixation? If so, this result could help explain the relatively broad fixation distributions obtained in different languages, because the optimal location to fixate in a word may deviate from the average location from the language as a function of context. The somewhat broad overall fixation location distributions may therefore in actuality reflect the averages of narrower fixation location distributions that are conditioned by the preceding word.

Our work shows that the observed behavioral effects in word recognition can be explained based on low-level information structures in the visual signal, without the need to resort to higher-level morphological structures. Higher-level structures can enter the visual-orthographic system in two ways: first, in shaping the word forms of a language, and second, as representations that mediate word recognition. The former is subsumed in our information-theoretic approach, which encompasses all the constraints that provided word forms with their actual shape, providing us with a quantitative comparative framework for a crosslinguistic perspective. The latter cannot be completely ruled out: although our model does not require morphological representations to succeed, the contribution of these representations should be assessed with more targeted experiments that aim to tease apart the visual/orthographic from morphological (e.g. looking at performance for regular and irregular morphemes such as *brothel/broth, corner/corn, farmer/farm*, Rastle, Davis, & New, 2004).

To sum up, this work offers a language-general and parsimonious account of how a specific type of statistical information drives performance in the perceptually-constrained word recognition system, complementing accounts based on the operation of the occulomotor system, as well as complementing or or subsuming accounts based on higher-level information. In so doing, this work reinforces the importance of studying how the structure of language itself interacts with the perceptual constraints of the visual/orthographic system (Lerner, Armstrong, & Frost, 2014) in shaping reading behaviors, and opens new avenues for combining isolated word and naturalistic reading research.

Acknowledgments

This project was funded by the ERC (ERC-2015-AdG-692502 to RF) and NSERC (DG-502584 to BCA).

References

Brysbaert, M., & Nazir, T. (2005). Visual constraints in written word recognition: evidence from the optimal viewing-position effect. *Journal of Research in Reading*, *28*(3), 216–228.

Deutsch, A., & Rayner, K. (1999). Initial fixation location effects in reading hebrew words. *Language and Cognitive Processes*, *14*(4), 393–421.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological review*, *112*(4), 777.

Kingma, D. P., & Ba, J. L. (2014). Adam: A method for stochastic optimization. In *Proc. 3rd Int. Conf. Learning Representations.*

Lambon Ralph, M. A., Lowe, C., & Rogers, T. T. (2007, 04). Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. *Brain*, *130*(4), 1127-1137.

Lerner, I., Armstrong, B. C., & Frost, R. (2014). What can we learn from learning models about sensitivity to letter-order in visual word recognition? *Journal of Memory and Language*, *77*, 40–58.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. an account of basic findings. *Psychological Review*, *88*(5), 375.

McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: Ii. frequency of refixating a word. *Perception & Psychophysics*, *46*(3), 245–253.

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

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the EZ reader model. *Vision research*, *39*(26), 4403–4411.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review*, *96*(4), 523.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Siegelman, N., Alhama, R. G., Bogaerts, L., Armstrong, B., Kuperman, V., & Frost, R. (2019). *Reading across writing systems: an information-theoretic perspective, in prep.*

Tiedemann, J. (2012). Parallel data, tools and interfaces in opus. In N. Calzolari et al. (Eds.), *Proc. 8th International Conf. on Language Resources and Evaluation.* Istanbul, Turkey: European Language Resources Association (ELRA).