To Catch a Snitch: Brain potentials reveal knowledge-based variability in the functional organization of (fictional) world knowledge during reading

Abstract

People vary in what they know, yet models of language processing do not take this variability into account. We harnessed the temporal sensitivity of event-related brain potentials alongside individual differences in Harry Potter (HP) knowledge to investigate the extent to which the availability and timing of information relevant for real-time word comprehension are influenced by variation in degree of domain knowledge. We manipulated meaningful (category, event) relationships between sentence contexts about HP stories and critical words (endings), assessed via behavioral ratings and by measuring similarity of word embeddings derived from a high-dimensional semantic model trained on HP texts. Individuals' ratings were sensitive to these relationships according to the degree of their domain knowledge. During reading, N400 amplitudes (neural measures of semantic retrieval) also reflected this variability, suggesting the degree to which information relevant for word understanding is available during real-time sentence processing varies as a function of individuals' domain knowledge.

Keywords: language processing, ERPs, knowledge, individual differences

Introduction

Across cognitive systems, world knowledge allows individuals to organize raw sensation into meaningful experiences. Understanding language is no exception—words cue world knowledge which can be rapidly brought to mind in real time (e.g., Hagoort et al., 2004), incrementally and sometimes even predictively (reviewed in Altmann & Mirković, 2009; Kutas, DeLong, & Smith, 2011). A more precise description of how this occurs—including which types of knowledge, their organization, and the timing of their use—requires a closer look at knowledge availability in real time. It is, however, experimentally challenging to capture the specifics of an individual's world knowledge with standard laboratory procedures.

Troyer and Kutas (2018; Troyer, Urbach, & Kutas, under review) provided a potential solution by focusing on a restricted domain of knowledge with the requisite properties for online language processing studies, including a large, rich set of verbal descriptions, wherein college-aged young adults differed in their degree of knowledge—the fictional world of Harry Potter (HP) by J.K. Rowling. Troyer & Kutas (2018) recorded EEG while participants with varying degrees of knowledge about HP read sentences that described general topics, followed by sentences that described events from the HP stories; sentences ended either in contextually supported or unsupported words. Across participants, and for both sentence types, the effect of contextual support was present on N400 amplitudes—a brain potential sensitive to factors impacting the ease of retrieval from semantic memory, with larger reductions in N400 (i.e., more positive-going potentials) associated with greater ease of retrieval (reviewed in Kutas & Federmeier, 2000). But critically, participants' degree of HP knowledge influenced the size of this effect only for the sentences about HP. More specifically, individuals' HP knowledge was correlated with N400 amplitudes to contextually supported, but not to unsupported, words. These results empirically demonstrate that the rapid influence of written sentence context, known to modulate N400 brain potentials, is a function of each individual's knowledge.

These findings are not surprising given the vast literature showing that people rapidly make use of a variety of word and world knowledge as they understand words in real time, such as orthographic neighborhood density (Laszlo & Federmeier, 2009), word frequency (Van Petten & Kutas, 1990), and non-linguistic knowledge including the organization of categories in semantic memory (Federmeier & Kutas, 1999), facts about the world (Hagoort et al., 2004), generalized event knowledge (Metusalem et al., 2012), personal preferences (Coronel & Federmeier, 2016), and fictional characters (Filik & Leuthold, 2013). It stands to reason that the structure and organization of individuals' knowledge would have consequences for the availability, contents, and timecourse of bringing to mind these varied sources of knowledge in real time.

One way to ask whether and, if so, when people bring different types of information to mind as they read sentences is to probe them with words that are linguistically anomalous, yet systematically related to the sentence context and/or a likely upcoming, linguistically licensed word. This related anomaly paradigm has been fruitfully employed to investigate the influence of the functional organization of semantic memory on sentence processing. For example, in sentence contexts setting up an expectation for the word *pines*, categorically related words (e.g., another type of tree, *palms*) elicited reduced N400 amplitudes compared to words from a different category (e.g., *tulips*), but which were larger than those to the expected word (Federmeier & Kutas, 1999, 2002). In a different study, where individuals read short paragraphs about common events (e.g., playing football) that set up linguistic expectations for a word (e.g., *touchdown*), unexpected and linguistically unlicensed words related to the event being described (e.g., *helmet*) also elicited reduced N400 amplitudes compared to unrelated words (e.g., *license*) (Metusalem et al., 2012). It is worth noting that the "related anomaly" in Federmeier & Kutas shared many features with an expected word whereas in Metusalem et al. the related anomaly was related in one or more of several ways to the generalized event being described in the context, but not did not share features with the linguistically expected word. Nonetheless, the related anomaly ERP effects in both studies had a similar timecourse and scalp topography, maximal around 400 ms over centro-parietal recording sites,

suggesting that people made quick use of both types of related information during real-time sentence processing.

The availability of related/relevant information stored in semantic memory during real-time language processing must, at least to some degree, be modulated by each individual's degree of domain knowledge. Indeed, the literature on expert knowledge proposes that the functional organization of information around themes, events, and categories is likely to depend on individuals' degree of expertise (reviewed in Ericsson et al., 2006). To investigate the extent to which variation in domain knowledge influences the nature and timing of the availability of knowledge stored in long-term memory during real-time sentence processing, we probed semantic memory using a related anomaly paradigm incorporating sentences describing the narrative world of Harry Potter.

Using freely available materials (including Wikipedia and HP fan sites) along with the text of the HP book series by J.K. Rowling, the first author created a set of 156 sentence pairs that accurately described events and entities from the series. Each sentence context ended either in (a) a contextually Supported (and linguistically expected) word; (b) a word which was factually incorrect and Unrelated to the context and to the supported word; and (c) a word which was factually incorrect but which was Related in one of two ways to the context and/or contextually supported word. For half of the materials, the related words were taken from the same category as the linguistically expected word, as in Federmeier & Kutas (1999). For the remaining materials, the related word was related in some way to the episode/event being described by the preceding sentence context, as in Metusalem et al. (2012). Based on the previous findings, we expected that both types of relationships would lead to N400 related anomaly effects which might be similarly influenced by the degree of individuals' domain knowledge. Three lists were constructed such that each sentence frame and each critical word appeared only once per list (examples provided in Table 1).

In order to verify that the words we deemed related via category or event to contextually supported words were indeed more closely related than the unrelated ending, we conducted a series of experiments to examine these relationships. First, we trained a high-dimensional semantics/language model directly on the text of the HP book series; we then asked whether the word embeddings learned by the model reflected the manipulation in our materials (e.g., with Supported-Related word embeddings being closer in semantic space than Supported-Unrelated word embeddings). Next, we conducted two experiments asking participants of varying degrees of HP knowledge to rate critical words from our materials for their similarity and relatedness, respectively. Finally, with these measures in hand, we conducted an EEG/ERP study to ask to what extent and when domain knowledge impacts the availability of contextually supported as well as contextually unsupported yet functionally (categorically, event-based) related knowledge during written sentence comprehension.

Experiment 1: Word embeddings

We trained a word2vec model (Mikolov et al., 2013a,b) on the text from the HP book series. This model uses a neural net to learn word embeddings (vectors) in high-dimensional semantic space from word co-occurrences in the input. The semantic "contents" of such embeddings can reflect various aspects of meaning, including category and event-based relationships (reviewed in Lenci, 2018). We could then use these embeddings to quantify relative similarities/differences between word pairs (or average vectors computed over sequences of words).

Methods

Word2vec model. We trained a word2vec model (distribution by D. Yaginuma, https://github.com/dav/word2vec) on the text from the seven books of the HP series, taken from the official electronic publication (https://usd.shop.pottermore.com)—a total of 1,125,854 words, with a vocabulary size of 8,046 words (subject to the constraint of each word appearing at least 5 times in the HP books). We used the continuous bag-of-words (CBOW) architecture, which learns to predict a word based on its context—in our case, a window of 10 words on either side. Each word from the HP books was modeled as a point (i.e., vector) in a 200-dimensional space.

Word-word similarity Using this model, we extracted word embeddings for critical words from each of our experimental conditions (Supported, Related, and Unrelated). For each

Table 1. Sample HP sentence materials.

Sentence frame	Supported	Related	Unrelated	Related Anomaly Type
Sybill Trelawney is a Hogwarts Professor. She teaches	*Divination*	*Transfiguration*	*basilisk*	**Category**
In Quidditch, games are usually won in one way. This is when the seeker catches the	*Snitch*	*Bludger*	*dragon*	**Category**
Harry has a patronus. It takes the form of a	*stag*	*dementor*	*Sectumsempra*	**Event**
When Harry is one year old, Hagrid brings him to the Dursleys'. For transportation, he uses a borrowed	*motorcycle*	*Sirius*	*Vow*	**Event**

item (156 total), we then computed the cosine similarity (angular distance) between the word embeddings for Supported-Related and Supported-Unrelated pairs of critical words. We expected that the similarity for the Supported word-Related word pair would be greater Supported word-Unrelated word pair.

Sentence context-word similarity We also extracted word embeddings for each word (where possible) of our sentence pair frames/contexts. To create a single embedding (i.e., vector) for each item's sentential context, we took the average of all its words' vectors. We then computed the cosine distance between this aggregate context vector and the vector for each ending type (Supported, Related, Unrelated). We expected this distance would be greatest for the Context-Supported pair, followed by Context-Related and finally Context-Unrelated.

Results

As expected, we found that word embeddings—derived from a corpus of the HP novels' text—for Supported words were more similar to Related words than to Unrelated words (Fig. 1a). This pattern held for both category- and event-related item subsets, though it was somewhat larger within category-related items ($p < .01$). Also as expected, average word embeddings for sentential contexts were most similar to Supported words, followed by Related and Unrelated words (Fig. 1b). This pattern held both for category- and event-related item subsets. These findings show that the high-dimensional semantic space learned by the word2vec model captured systematic, meaningful differences in the relationships between the sentence context and the Supported, Related, and Unrelated endings.

Experiments 2a-b: Ratings studies

To further assess the manipulation in the HP sentences, and to examine the extent to which the manipulation was dependent on HP knowledge, we conducted two behavioral studies, asking participants to rate critical word-pairs (Supported-Related and Supported-Unrelated) on similarity (Exp. 2a) or relatedness (Exp. 2b). These criteria were chosen specifically to examine the two types of relationships we targeted in our HP sentence materials, namely categorical relationships (words share many similar features) and event relationships (words are related via an event/episode from the HP books). In addition, these experiments allowed us to assess the ratings of similarity/relatedness as a function of individuals' degree of HP knowledge. We expected that individuals with greater knowledge would be more sensitive to our experimental manipulations—i.e., that more knowledgeable individuals would indicate relatively greater similarity/relatedness for Supported-Related than for Supported-Unrelated word pairs.

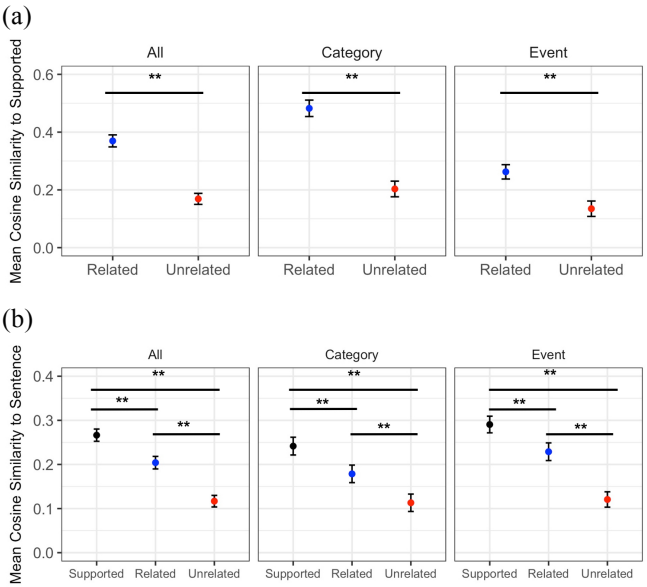


Fig. 1. (a) Across items and within each subset (Category-related, Event-related), mean cosine similarity for Supported & Related endings is greater than for Supported & Unrelated (all ps < .01). (b) For all 156 items (and within each subset) there was a significant three-way difference between cosine similarity of averaged word embeddings for sentences and Supported < Related < Unrelated endings.

Methods

Participants 24 participants completed similarity ratings; 25 different participants completed relatedness ratings. All participants were UCSD students; they received partial course credit as compensation.

Procedure For the similarity ratings experiment, participants were asked to consider word-pairs in the context of the Harry Potter stories and to judge their similarity in meaning using a scale ranging from 1 ("not similar at all") to 7 ("nearly the same meaning"). They were given the following guide to judging similarity of word meanings:

(1) Do the two word meanings behave similarly (e.g., do they perform the same actions)?
(2) Do the two word meanings share physical / sensory properties (e.g., do they look, taste, smell, sound or feel similarly)?
(3) Do the two word meanings share many functional properties (e.g., are they used in similar ways, or do they serve a similar purpose)?
(4) Do the two word meanings share any other properties and/or features in common?

For the relatedness ratings experiment, instructions were similar, except participants were asked to judge words on how *related* they were using a scale ranging from 1 ("not related at all") to 7 ("very closely related)". They were given

the following guide to judge whether the pairs were meaningfully related:

(1) How likely are the words to show up within the same context (that is, in/around the same part of the HP stories)?
(2) How important does one word seem to be for understanding the meaning of the other?
(3) Are the two words related via some theme, topic, event, or episode/scenario in the HP stories?
(4) Are the two words related via any other relationship?

Participants in Exp. 2a-b also completed a 10-question trivia quiz assessing their HP knowledge and a questionnaire about their HP experience.

Results

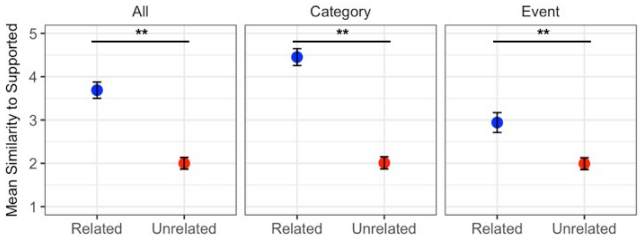
As expected, mean similarity ratings for Supported-Related word pairs were greater than those for Supported-Unrelated word pairs (Fig. 2a). This pattern was similar for both the category- and event-related item subsets, but was larger for the category-related subset, which might be expected based on greater similarity due to feature overlap between members of the same category (compared to words related via an event or episode).. Also as expected, HP knowledge was positively correlated with the size of the effect (i.e., similarity for Supported-Related word pairs minus similarity for Supported-Unrelated word pairs) at $r = .51$, $p < .05$.

In addition, mean relatedness ratings for Supported-Related word pairs were greater than those for Supported-Unrelated word pairs (Fig. 2b). This pattern was similar for both the category- and event-related item subsets. Also as expected, HP knowledge was positively correlated with the size of the effect (i.e., relatedness for Supported-Related word pairs minus relatedness for Supported-Unrelated word pairs) at $r = .68$, $p < .001$.

We also examined the correlation between the word2vec cosine similarity measures (Exp. 1) and the similarity and relatedness ratings for Supported-Related and Supported-Unrelated word pairs, respectively (Exp. 2). Cosine similarity was positively correlated with both similarity ($r = .43$, $p < .0001$) and relatedness ($r = .26$, $p < .01$) ratings for the Supported-Related word pairs, but not for the Supported-Unrelated word pairs (n.s.).

These results empirically indicate that our Supported sentence endings were indeed more similar/related to our Related, compared to Unrelated, endings. Moreover, that the size of these effects was positively correlated with HP knowledge further supports the notion that sensitivity to the relatedness manipulation depends on knowledge specific to the HP book series. Next we describe an ERP/EEG study designed to investigate the extent to which individual differences in domain knowledge influence the availability of information relevant for word processing—i.e., information cued by categorically / event related words—during real-time sentence processing.

(a)



(b)

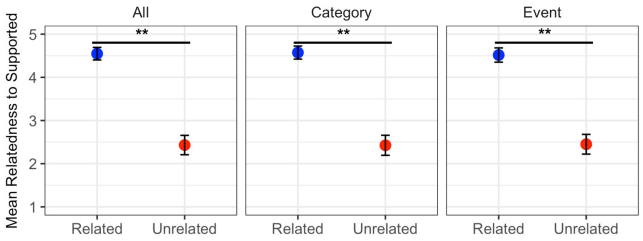


Fig. 2. (a) Across items and within each subset (Category-related, Event-related), similarity ratings for Supported & Related endings are greater than Supported & Unrelated (all ps < .01); this effect was larger for the category-related subset of items compared to the event-related subset ($p < .0001$). (b) Across items and within each subset, relatedness ratings for Supported & Related endings are greater than Supported & Unrelated (all ps < .001).

Experiments 3a-b: ERP studies

In Experiment 3, we asked whether certain aspects of the functional organization of semantic memory—namely organization of words/concepts via categories (wherein members share many similarities/features) and event/episode relationships—would be available to comprehenders as they read sentences about a fictional domain. We were particularly interested in whether the availability of not just the contextually supported information, but also the contextually unsupported, related information would be a function of the degree of individuals' domain knowledge.

To this end, participants of varying degrees of HP knowledge read 156 sentence pairs about the fictional world of HP while we recorded EEG (Experiment 3b). Sentences ended in a critical word that was either Supported, Related, or Unrelated (HP sentence materials described above). Three lists were then constructed so that every participant read each sentence frame and each critical word only once. That is, even though the same critical word appeared in other conditions on other lists, it never appeared in the critical position more than once in the same list. All but three words appeared as critical words in two or all three conditions. We expected that for individuals knowledgeable about HP, we would see a three-way difference in the amplitude of N400 potentials to the critical words, with the largest amplitude for Unrelated, the most reduced amplitude for Supported, and an intermediate amplitude for Related. We also expected that

N400 amplitude to Supported words would be positively correlated with HP knowledge (replicating Troyer & Kutas, 2018 and Troyer, Urbach, & Kutas, 2018). Moreover, we expected that N400 amplitude to Related words would be positively correlated with HP knowledge, consistent with the real-time use of differential functional organization of long-term memory as a function of degree of domain knowledge. For visualization of ERPs, we present subgroups of participants based on a median split on HP knowledge.

To demonstrate (a) that individuals spanning a range of HP knowledge scores could elicit N400 effects, more generally, and (b) that the relationship between HP knowledge and N400 effects was specific to sentences about HP, we also recorded EEG while the same participants read sentences about general topics (Experiment 3a). Due to time constraints, we included only 40 such sentences, half of which ended in a contextually Supported word / the best completion (determined by an offline cloze norming task in which participants provided completions to sentence frames) and the other half of which ended in a contextually Unsupported word (a plausible word that was low-cloze).

Methods

Participants 48 students from the UCSD community participated in the EEG study (Experiments 3a-b).

Experimental procedures Participants were instructed to silently read sentences for comprehension, first about general topics (Experiment 3a) and then about Harry Potter (Experiment 3b). In each experiment, the whole first sentence appeared in the center of the screen. When ready, participants pressed a button to move on to the second sentence, which was presented one word at a time in the center of the screen with a 500 ms SOA (200 ms on, 300 ms off). Following the ERP study, participants completed a 10-question HP trivia quiz and a questionnaire about their HP experience.

In addition, we collected several other measures of individual differences to better understand group differences among participants (see Troyer & Kutas, 2018, for more details). We combined measures of general print/reading experience (media and reading habits questionnaire, author and magazine recognition tests; Stanovich & West, 1989) for an aggregate reading experience score, and we also collected a measure of general knowledge (trivia quiz developed from freely available materials), and verbal working memory (sentence span, Daneman & Carpenter 1980). Finally, we administered a debriefing questionnaire.

ERP recording and data analysis The electro-encephalogram (EEG) was recorded from 26 tin electrodes geodesically arranged in an ElectroCap, with impedances kept below 5 KΩ. Recordings were referenced online to the left mastoid and re-referenced offline to an average of the left and right mastoids. EEG was recorded by Grass bio-amplifiers with a bandpass of .01-100 Hz at a sampling rate of 250 Hz. Trials contaminated by artifacts (e.g., eye movements or blinks) were not included in analyses.

Grand average ERPs to sentence-final words were computed across all 26 recording sites for each experiment and by Ending Type (3a: Supported / Unsupported; 3b: Supported / Related / Unrelated). For statistical analyses, we used linear mixed effects models and focused on a region of interest (ROI) where N400 effects are typically largest, including an average of 8 centro-parietally distributed channels (MiCe, LMCe, RMCe, MiPa, LDPa, RDPa, LMOc, and RMOc) in a canonical N400 time period (250-500 ms) relative to a 200 ms pre-stimulus baseline.

Results

ERPs from our centro-parietal ROI are shown in Fig. 3. ERPs to critical words are characterized by N1 and P2 sensory components. Across all participants, the P2 is followed by a relative negativity (N400), which is most reduced for Supported words compared to Unsupported (Control) and Unrelated/Related (HP) words.

Experiment 3a: Control sentences. Our primary aim in analyzing the control experiment was to ask (a) whether individuals, irrespective of their degree (or depth) of HP knowledge, would elicit standard N400 effects to contextually supported vs. unsupported words in sentences about general topics and (b) to determine whether HP knowledge influenced the size of this effect. We predicted that HP knowledge would have a specific influence on sentences about HP, but not on the size of the effect for control sentences. Our results confirmed this prediction. We observed main effects of ending type (Supported < Unsupported; $p < .0001$) and HP knowledge (individuals with greater HP knowledge tended to yield overall somewhat more positive-going N400 potentials; $p < .01$), but critically, no interaction between ending type and HP knowledge. That is, the size of individuals' N400 reduction to contextually supported, compared to unsupported, words did not differ as a function of individuals' degree of HP knowledge.

Experiment 3b: HP sentences. We expected that HP knowledge would modulate N400 amplitude to contextually supported words in sentences about HP, as in previous studies. In addition, we asked whether HP knowledge would also modulate N400 amplitude to contextually unsupported, but related, words in HP sentences. We observed a reliable interaction between HP knowledge and ending type ($p < .05$) and followed up using planned comparisons examining (a) the Unrelated vs. Supported endings and (b) the Unrelated vs. Related endings, finding that HP knowledge interacted with ending type in both cases (both ps < .05). Follow-up analyses revealed that HP knowledge was correlated with N400 amplitudes to Supported words (Pearson's $r = .57$, $p < .0001$) and Related words (Pearson's $r = .47$, $p < .001$), but not to Unrelated words (n.s.).

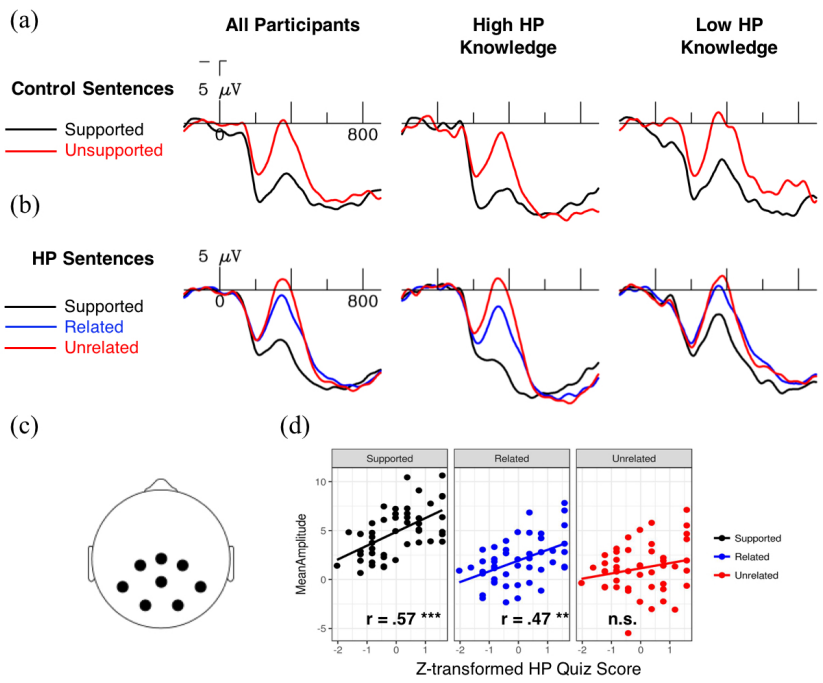


Fig 3. ERPs are plotted to critical words from Control (a) and HP (b) sentences for an ROI based on an average of 8 centro-parietal channels (c). For Control sentences, across all participants, Supported words elicited a reduced N400 compared to Unsupported words; there was no interaction between ending type and degree of HP knowledge. For HP sentences, across all participants, Supported words showed reduced N400 amplitude compared to Unrelated words, with Related words eliciting an intermediate N400. Reductions in N400 amplitude for both Supported and Related words were largest for individuals with high HP knowledge and smallest for individuals with low HP knowledge. (d) N400 amplitude is plotted against HP knowledge; the correlation between the two was strong for Supported and moderate for Related words, but was not significant for Unrelated words.

Due to some differences present in similarity metrics for category- vs. related-anomaly subgroups of items (Exp. 1-2), we also tested whether there were systematic differences in the N400 response between subgroups of materials. However, linear mixed effects models revealed no interaction of ending and related anomaly type on N400 potentials nor any interaction between these predictors and HP knowledge.

To rule out the possibility that other existing individual differences (namely reading experience, general knowledge, and verbal working memory scores) could better account for the observed variability in N400 ERPs, we tested a model that incorporated fixed effects of ending type, HP domain knowledge, general knowledge scores, reading span scores, and aggregate reading experience scores along with interaction terms for each individual differences measure with ending type. We compared this model and a similar model that did not incorporate interaction terms with any individual differences measures (except for the HP domain knowledge-by-ending type interaction term), and found that the more complex model did not explain additional variance.

Discussion

We asked whether, when, and to what extent individuals' degree of domain knowledge of the fictional world of HP would reliably influence the availability of meaningfully relevant information during written language comprehension, even when it was linguistically unexpected. To that end, we assessed a set of materials in which sentence contexts set up expectations for contextually supported words, along with sentence endings that were contextually unsupported, but were meaningfully related or unrelated to the sentence contexts and/or to the supported endings. In a word-by-word reading ERP study, we probed the extent to which real-time access to the same sentence endings was modulated by domain knowledge.

Importantly, individuals' degree of HP knowledge did not influence the size of the contextual support effect for Control sentences about general topics. Replicating Troyer & Kutas (2018), we found that N400 reduction to supported words was strongly predicted by degree of domain knowledge. Moreover, we observed a similar pattern for critical words that were contextually unsupported, yet related to the sentence context and/or supported word.

These results suggest that variation in knowledge—even of a fictional narrative world—influences what knowledge is retrieved in real time, which we believe is likely to reflect the way that knowledge is functionally organized. Our results further suggest that having relatively more knowledge, and thereby more organization around categories and/or events, allows for quick availability of this relevant organization during real-time reading. That is, knowledgeable individuals can quickly (pre)-activate relevant featural and/or thematic (as in the event-related subset of items) information—the very knowledge that is needed to make sense of words in real time. Moreover, individuals' degree of knowledge seems to predict the likelihood with which and/or extent to which such information becomes available for use. These methods and findings invite new research using knowledge-based individual differences to better understand how language processing interfaces with knowledge in real time. For example, future work could combine subject-level domain knowledge and sentence-and-word-level similarity and/or relatedness measures (e.g., based on computational models (Exp. 1) or human judgments (Exp 2.)) to investigate their joint influences at the individual trial level.

References

Altmann, G.T.M. & Mirkovic, J. (2009). Incrementality and prediction in sentence processing. *Cognitive Science*, *33*, 583-609.

Coronel, J.S. & Federmeier, K.D. (2016). The N400 reveals how personal semantics is processed: Insights into the nature and organization of self-knowledge. *Neuropsychologia*, *84*, 36-43.

Daneman, M. & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466.

Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R. (Eds). (2006). *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK: Cambridge University Press.

Federmeier, K.D. & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469-495.

Federmeier, K.D., McLennan, D.B., de Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*, 133-146.

Filik, R. & Leuthold, H. (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs. *Brain Research*, *1506*, 94-104.

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438-441.

Kutas, M., DeLong, K.A., & Smith, N.J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the Brain*. Oxford, UK: Oxford University Press, pp. 190-207.

Kutas, M. & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463-470. doi: 10.1016/S1364-6613(00)01560-6

Laszlo, S. & Federmeier, K.D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*, 326-338.

Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, *4*, 151-171.

Metusalem, R., Kutas, M., Urbach, T.P., Hare, M., McRae, K., & Elman, J.L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*(4), 545-567.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of Workshop at ICLR*.

Mikolov, T., Sutskever, I, Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances on Neural Information Processing Systems*.

Stanovich, K.E. & West, R.F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, *112*(1), 1-36.

Troyer, M. & Kutas, M. (2018). Harry Potter and the Chamber of *What*?: The impact of what individuals know on word processing during reading. *Language, Cognition, and Neuroscience*, 1-17.

Van Petten, C. & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, *18*(4), 380-393.