4-2020

Underspecification in Language Processing

Adam An

Underspecification in Language Processing

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Arts in Linguistics from
The College of William and Mary

By

Adam An

Accepted for Honors

Dan Parker
____
Dan Parker, Director

____
Kaitlyn Harrigan

Anya Lunden
____
Anya Lunden

Timothy A Davis
____
Timothy Davis

Williamsburg, VA
April 27, 2020

Abstract

Memory retrieval during sentence comprehension is sensitive to the degree of match between a retrieval's cues and the cues of a candidate word encoded in memory. Should a candidate word's cues mismatch with the cues of the retrieval, a disruption in retrieval latency may be experimentally evidenced by a measured disruption in reading time. However, it is unclear how this degree of match is determined for words that are encoded in memory with underspecified cues, and it is unclear how the retrieval mechanism responds to these. The current study tests how the retrieval mechanism is affected when presented with words with underspecified cues. Results from a self-paced reading activity show that such words do not cause delays in reading time; comparing these results to computationally generated predictions suggests that underspecified cues are treated like matching cues, not mismatching cues, in the retrieval process. The findings augment our understanding of the retrieval mechanism by accounting for its response to a greater variety of words.

Introduction

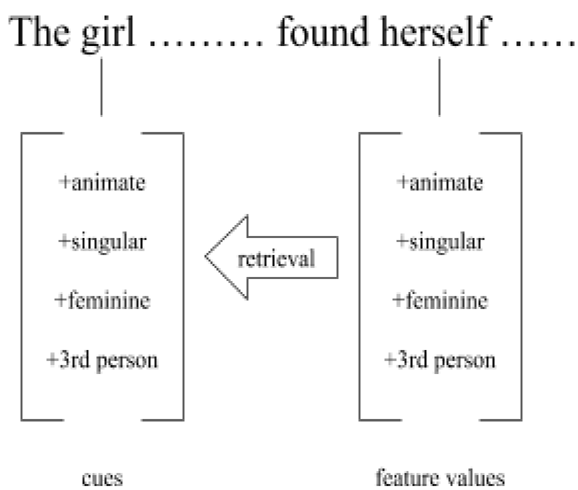
Real-time language processing and comprehension require the use of working memory (Chomsky & Miller, 1963; Fodor, Bever, & Garrett, 1974; Frazier & Fodor, 1978; Jarvella, 1971; Jarvella & Herman, 1972; Kimball, 1973, 1975; McElree, 2000; McElree, Foraker, & Dyer, 2003; Miller & Chomsky, 1963; Miller & Isard, 1963; Van Dyke, 2007; Van Dyke &

Lewis, 2003; Van Dyke & McElree, 2006, 2007, 2011; Wundt, 1904), as exemplified by "long-distance dependencies" involving relationships between nonadjacent words. A typical long-distance dependency, a reflexive-antecedent dependency, may be seen in (1):

(1) **The girl**, separated from the tour group, suddenly found **herself** in an unfamiliar place.

There is strong precedent for using reflexive-antecedent dependencies like (1) to investigate the cognitive properties of memory retrieval (Chen, Jäger, & Vasishth, 2012; Cunnings & Felser, 2013; Cunnings & Sturt, 2014; Dillon, 2014; Dillon, Chow, & Xiang, 2016; Jäger, Engelmann, & Vasishth, 2017; Kush & Phillips, 2014; Parker & Phillips, 2017; Patil, Vasishth, & Lewis, 2016; Sturt, 2003; Xiang, Dillon, & Phillips, 2009). To process these sentences, a comprehender needs to deduce that the word *herself* is related to an earlier "licensor" or "antecedent", *the girl*. Current models of how words are cognitively related to earlier licensors rely on a retrieval mechanism. During initial sentence processing, words are linguistically encoded in memory. Then, words like *herself* that require licensors trigger a retrieval to recover the mental encoding of that licensor among the other encodings in memory (Badecker & Straub, 2002; Clifton, Frazier, & Deevy, 1999; Cunnings & Felser, 2013; Cunnings & Sturt, 2014; Dillon, Mishler, Sloggett, & Phillips, 2013; Nicol & Swinney, 1989; Parker & Phillips, 2017; Runner, Sussman, & Tanenhaus, 2006; Sturt, 2003; Xiang et al., 2009).

More specifically, current models propose that the retrieval mechanism is *content-addressable* (Dillon, Chow, & Xiang, 2016; Dillon, Mishler, Sloggett, & Phillips, 2013; Foraker & McElree, 2007; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Martin & McElree, 2008, 2009, 2011; Martin, Nieuwland, & Carreiras, 2012; McElree, 2000; McElree, Foraker, & Dyer, 2003; Parker & Lantz, 2017; Parker & Phillips, 2016, 2017; Tanner, Nicol, & Brehm, 2014; Tucker, Idrissi, & Almeida, 2015; Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006, 2011; Vasishth, Brüssow, Lewis, & Drenhaus, 2008; Wagers, Lau, & Phillips, 2009). A retrieval triggered by a certain word contains a subset of syntactic and semantic features ("cues") of that word's licensor (e.g. the licensor of *herself* must have +subject, +3rd person, +singular, +feminine, etc.) (Parker & Phillips, 2017). This collection of cues is compared in parallel to the features of each preceding word in the sentence, and the number of matching cues between the retrieval and each preceding word contributes to that word's "activation"[1]. The most-activated word is then selected as the licensor and returned as the result of the retrieval.



[1] A word's activation also depends on other conditions, such as the time since hearing the word and the presence of "distractor items" that resemble the intended licensor (Van Dyke & McElree, 2011; Wagers, Lau, & Phillips, 2009).

Each match between a retrieval feature and a candidate word feature (e.g. +feminine in both *the girl* the retrieval triggered by *herself*) contributes to the candidate word's activation. Parker (2019) provided evidence that the effect of multiple cues on activation is multiplicative: the contributions of each cue's match (or lack thereof, in which case the contribution is 0 or near-zero) are multiplied, meaning that a single mismatch between any corresponding features of the retrieval and candidate word causes a sharp reduction in activation. Consider (2):

(2) * **The girl** found **himself** in an unfamiliar place.

Here, the retrieval triggered by *himself* contains +masculine, but *the girl* is +feminine. Despite all other features between *himself* and *the girl* being identical, this single mismatch on the gender feature causes a decrease in activation on *the girl*, which hinders further consideration of *the girl* as the licensor of *himself* as the retrieval continues (more specifically reducing the probability that the word would be retrieved, as reviewed in Caplan & Waters, 2013; Clark & Gronlund, 1996, and Jonides et al., 2008).

The question of how exactly cue matching (and mismatching) impacts the dynamics of memory retrieval in sentence comprehension has only started to be explored systematically in recent work. To date, most studies have focused on contexts involving cases where all of the relevant feature values are known and fully specified in memory (Parker, 2019; Parker & Phillips, 2017). For example, in (1) and (2), all relevant properties of *the girl* are known: it is +subject, +3rd person, +feminine, etc.

Currently, it is unknown how this matching process at retrieval handles cases when a feature of a candidate word has an underspecified value, as in (3):

(3) **The child** found **herself** in an unfamiliar place.

Specifically, in (3) it is unclear whether *the child* is +feminine or +masculine. +feminine must be in the retrieval triggered by the word *herself*, as its mismatch with the +masculine of *the boy* in (2) is the cause of *the boy*'s lowered activation. Therefore, when processing (3), the matching process is forced to compare the +feminine of *herself*'s expected licensor with the underspecified gender feature of *the child*.

Such cases are prominent in natural language. A common case is when animals or inanimate objects, with no inherent gender, are referred to with (necessarily-gendered) animate pronouns, such as in children's literature or descriptions of fictional characters:

Train *reaches a broken fence with a Do Not Enter sign.* ***He*** *pauses momentarily to consider whether to trespass or not.* (Mai, 2020)

Any theory of memory retrieval for sentence comprehension must be equipped to handle such cases, but it is unclear how current models might handle such cases, as existing empirical research has largely ignored the impact of feature underspecification in the encoding of linguistic structure in memory. This study seeks to better understand how feature underspecification impacts real-time language processing using a combination of computational and behavioral methods.

Previous work on retrievals

Computational models of memory retrieval seek to more precisely define the general concept of activation and calculate the hypothesized activation values for words at the point of retrieval. One such model is the ACT-R model of memory retrieval (Lewis & Vasishth, 2005). Under ACT-R, the contribution that each retrieval feature Q_j, compared with a candidate item I_i, makes towards activation is its strength of association, written as $S(Q_j, I_i)$; features that match with the item I_i have a higher strength of association than those that do not, typically defined in terms of maximal similarity (e.g. 1) and maximal dissimilarity (e.g. 0). These strengths of association may be given weights W_j, producing the final contribution to activation $W_jS(Q_j, I_i)$ (although typically it is assumed that all strengths of association are weighted equally). The total activation of an item I_i is a function of all of these strengths of association from each cue in a retrieval; the strengths of activation must somehow be combined to produce a final activation value A_i.

Some models use a linear cue combination method where strengths of association are summed, as in Eq. 1 (used in Anderson, 1990; Anderson et al., 2004; Lewis & Vasishth, 2005; Vasishth, Brüssow, Lewis, & Drenhaus., 2008):

(1) $$A_i = \sum_{j=1}^{n} W_j S(Q_j, I_i)$$

In these models, each matching cue contributes directly to activation, with the implication that an item with only some of its features matching a retrieval's features will still be activated. A linear increase in the number of mismatches corresponds to a linear decrease in activation.

However, nonlinear cue combination methods instead use the product of strengths of association, as in Eq. 2 (used in Gillund & Shiffrin, 1984; Hintzman, 1984; Nairne, 1990; Parker, 2019; Raaijmakers & Shiffrin, 1981; Van Dyke, 2007; Van Dyke & McElree, 2006):

$$(2) \qquad A_i = \prod_{j=1}^{n} W_j S(Q_j, I_i)$$

In these models, a single mismatching cue, with its low strength of association, can drastically lower the overall product, even in the presence of many matching cues with high strengths of association. In these models, a linear increase in the number of mismatches corresponds to an exponential decrease in activation. Effectively, activation is "all-or-nothing": items whose features do not match all of a retrieval's features exactly will have little to no activation.

An item's activation may be observed through its effect on its reading time (Anderson & Milson, 1989). A typical function that relates activation and retrieval time is Eq. 3, where T_i is the time needed to retrieve and process an item i. F and f are scaling constants that scale predicted retrieval times to better reflect the scale of any experimental reading times that the predictions may be compared to. Roughly, the lower an item's activation is, the higher the time it takes to read it.

(3) $T_i = F e_i^{-(f \times A_i)}$

Parker (2019) found evidence for a nonlinear cue combination model by examining sentences that prompted retrievals involving no, one, or two mismatches, e.g. (3a), (3b), and (3c).

(3) a. **The man** recently hurt **himself** at the park.

b. * **The woman** recently hurt **himself** at the park.

c. * **The women** recently hurt **himself** at the park.

In all of the above sentences, the word *himself* triggers a retrieval that looks for a +singular, +masculine antecedent. In (3a), the target item *the man* matches all of these cues exactly and receives high activation. However, in (3b), the target item *the woman* is +feminine and therefore has one feature mismatch with the retrieval. In (3c), the target item *the women* is both +feminine and +plural and therefore has two feature mismatches with the retrieval.

A linear cue combination model would predict that the activation of the target items would decrease linearly with an increasing number of feature mismatches, while a nonlinear cue combination model would predict that the activation of the target items would decrease exponentially. By relating activation to reading time delay, Parker showed that the nonlinear cue combination model more accurately reflected the target items' actual activations.

However, Parker also tested sentences like (4):

(4) **We** recently hurt **himself** at the park.

This target item *we* has feature mismatches with the retrieval from *himself* and patterned like (3b) and (3c) in that it received little to no activation. However, a challenge in analyzing *we* is that its gender feature is underspecified, making it unclear what its strength of association is, or even how many mismatches *we* has in the retrieval from *himself*. As discussed above, items with similarly underspecified features are common in language; it is important to have an account of their handling to better understand what determines success or failure in language comprehension and to guide elaboration of the computational model.

Current study

This study uses behavioral and computational methods to investigate the behavior of retrieval probes when they encounter candidate words with underspecified features. More specifically, we seek to better understand the effect that an underspecified feature has on an item's activation, as compared to that of fully-defined features that either match or mismatch with a retrieval's cues, using reading time data as a proxy measure for retrieval time and activation. If an underspecified feature behaves like a matching feature, then its presence should not incur an activation penalty. However, if an underspecified feature behaves like a mismatching feature, then its presence should prompt an activation penalty of similar degree to the penalty seen with fully-defined mismatching features. It is also possible that underspecified features trigger an activation penalty that is not as severe as that of a mismatching feature penalty. In effect, the effects on activation of fully-defined matching and mismatching features form the two endpoints on a spectrum; this study seeks to place the effect of underspecified features on that spectrum.

To address this question, this study used reflexive-antecedent dependencies to trigger retrievals involving underspecified features. Experiment 1 used computational modeling to predict delays in retrieval time caused by underspecified features when treated as matches or mismatches. Experiment 2b measured these retrieval time delays using a self-paced reading activity. This experiment compared target items that either matched a retrieval's cues fully, had one mismatching feature, or had one underspecified feature, for comparison with the modeled predictions from Experiment 1. Experiment 2b showed that target items with underspecified features patterned like target items with no mismatch at all, suggesting that underspecified features do not count as mismatching features and that their strengths of association are comparably high to those of matching features.

Experiments

Experiment 1: computational modeling

Procedure

The ACT-R model explained above was implemented in the R software environment (R Development Core Team, 2018). Then, retrievals were done on sentences with reflexive-antecedent dependencies after manipulating the degree of match between their reflexives and antecedents, as shown in Table 1. Input into the model only consisted of the target word and the retrieval-triggering word, along with the cues of lexical category, gender, number, and animacy[2], and a time in milliseconds representing the time after the start of a sentence a word becomes encoded in a comprehender's memory. Target words were given an encoding time of 100 ms; retrieval-triggering words were given an encoding time of 2000 ms. Example input files are shown in Table 2.

Predictions for the effect of target items with no or one feature mismatches on retrieval time were generated first. Then, two more retrieval-time predictions were generated for two possible treatments of target items with an underspecified feature: in one scenario, the underspecified feature was treated as a match, and in the other, the underspecified feature was treated as a mismatch.

To model the possibility that underspecified features were treated as matches, the underspecified gender feature was given an arbitrary filler value (the string "NULL"). Then, the model code was modified to look for instances of this filler value and always count them as matches. To model the possibility that underspecified features were treated as mismatches, the underspecified features were given a different arbitrary filler value (the string "mismatch"), which was avoided in the retrieval cues to ensure that when the retrieval cues and target item features were compared, a mismatch would be found.

---

[2] Lexical category is a cue required in inputs given to the implementation of the ACT-R model that was used. Number and animacy were arbitrarily chosen to represent the unknown number of (matching) cues that are actually present in retrievals of this type and to demonstrate the effects of their presence on activation (i.e., no penalty).

Table 1

Sample items for Experiment 1

<table>
<tr><th>Degree of match between reflexive and antecedent</th><th>Sentence</th><th>Target item gender feature</th></tr>
<tr><td>Full match</td><td>The girl found herself in an unfamiliar place.</td><td>feminine</td></tr>
<tr><td>One feature mismatch</td><td>* The boy found herself in an unfamiliar place.</td><td>masculine</td></tr>
<tr><td>One underspecified feature treated as a match</td><td rowspan="2">The child found herself in an unfamiliar place.</td><td>NULL (filler value specially treated as match)</td></tr>
<tr><td>One underspecified feature treated as a mismatch</td><td>mismatch (mismatching filler value)</td></tr>
</table>

Notes. The symbol “*” indicates ungrammaticality. The gender feature of the retrieval triggered by “herself” was always feminine.

Table 2

Sample input files for Experiment 1

Target word (“girl”)	Retrieval-triggering word (“herself”)

`name       woman`
`created    100`
`cat        NP`
`gender     f`
`number     sg`
`animacy    true`
	`moment     600   2000`
	`cat        NP    NP`
	`gender     m     m`
	`number     sg    sg`
	`animacy    true  true`

Notes. For the implementation of the ACT-R model that was used, input files defining the retrieval-triggering word require at least two words or columns to be defined, hence the two duplicate columns that define the same word, differing only in the time created. The program effectively ignores the first instance of the word definition; therefore, this input file models only one retrieval.

Following Parker (2019), matching cues were assigned high strengths of association (1), and mismatching cues were assigned low strengths of association (0). The resulting activation values were normalized[3] to give a fully-matching item an activation of 1 before retrieval times were calculated. It was assumed that the retrieval time predictions were monotonically related to experimental reading time measures (following precedent set by Anderson, Budiu, & Reder, 2001; Boston, Hale, Vasishth, & Kliegl, 2011; Dillon, Chow, Wagers, et al., 2016; Dillon et al., 2013; Jäger, Engelmann, & Vasishth, 2015; Kush & Phillips, 2014; Lewis & Vasishth, 2005; Nicenboim, Logacev, Gattei, & Vasishth, 2016; Nicenboim & Vasishth, 2018; Patil, Vasishth, & Lewis, 2016; Tucker, Idrissi, & Almeida., 2015; Vasishth, Brüssow, Lewis, & Drenhaus, 2008, and others). Other factors do affect reading times (interpretation time, reanalysis, etc.), but these can be assumed to not affect the monotonic relationship between retrieval time and reading time.

---

[3] Most cue-based memory models assume that, in sentences with multiple candidate items, the probability of retrieving an item is proportional to its activation (Clark & Gronlund, 1996; Lewis & Vasishth, 2005). Normalizing activation values this way allows them to be interpreted as probabilities [0, 1].

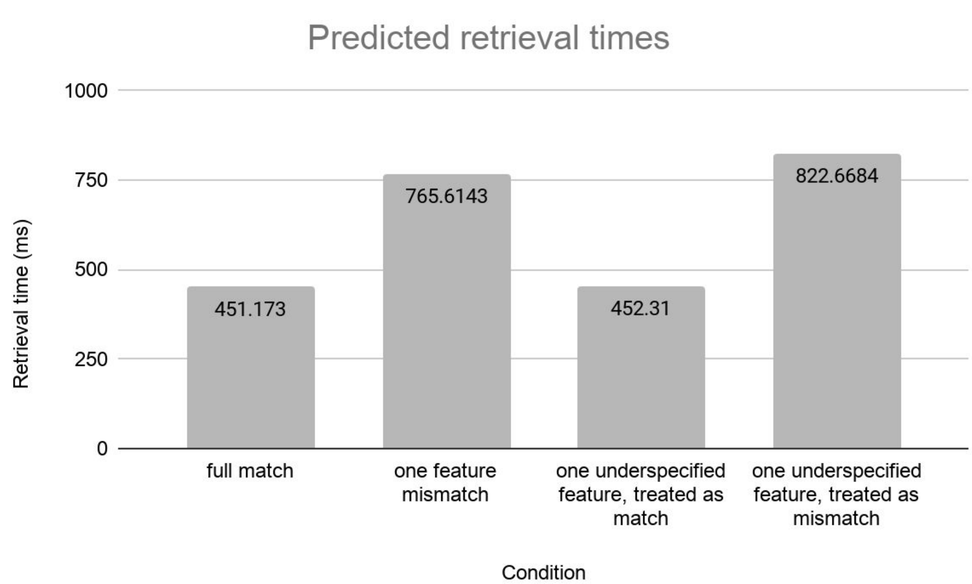
Results and discussion

Figure 1 shows predicted retrieval times for the four conditions. We notice that retrievals for words with fully-matching features have a low retrieval time, while retrievals for words with a mismatching feature have a higher retrieval time.

We may also compare predictions for the two accounts of retrieval for underspecified features. If underspecified features are treated as matches, we predict that retrieval times will be low and resemble those from sentences with full matches; if underspecified features are treated as mismatches, we predict that retrieval times will be high and resemble those from sentences with one feature mismatch.

Figure 1

Predicted retrieval times for conditions tested by Experiment 1



Experiment 2a: Verifying underspecification for gender

To investigate the behavior of nouns that were underspecified for gender, it was first necessary to verify which nouns a majority of the population considers to be underspecified gender. To do this, 25 native speakers of American English were given a list of human nouns and asked to indicate whether they thought the noun was “Male”, “Female”, or “No gender”. Responses were collected via Google Forms; participants were not compensated for their participation.

Nouns that 80% or more of participants selected “No gender” for are assumed to be underspecified for gender in future experiments.

Experiment 2b: Self-paced reading

Experiment 1 laid out predictions for two possible treatments of underspecified features and their effects on activation. Previous studies have shown that reading times in self-paced reading activities can serve as an accurate proxy for retrieval time (and, by extension, activation) (Anderson & Milson, 1989); therefore, Experiment 2 used self-paced reading to investigate which prediction more closely resembled the actual cognitive treatment of these underspecified features.

Methods

One hundred sixteen native speakers of American English were recruited from the College of William & Mary undergraduate population. Participants received participant pool credit.

The sentences used in the experiment were 18 three-item sets of the form:

a) **The girl** accidentally pricked **herself** on the cactus on the windowsill.
b) **The boy** accidentally pricked **herself** on the cactus on the windowsill.
c) **The child** accidentally pricked **herself** on the cactus on the windowsill.

Each sentence consisted of the word “the”, a subject noun that was the antecedent of the following reflexive, a filler adverb, a verb, a reflexive pronoun, and a section of words at the end of the sentence of variable length and structure (“spillover region”). The reflexive was always a gendered singular third-person form (i.e., either *himself* or *herself*). The specific reflexive was constant across the conditions of an item set; half of the item sets used *himself* while the other half used *herself*.

Within each item set, the subject noun varied in its degree of match with the reflexive. Item a) matched on every feature; item b) mismatched on the gender feature. Item c) used a noun from Experiment 1 assumed to be underspecified for gender.

The 18 item sets were divided into three lists following a Latin square design, ensuring each list contained equal numbers of *himself* and *herself* item sets. The 18 items were then combined with 36 grammatical filler sentences that resembled the experimental sentences in length and complexity.[4] To prevent participants from only partially reading the sentences, all items were followed by a yes-or-no comprehension question that asked about different parts of the sentence.

A full list of items and filler sentences is provided in the appendix.

Procedure

This experiment was conducted using Ibex (http://spellout.net/ibexfarm), an online experiment platform that allows running self-paced reading experiments in a web browser. Sentences were first displayed with each word replaced with a dash, though with whitespace and punctuation unchanged. Participants pushed the space bar to reveal each word of the sentence in order; when one word was revealed, the previous word was again concealed with dashes, such that the participant only saw one word at a time. The time each participant spent on each word was recorded. Each sentence was followed with its comprehension question; participants were asked to respond with "yes" or "no" and provided with an on-screen message if their answer was incorrect. Presentation was randomized for each participant.

The experiment included an instructional manipulation check to ensure that participants completed the task correctly (Oppenheimer, Meyvis, & Davidenko, 2009). Prior to the experiment, participants were asked to respond to questions about the instructions (e.g. *What*

[4] Each participant saw 6 ungrammatical sentences (item b)) and 6 sentences with nouns underspecified for gender (item c), for which I reserve any grammaticality judgements). The remaining 42 sentences were grammatical.

button do I press to advance a word? or *What button do I press to respond "Yes" to a comprehension question?*).

Analysis

Data from all participants were included in the analysis. Average reading times were compared across conditions in the following regions: the retrieval-triggering reflexive pronoun itself (*critical region*), and the two words immediately following the reflexive pronoun (*spillover regions 1 and 2*). Statistical analyses were performed with linear mixed-effects models over the log-transformed reading times to account for the log-normal distribution of reading times (Box & Cox, 1964; Ratcliff, 1993; Vasishth & Nicenboim, 2016). Linear mixed-effects models were constructed using the *lme4* package (Bates, Maechler, & Bolker, 2011) in the R software environment (R Development Core Team, 2018). Models were defined to compare the effects of cue match between the three experimental conditions:

Cue Match: underspecified vs. match (item c) vs. item a))

Cue Mismatch: underspecified vs. mismatch (item c) vs. item b)).

The structure of the models was as follows:

lmer(RT ~ C1 + C2 + (1 | Item_number) + (1 | Subject))

Fixed effects with absolute t values > |2| were considered significant, since their 95% confidence intervals did not include 0 (Gelman & Hill, 2007). Reading times were then compared to the retrieval times predicted by the model in Experiment 1.

Results

Figure 2 shows the average raw reading times by region and condition and Table 3 shows the results of the statistical analyses. Results across all three regions of interest showed a main effect of cue mismatch, such that sentences with a cue mismatch were read more slowly than sentences with an underspecified mismatch. No effect of cue match (i.e., match vs. underspecified) was observed in any region of interest, as the reading times for the underspecified condition did not diverge from the match condition in a statistically significant manner.

Figures 3.1 and 3.2 compare the reading times collected in Experiment 2b with the predicted retrieval times generated in Experiment 1. Visually inspecting the two graphs, for sentences with matching features and mismatching features, the predicted retrieval times and experimental reading times pattern alike: matching features low, mismatching features high. The low / high profile is seen again in the predicted retrieval times for the two treatments of underspecified features: low if treated as matches, high if treated as mismatches. The experimental reading time for underspecified features is low, aligning with the predicted retrieval time for underspecified features treated as matches.

Figure 2.2

Mean raw reading times by region and condition for Experiment 2b.

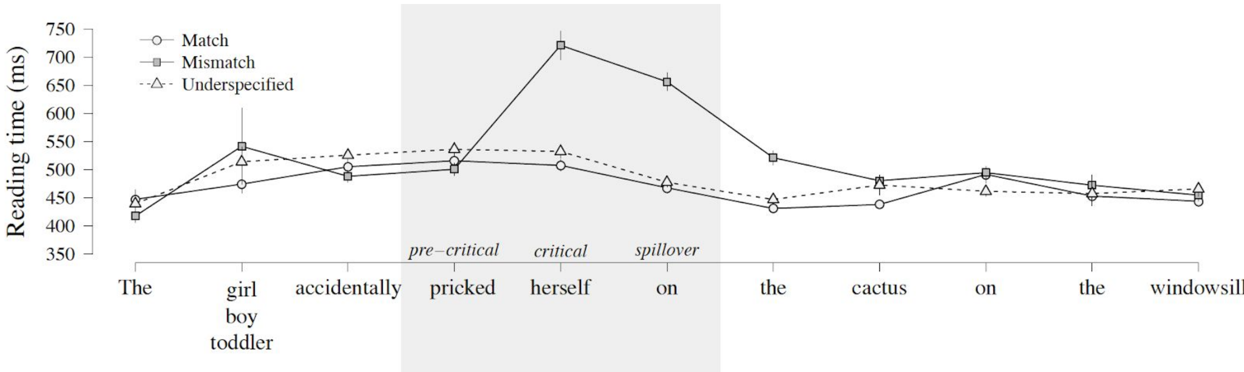


Figure 2.2

Mean logged reading times by region and condition for Experiment 2b.

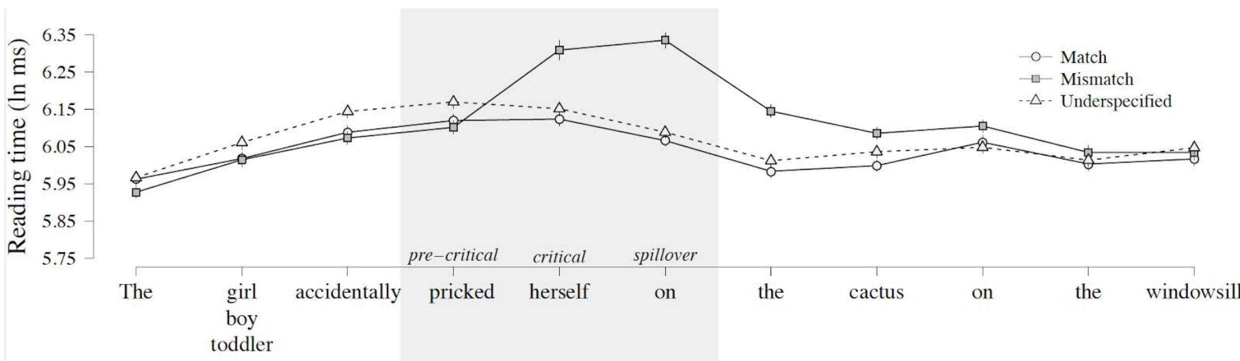


Table 3

Summary of statistical analyses for conditions tested in Experiment 2b.

	critical (word 5)			spillover 1 (word 6)			spillover 2 (word 7)		
	$\widehat{\beta}$	SE	t	$\widehat{\beta}$	SE	t	$\widehat{\beta}$	SE	t
Cue Match	0.02809	0.02352	1.19	0.02267	0.02016	1.12	0.02866	0.01790	1.601
Cue Mismatch	**0.18497**	**0.02352**	**7.86**	**0.26925**	**0.02016**	**13.36**	**0.16082**	**0.01790**	**8.985**

Significant coefficients (|t| > 2) are shown in bold.

Figure 3.1

Raw reading times from Experiment 2b (top).

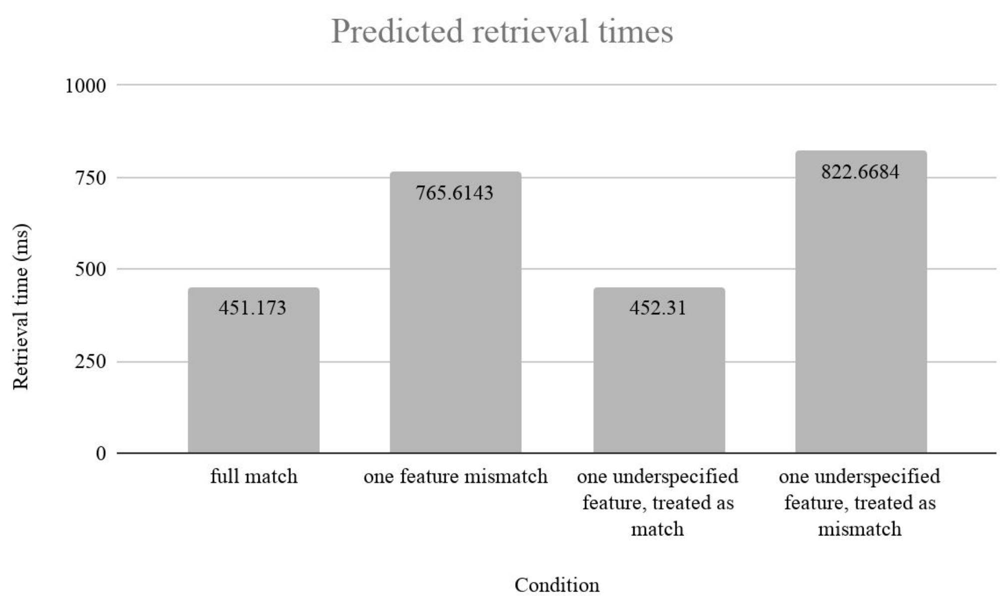
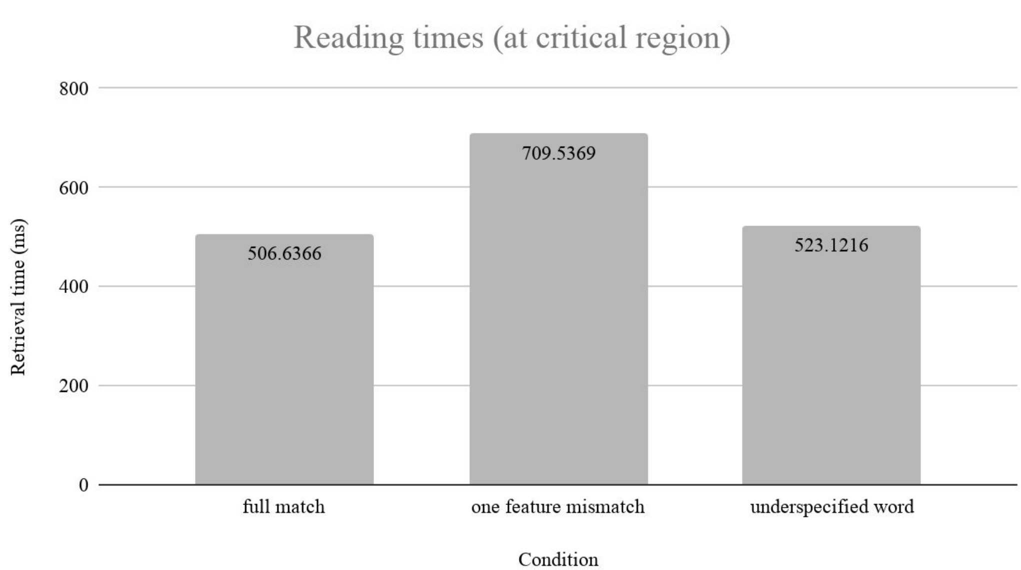


Figure 3.2

Predicted retrieval times from Experiment 1 (bottom).



Discussion

This experiment tested the behavior of retrieval for reflexive licensing encountering candidate items with underspecified features and compared that behavior with the predictions generated in Experiment 1 to determine whether underspecified features were treated more like matches or mismatches. The results of Experiment 2b first demonstrate that sentences with feature mismatches cause delays in reading time that are not seen in those without. This replicates the effect of feature mismatches seen in previous studies (Parker, 2019; Parker & Phillips, 2017). Secondly, these results suggest that underspecified features on words do not cause delays in reading time like mismatching features do. Rather, retrievals encountering these words pattern like retrievals encountering words with matching features.

Turning our attention to the predicted retrieval times generated in Experiment 1, we notice that retrievals for words with fully-matching features are predicted to have a relatively low retrieval time; retrievals for words with mismatching features are predicted to have a relatively high retrieval time. We then see a difference in predicted retrieval times for two different accounts of retrieval: Under an account that treats underspecified cues as matches, we see predicted retrieval times for underspecified features pattern like the lower times of fully-matching features. But under an account that treats underspecified cues as mismatches, we instead see predicted retrieval times for underspecified features pattern like the higher times of mismatching features.

Comparing the two sets of results, we see that for both predicted retrieval times and experimental reading times, times for retrievals with fully-matching features are low, while times for retrievals with mismatching features are high. This supports the use of experimental reading time as a proxy for experimental *retrieval* time, and we may assume that experimental retrieval times have a similar profile: fully-matching features low, mismatching features high.

We then notice that experimental reading times for retrievals with underspecified features are low, just like those of fully-matching features. This suggests that the experimental retrieval time for underspecified features was low and also patterned with the fully-matching features. In the modeling, of the two accounts of retrieval for underspecified features, only the retrieval procedure that treated underspecified cues as matches predicts reading times that are similarly low. The experimental results are therefore consistent with an account of retrieval that treats underspecified cues as matches, not mismatches.

General Discussion

The goal of the current study was to investigate words whose representations in memory require certain cues to be underspecified and test their effect on the retrieval process during sentence processing. Experiment 1 proposed two different possible effects on the retrieval process; Experiment 2b behaviorally observed the actual effect. Results were consistent with a model of the memory retrieval process that does not treat underspecified cues as mismatches, suggesting that the presence of underspecified cues on a candidate word being considered by a retrieval does not cause a decrease in that word's activation and a delay in the retrieval latency.

One concern with the current results is the possibility of different processing times or retrieval latencies for the pronouns *himself* and *herself*. Each participant received equal numbers of both pronouns across test conditions, motivating the assumption that any effect of pronoun gender would not be reflected in the differences seen in reading times seen across conditions. However, we may also examine reading times at the critical region (i.e., the reflexive pronoun) to look for a difference between *himself* and *herself* reading times:

Table 4

Mean raw critical region reading times (ms)

	himself	***herself***
Condition a) (full match)	502	501
Condition b) (one feature mismatch)	716	679
Condition c) (underspecified mismatch)	502	536

A visual inspection of the reading times does not suggest a difference between the two pronouns beyond what would be expected from typical sampling variation.

Another concern with the current findings is how the proposed model would extend to cases of underspecified cues *at retrieval*. For example, the reflexive pronoun *itself* is underspecified for gender, prompting the question of how (or whether) this underspecified cue is involved in any retrievals triggered by *itself*. This issue is complicated by the fact all standard English ungendered pronouns must have ungendered antecedents that are also underspecified for gender; these underspecified cues on candidate words have been shown to cause no delay in retrieval latency. Therefore, we would expect retrieval latency to be the same if the underspecified cue were used in a retrieval or if it were ignored, since the retrieval would always be met with a candidate word with a corresponding underspecified cue. However, the nonstandard singular *they* may present an opportunity to test retrievals with underspecified cues against candidate words with defined cues in sentences like (5):

(5) **The girl** found **themself** in an unfamiliar place.

In the above sentence, *themself* is underspecified for gender and triggers a retrieval for the +feminine *girl*. If the *themself* retrieval experiences more delay than that of a gendered pronoun retrieval, it could be attributed to the pronoun's underspecification for gender being crucially encoded in the retrieval cues and mismatching with the +feminine of the candidate word. On the other hand, a lack of delay could point to a lack of mismatch between the underspecification of the retrieval and the +feminine of the candidate word, or indicate that the underspecification was never encoded in the retrieval to begin with.

Davenport (2020) tested sentences like (5) in self-paced reading studies. While Davenport's study did not compare sentences with underspecified-for-gender pronouns directly against sentences with gendered reflexive pronouns, the general profile of reading times across sentences with underspecified retrievals did not see spikes in delay comparable to the spikes caused by fully-specified, mismatching retrievals in Experiment 2b. Therefore, Davenport's study provides preliminary evidence for a lack of delay caused by underspecified retrievals.

Another concern with the results of Experiment 2b is that self-paced reading provides a coarse-grained measure of reading time. It is possible, for instance, that there was indeed a decrease in activation associated with feature underspecification, which might be readily overcome, but the current behavioral measure was not sensitive enough to detect such effects. Reading time may be more precisely measured with methods such as eye tracking; such methods would allow any smaller disruptions in reading time to be discerned.

Possible "fill-in" effects provide further opportunities for study. In sentences like "the **child** found **herself** in an unfamiliar place", the pronoun *herself* not only triggers a retrieval for the noun *child* but also logically indicates that the child is female. We may speculate whether this new information, a +feminine cue, becomes part of the encoding of the instance of *child* that was heard earlier. This could be tested by conducting similar behavioral studies with sentences like (6):

(6) * The **child** found **herself** in an unfamiliar place, but **he** eventually found the way back.

Both pronouns, *herself* and *he,* trigger retrievals for the encoding of *child* in memory. At the retrieval triggered by *herself*, the encoding of *child* is underspecified for gender, and as such, we would expect no retrieval or reading delay on *herself*, as previously established. More interesting is the retrieval triggered by *he*, which also should retrieve the encoding of *child*. However, the behavior of the *he* retrieval depends on whether the *herself* retrieval has updated the encoding of *child*. It could be the case that *child* still lacks a gender cue; if this is the case, we would expect no retrieval or reading delay on *he*. However, if a reading delay on *he* were to be seen, it could be attributed to the encoding of *child* having gained a +feminine cue that mismatched with the +masculine of the *he* retrieval, which would more generally suggest that mental encodings of words with underspecified cues could be augmented by further information presented by words later in a sentence in a way that could be utilized by future retrievals.

If fill-ins do occur, it is unclear whether they cause any delays in processing time. If there are indeed delays that the current study failed to detect, the fill-in process described above could be a cause. Another potential cause, independent of fill-ins, could be that underspecified cues cause delay in the process of integrating a word back into the primary processing stream after a word has been retrieved. If further experiments were to show delays in processing time but a lack of fill-in effects, delay of integration into the primary processing stream could be a potential explanation.

Conclusion

The current study showed that underspecified features did not cause disruptions in retrieval like those of mismatching features. These results suggest that underspecified features pattern like matching features in that they contribute to a word's activation instead of detracting from it, despite not strictly matching the features of a retrieval. These findings contribute to a more complete picture of the memory retrieval mechanism and its behavior when processing the variety of words encountered in natural language.

Appendix: Stimuli

Sentences

Condition a)

The schoolchild quickly made himself a sandwich before running off to catch the bus.
The toddler quietly let herself out of the playpen her parents had forgotten to lock.
The kid quickly made himself comfortable on the big couch in the corner.
The infant finally calmed herself down after crying for hours and hours.
That kid unfortunately burned himself using the stove for the first time.
The toddler accidentally pricked herself on the cactus on the windowsill.

The child suddenly found himself lost in the middle of the grocery store.
The toddler amazingly taught herself how to read in two different languages.
The kid really enjoyed himself at school today.
The student hurriedly prepared herself for the huge test the next day.
The baby thankfully behaved himself in the busy waiting room.
The child suddenly saw herself in the huge mirror.
The kid quietly busied himself with coloring in the coloring page.
The baby accidentally hurt herself crawling down the stairs.
The schoolchild repeatedly pinched himself to stay awake during the long, boring lecture.
The infant seemingly amused herself with the colorful toys in the playpen.
The toddler suddenly shocked himself while playing with an outlet.
The child carefully cleaned herself off in the river.

Condition b)

The schoolboy quickly made himself a sandwich before running off to catch the bus.
The girl quietly let herself out of the playpen her parents had forgotten to lock.
The boy quickly made himself comfortable on the big couch in the corner.
The girl finally calmed herself down after crying for hours and hours.
That boy unfortunately burned himself using the stove for the first time.
The girl accidentally pricked herself on the cactus on the windowsill.
The boy suddenly found himself lost in the middle of the grocery store.
The girl amazingly taught herself how to read in two different languages.
The boy really enjoyed himself at school today.
The schoolgirl hurriedly prepared herself for the huge test the next day.
The boy thankfully behaved himself in the busy waiting room.
The girl suddenly saw herself in the huge mirror.
The boy quietly busied himself with coloring in the coloring page.
The girl accidentally hurt herself crawling down the stairs.
The schoolboy repeatedly pinched himself to stay awake during the long, boring lecture.
The girl seemingly amused herself with the colorful toys in the playpen.
The boy suddenly shocked himself while playing with an outlet.
The girl carefully cleaned herself off in the river.

Condition c)

The schoolgirl quickly made himself a sandwich before running off to catch the bus.
The boy quietly let herself out of the playpen her parents had forgotten to lock.
The girl quickly made himself comfortable on the big couch in the corner.
The boy finally calmed herself down after crying for hours and hours.
That girl unfortunately burned himself using the stove for the first time.
The boy accidentally pricked herself on the cactus on the windowsill.
The girl suddenly found himself lost in the middle of the grocery store.
The boy amazingly taught herself how to read in two different languages.
The girl really enjoyed himself at school today.
The schoolboy hurriedly prepared herself for the huge test the next day.

The girl thankfully behaved himself in the busy waiting room.
The boy suddenly saw herself in the huge mirror.
The girl quietly busied himself with coloring in the coloring page.
The boy accidentally hurt herself crawling down the stairs.
The schoolgirl repeatedly pinched himself to stay awake during the long, boring lecture.
The boy seemingly amused herself with the colorful toys in the playpen.
The girl suddenly shocked himself while playing with an outlet.
The boy carefully cleaned herself off in the river.

Filler items

The dog jumped in the pile of leaves.
The cops chased the criminal.
The doctor accidentally injured the patient.
The door to the bathroom loudly closes.
The boy under the blanket peacefully sleeps.
The baby at the movie annoyingly cries.
The robber at the bank always escapes.
The comedian in the performance frequently laughs.
The traveler at the airport patiently waits.
The pizza in the oven always burns.
The vaccine for the disease fortunately works.
The actor in the play occasionally sings.
The chicken in the pot slowly braises.
The representative for the company often works.
The light in the bedroom ominously flickers.
The tea in the kettle never cools.
The glue for the fabric never sticks.
The billboard on the highway never changes.
The raft on the water never sinks.
The officer on the case never sleeps.
The strategy for the game rarely works.
The lamb quickly pulled itself free from the barbed-wire fence.
The crow carefully preened itself with a twig.
The snake suddenly bit itself on the tail.
The princess recently inspired her servants to be kinder.
The schoolgirl carefully disciplined herself to study more often for the exam.
The attorney accidentally misrepresented his client in the controversial court case.
The seamstress accidentally pinched her finger in the sewing machine.
The reporter openly characterized her congressman as an idiot.
The safety net that the brave policeman used was made of synthetic material.
The secretary accidentally incriminated herself in the office scandal.
The security system that the police officer used was recently installed in the airport.
The slanderous accusation that the wealthy executive disregarded was full of inaccurate information.
The brief memo that the studious schoolgirl noticed was posted on the notice board.
The teacher blamed himself for the students' poor grades.

The newspaper article that the rich executive read was quite misleading about the facts.
The new medication that the sick patient took was very expensive for the insurance companies.
The mural that had been painted by the students unfortunately was painted over by the start of the next semester.
The paragraph that had been removed by the editor ultimately was put back in to the article.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychology Review, 96*, 703–719. https://doi.org/10.1037/0033-295X.96.4.703.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychology Review, 111*, 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036.

Anderson, J. R., Budiu, R., & Reder, L. M. (2001). A theory of sentence memory as part of a general theory of memory. *Journal of Memory and Language, 45*, 337–367. https://doi.org/10.1006/jmla.2000.2770.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 748– 769. https://doi.org/10.1037/0278-7393.28.4.748.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. Available at http://CRAN.R-project.org/package=lme4.

Boston, M. F., Hale, J., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes, 26*, 301–349. https://doi.org/10.1080/01690965.2010.492228.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological), 26*, 211–252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Caplan, D., & Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin & Review, 20*, 243–268. https://doi.org/10.3758/s13423-012-0369-9.

Chen, Z., Jäger, L., & Vasishth, S. (2012). How structure-sensitive is the parser? Evidence from Mandarin Chinese. In B. Stolterfoht & S. Featherston (Eds.), *Empirical approaches to linguistic theory: Studies of meaning and structure (studies in generative grammar)* (pp. 43–62). Berlin: Mouton de Gruyter.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. Duncan Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York, NY: Wiley.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3*, 37–60. https://doi.org/10.3758/BF03210740.

Clifton, C. J., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Rivista di Linguistica, 11*, 11–39.

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. Language and *Cognitive Processes, 28*, 188–219. https://doi.org/10.1080/01690965.2010.548391.

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory & Language, 75*, 117–139. https://doi.org/10.1016/j.jml.2014.05.006.

Davenport, C. M. (2020). *The Use of* Themself *to Refer to a Singular Antecedent* [Unpublished honors thesis]. College of William & Mary.

Dillon, B. (2014). Syntactic memory in comprehension of reflexive dependencies: An overview. Language and Linguistics Compass, 8, 171–187. https://doi.org/10.1111/lnc3.12075.

Dillon, B., Chow, W. Y., & Xiang, M. (2016). The relationship between anaphor features and antecedent retrieval: Comparing Mandarin Ziji and Ta-Ziji. *Frontiers in Psychology, 6*, 1–16. https://doi.org/10.3389/fpsyg.2015.01966.

Dillon, B., Chow, W. Y., Wagers, M., Guo, T., Liu, F., & Phillips, C. (2016). The structure-sensitivity of search: Evidence from Mandarin Chinese. *Frontiers in Psychology, 5*, 1–16

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language, 69*, 85–103. https://doi.org/10.1016/j.jml.2013.04.003.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). The psychology of language. New York: McGraw-Hill.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language, 56*, 357–383. https://doi.org/10.1016/j.jml.2006.07.004.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291– 325. https://doi.org/10.1016/0010-0277(78)90002-1.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge, UK: Cambridge University Press.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67. https://doi.org/10.1037/0033-295X.91.1.1.

Hintzman, D. L. (1984). MINERVA2: A simulation model of human memory. *Behavior Research Methods, Instruments, and Computers, 16*, 96–101. https://doi.org/10.3758/BF03202365.

Jarvella, R. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior, 10,* 409–416. https://doi.org/10.1016/S0022-5371(71)80040-3

Jarvella, R., & Herman, S. (1972). Clause structure of sentences and speech processing. *Perception and Psychophysics, 11*, 381–384. https://doi.org/10.3758/BF03206272.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*, 193–224. https://doi.org/10.1146/annurev.psych.59.103006.093615.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language, 94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.004.

Jäger, L., Engelmann, F., & Vasishth, S. (2015). Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology, 6*, 1–24. https://doi.org/10.3389/fpsyg.2015.00617.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition, 2*, 15–47. https://doi.org/10.1016/0010-0277(72)90028-5.

Kimball, J. (1975). Predictive analysis and over-the-top parsing. In J. Kimball (Ed.), *Syntax and semantics* (pp. 155–179). New York: Academic Press.

Kush, D., & Phillips, C. (2014). Local anaphor licensing in an SOV language: Implications for retrieval strategies. *Frontiers in Psychology, 5*, 1–12. https://doi.org/10.3389/fpsyg.2014.01252.

Lago, S., Shalom, D., Sigman, M., Lau, E., & Phillips, C. (2015). Agreement processes in Spanish comprehension. *Journal of Memory and Language, 82*, 133–149. https://doi.org/10.1016/j.jml.2015.02.002.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375–419. https://doi.org/10.1207/s15516709cog0000_25.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. Trends in Cognitive Science, 10, 447–454. https://doi.org/10.1016/j.tics.2006.08.007.

Mai, E. (2020, March 25). *Curiosity Killed the Train (Project 4)*. Computer Animation. https://edmai48.wixsite.com/website/post/curiosity-killed-the-train-project-4.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language, 58*, 879–906. https://doi.org/10.1016/j.jml.2007.06.010.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1231–1239. https://doi.org/10.1037/a0016271.

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: Evidence from Sluicing. *Journal of Memory and Language, 64*, 327–343. https://doi.org/10.1016/j.jml.2010.12.006.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage, 59*, 1859–1869. https://doi.org/10.1016/j.neuroimage.2011.08.057.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*, 155–200. https://doi.org/10.1023/A:1005184709695; https://doi.org/10.1023/a:1005140927442.
McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*, 155–200. https://doi.org/10.1023/A:1005184709695; https://doi.org/10.1023/a:1005140927442.
McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*, 67–91. https://doi.org/10.1016/S0749-596X(02)005

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. Duncan Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 419–491). New York: Wiley.

Miller, G. A., & Isard, S. D. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior, 2*, 217–228. https://doi.org/10.1016/S0022-5371(63)80087-0.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251–269. https://doi.org/10.3758/BF03213879.

Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language, 99*, 1–34. https://doi.org/10.1016/j.jml.2017.08.004.

Nicenboim, B., Logacev, P., Gattei, C., & Vasishth, S. (2016). When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in Psychology, 7*, 1–24. https://doi.org/10.3389/fpsyg.2016.00280.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*, 5–19. https://doi.org/10.1007/BF01069043.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867–872. https://doi.org/10.1016/j.jesp.2009.03.009.

Parker, D. (2019). Cue Combinatorics in Memory Retrieval for Anaphora. *Cognitive Science, 43*, 1-30. https://doi.org/10.1111/cogs.12715.

Parker, D., & Lantz, D. (2017). Encoding and accessing linguistic representations in a dynamically structured holographic memory system. *Topics in Cognitive Science, 9*, 51–68. https://doi.org/10.1111/tops.12246.

Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition, 157*, 321–339. https://doi.org/10.1016/j.cognition.2016.08.016.

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language, 94*, 272–290. https://doi.org/10.1016/j.jml.2017.01.002.

Patil, U., Vasishth, S., & Lewis, R. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology, 7*, 1–18. https://doi.org/10.3389/fpsyg.2016.00329.

R Development Core Team. (2018). R: *A language and environment for statistical computing.* R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134. https://doi.org/10.1037/0033-295X.88.2.93.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510. https://doi.org/10.1037/0033-2909.114.3.510.

Runner, J., Sussman, R. S., & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrases. *Cognitive Science, 30*, 193–241. https://doi.org/10.1207/s15516709cog0000_58.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*, 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3.

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language, 76*, 195–215. https://doi.org/10.1016/j.jml.2014.07.003.

Tucker, M. A., Idrissi, A., & Almeida, D. (2015). Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in Psychology, 6*, 1–21. https://doi.org/10.3389/fpsyg.2015.00347.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 407–430. https://doi. org/10.1037/0278-7393.33.2.407.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49, 285–316*. https://doi.org/10.1016/S0749-596X(03)00081-0.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. Journal of Memory and Language, 55, 157–166. https://doi.org/10.1016/j.jml.2006.03.007.

Van Dyke, J. A., & McElree, B. (2007). Similarity-based proactive and retroactive interference reduces quality of linguistic representations. Poster presented at the CUNY Conference on Human Sentence Processing. San Diego, CA.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65*, 247–263. https://doi.org/10.1016/j.jml.2011.05.002.

Vasishth, S., & Nicenboim, B. (2016). Statistical Methods for Linguistic Research: Foundational Ideas - Part 1. *Language and Linguistics Compass, 10*, 349–369.

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science, 32*, 685–712. https://doi.org/10.1080/03640210802066865.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*, 206–237.

Wundt, W. (1904). The psychology of the sentence. Reprinted in A. L. Blumenthal (Ed.). (1970). *Language and psychology: Historical aspects of psycholinguistics.* (pp. 9–32). New York: Wiley.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language, 108*, 40–55.

Xiang, M., Grove, J., & Giannakidou, A. (2013). Dependency dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology, 4*, 1–19.