



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Cognitive Psychology

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

Encoding interference effects support self-organized sentence processing

Garrett Smith^{a,*}, Julie Franck^b, Whitney Tabor^{a,c}

^a University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269, USA

^b Université de Genève, Uni-Pignon bureau 610, Boulevard du Pont d'Arve 42, 1205 Geneva, Switzerland

^c Haskins Laboratories, 300 George Street, New Haven, CT 06511, USA

ARTICLE INFO

Keywords:

Sentence comprehension
Encoding interference
Semantic similarity
Agreement attraction
Dynamical systems models
Self-organized sentence processing

ABSTRACT

According to cue-based retrieval theories of sentence comprehension, establishing the syntactic dependency between a verb and the grammatical subject is susceptible to interference from other noun phrases in the sentence. At the verb, the subject must be retrieved from memory, but non-subject nouns that are similar on dimensions that are relevant to subject-verb agreement, like number marking, can make the retrieval more difficult. However, cue-based retrieval models fail to account for a class of interference effects, conventionally called “encoding interference,” that cannot be due to retrieval interference. In this paper, we implement a self-organized sentence processing model that provides a more parsimonious explanation of encoding interference effects than otherwise reasonable extensions that could be made to the cue-based retrieval approach. We first also present new behavioral evidence for encoding interference using a semantic similarity manipulation in two self-paced reading studies of subject-verb number agreement. The results of these experiments are more compatible with the self-organizing account. We argue that self-organization, which reduces all parsing to fallible feature match optimization and makes no a priori distinction between encoding and retrieval, can provide a unifying approach to similarity-based interference in sentence comprehension.

1. Introduction

In sentence comprehension, the syntactic dependency between a verb and its grammatical subject has to be established in order to build an interpretable structure. Cue-based retrieval models of parsing hypothesize that the dependency formation process is susceptible to similarity-based interference. For example, *the key in the key to the cabinets is...* has a dependency with the verb *is*, but it is separated from it by the prepositional phrase *to the cabinets*. According to leading theories of cue-based retrieval in sentence processing, after reading *is*, the parser is hypothesized to retrieve a subject from memory on the basis of cues at the verb, such as +SINGULAR and +NOUN (Lewis & Vasishth, 2005; McElree, Foraker, & Dyer, 2003). When multiple noun phrases (NPs) in memory share retrieval features, the retrieval of the correct word can be delayed, or in some cases, the wrong word can be retrieved. An example of this comes

* Corresponding author at: University of Potsdam, Department of Linguistics, Haus 14, Karl-Liebknecht-Straße 24–25, 14476 Potsdam, Germany.
E-mail address: gasmith@uni-potsdam.de (G. Smith).

<https://doi.org/10.1016/j.cogpsych.2020.101356>

Received 26 November 2019; Received in revised form 31 August 2020; Accepted 14 October 2020

Available online 4 December 2020

0010-0285/© 2020 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

from studies of agreement processing in ungrammatical structures like (1).

- (1) a. The key to the cabinet are...
b. The key to the cabinets are...

Typically, participants read the verb *are* in (1-b) more quickly than in (1-a) (Dillon, Mishler, Sloggett, & Phillips, 2013; Jäger, Engelmann, & Vasishth, 2017; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Pearlmutter, Garnsey, & Bock, 1999; Wagers, Lau, & Phillips, 2009). The cue-based retrieval model of Lewis and Vasishth (2005), which is based on the Adaptive Control of Thought-Rational (ACT-R; Anderson, 1990; Anderson et al., 2004), can explain this effect: In ACT-R, words in these sentences effectively race to reach an activation threshold in order to be retrieved. When a word matches a retrieval cue, its activation is increased, causing it to reach the retrieval threshold more quickly. In (1-a), *key* receives a small activation boost because it matches the verb's +SUBJECT (but not the +PLURAL) cue. *Cabinet*, on the other hand, is neither the subject nor plural, so it receives no activation boost. Retrieval takes a relatively long time because neither noun's activation started near the threshold, and the one noun that is boosted by a feature match (*cabinets*) is only a partial feature match, so it is not boosted very much. For (1-b), *key* (with its +SUBJECT feature) and *cabinets* (with its +PLURAL feature) are both partial matches for the verb's cues, so both nouns get a boost. Both nouns are now roughly equally activated, so a noisy race process ensues as both nouns' activations approach the threshold for retrieval. As discussed below, this leads, on average, to faster processing and more incorrect retrievals than in (1-a). The similar feature match between the nouns thus leads to facilitatory interference compared to (1-a).

Cue-based retrieval provides an elegant explanation for similarity-based interference when items are similar with regard to retrieval cues. However, not all interference effects can be explained as retrieval interference because the items involved are similar on features that are not relevant for retrieval. For example, Gordon, Hendrick, and Johnson (2001) studied sentences like (2) in self-paced reading.

- (2) a. The banker that the barber praised climbed the mountain.
b. The banker that praised the barber climbed the mountain.
c. The banker that you praised climbed the mountain.
d. The banker that praised you climbed the mountain.

Gordon et al. (2001) found that the object-relative clause in (2-a) was more difficult to process than the subject-relative clause in (2-b), replicating previous findings (e.g. King & Just, 1991). However, the difference was attenuated in (2-c) and (2-d), which differed only in the type of noun phrase (NP) used in the embedded clause (definite description vs. pronoun).¹ Gordon, Hendrick, and Johnson (2004) and Gordon, Hendrick, Johnson, and Lee (2006) report similar findings. Definite descriptions and pronouns do not differ in their match to the verb's retrieval cues—verbs just need something nominal in the correct syntactic position—so this effect cannot be explained using cue-based retrieval. Instead, many researchers have hypothesized that similarity between two elements in a sentence can have an effect on their representations as soon as the second one is encoded, whether or not a later-arriving element prompts the retrieval of one of them. Effects that show this pattern have thus been called *encoding interference effects*. Encoding interference includes all interference that is caused by feature overlap between items where the features are not relevant for retrieval in the current context (see also Kush, Johns, & Van Dyke, 2015; Laurinavichyute, Jäger, Akinina, Roß, & Dragoy, 2017; Villata, Tabor, & Franck, 2018).²

Hofmeister (2011) and Hofmeister and Vasishth (2014) also found encoding interference in comprehension: They observed significant reading time slowdowns at the retrieval site in long-distance dependencies when the retrieval target was syntactically and semantically more similar to competitors. Semantic similarity can also affect agreement production. Barker, Nicol, and Garrett (2001) reported encoding interference in agreement attraction, with higher rates of incorrect plural verb number productions following subject noun phrases like (3-b), where both nouns have features related to being a boat, than for (3-a), where they do not.

- (3) a. The canoe by the cabins...
b. The canoe by the sailboats...

Villata et al. (2018) tested for encoding interference in subject-verb agreement in the comprehension of grammatical English and Italian sentences. In the Italian experiment, they used semantically reversible verbs (like *surprised*) that are not marked for gender and animate NPs that either matched or mismatched in gender, e.g., *The dancer-MASC that the waiter-MASC/waitress-FEM surprised drank a cocktail with alcohol*. The English materials used number (mis-) match instead of gender: *The dancer-SG/dancers-PL that the waiter-SG strongly criticized most of the time ordered a rum cocktail*. The past tense verb *criticized* is not overtly marked for number, so number

¹ One might worry that pronouns (but not nouns) are case-marked in English (making the pronouns distinctive on a retrieval cue), and so this effect might be explainable using retrieval cues like case. But Gordon et al. (2001)'s third experiment used proper names, which are not case-marked, instead of pronouns. This experiment produced similar results, ruling out this possible confound.

² In certain models of memory encoding and retrieval, encoding interference can affect retrieval. The models of Nairne (1990) and Oberauer and Kliegl (2006), for example, say that if two items share features, those features are effectively removed from consideration, making the items more difficult to distinguish from others in memory and therefore more difficult to retrieve. As Laurinavichyute et al. (2017) note, items can also become less distinct after they are encoded but before they are retrieved. This is difficult experimentally to distinguish from encoding interference, so we do not consider it separately here.

cannot be used as a cue that would differentiate between possible subjects. In both languages, they found evidence of interference in the comprehension questions—participants were significantly more likely to correctly answer the question in the feature mismatch conditions than in the match conditions. In the Italian experiment, there was also a weak effect in reading times at the verb (*surprised*), with faster reading times in the mismatch condition compared to the match condition. These results suggest that similarity-based interference from agreement features that are not retrieval cues has a clear effect in comprehension question accuracy, as well as a milder effect on online sentence processing in grammatical sentences. Laurinavichyute et al. (2017), Lewis and Vasishth (2005), Franck, Colonna, and Rizzi (2015), and Adani, Van der Lely, Forgiarini, and Guasti (2010, off-line, in children) report similar mismatch advantage effects in grammatical sentences, although such effects are typically much smaller and less reliable in grammatical sentences than in ungrammatical ones (which is typical for reading time effects in subject-verb agreement, see Jäger et al., 2017).

There is no current leading explanation for these effects in sentence processing, and the cue-based retrieval model of Lewis and Vasishth (2005), while able to capture many (but not all, see Jäger et al., 2017; Engelmann, Jäger, & Vasishth, 2019) retrieval interference effects, does not predict encoding interference effects. However, a simple extension of cue-based retrieval does allow that theory to account for these results.³ First, we can assume that semantically similar chunks in memory spread activation to each other in proportion to their own activation (as implemented in van Maanen & van Rijn, 2007; van Maanen, van Rijn, & Borst, 2009; van Maanen, van Rijn, & Taatgen, 2012, for example). This means that the activation of similar chunks will be higher than the activation of dissimilar chunks. Because this activation spreading is proportional to the activation of the sending chunk, the activations of two similar chunks will tend to be more equal, in addition to being higher, compared to the activations of two dissimilar chunks.⁴ When combined with noise in activation values, this predicts more incorrect retrievals, i.e., agreement attraction, in (3-b) compared to (3-a), just as Barker et al. (2001) found.

This extension also makes predictions for reading times at the verb, where the parser must retrieve a subject from memory. In ACT-R, the speed of retrieval—and therefore reading speed—is inversely related to the retrieved chunk's activation. The higher activations and the smaller difference in activations between semantically similar chunks predicts faster reading times via *statistical facilitation* (Raab, 1962) compared to sentences with semantically dissimilar chunks. Statistical facilitation is a phenomenon that occurs in race processes (Heathcote & Love, 2012; Logačev & Vasishth, 2015; Nicenboim & Vasishth, 2018; Rouder, Province, Morey, Gomez, & Heathcote, 2015; van Gompel, Pickering, & Traxler, 2000). To understand how it works, consider first a situation with two chunks, A and B, but where chunk A always wins. This means its activation was (much) higher than chunk B's, and the average retrieval time is just the average of all retrievals of A. Now, what would happen instead if chunk B's activation were approximately equal to chunk A's so that noise in the activations would allow B to win in half of the trials? This would mean that chunk B's activation was higher than chunk A's on each of those trials, and so the retrieval time is faster than if A had won. In this case, half of the trials would be faster than they would have been if chunk A had won every time. Note that this holds no matter which chunk we label A or B; if a chunk wins, it must have been faster than the other one, thus speeding up average finishing times across the board. Thus, over many trials, having chunks with closer activations must have a faster average retrieval time than when the chunks have activations that are further apart. This predicts that reading times will be faster on average when two nouns are semantically similar (contrary to what Villata et al., 2018, found).⁵

This spreading activation race model based on ACT-R can explain encoding interference effects like Barker et al. (2001)'s agreement production data, and it makes a new testable prediction for reading times. However, it requires positing an additional mechanism—proportional activation spreading between similar chunks—beyond what is needed simply to account for retrieval interference. In the next section, we present a different approach to sentence processing, self-organized sentence processing (SOSP; Smith & Tabor, 2018; Smith, Franck, & Tabor, 2018) that can also explain the Barker et al. (2001) results. As discussed below, SOSP takes a more parsimonious approach by treating encoding and retrieval as the product of an error-prone search for structures with good feature match. Importantly, the SOSP simulations below show that the two approaches make opposite predictions for reading times. We test the opposing predictions in Experiments 1 and 2.

2. A self-organizing approach to encoding interference

Self-organized sentence processing (SOSP; Smith, 2018; Smith & Tabor, 2018; Villata et al., 2018; Villata & Franck, 2019) is a dynamical systems approach to sentence processing. In contrast to other theories that apply symbolic rules sequentially (like ACT-R), SOSP builds on models that are defined by continuous-time differential equations (Cho & Smolensky, 2016; Cho, Goldrick, & Smolensky, 2017; Cho, Goldrick, Lewis, & Smolensky, 2018; Green & Mitchell, 2006; Kempen & Vosse, 1989; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Smith et al., 2018; Stevenson, 1994; Stevenson & Merlo, 1997; Tabor & Hutchins, 2004; Vosse & Kempen, 2000; Vosse & Kempen, 2009; van der Velde & de Kamps, 2006). Like these other dynamical approaches, parses in SOSP gradually emerge through local interactions between individual words, with incompatible alternative parses competing based on how well-formed the resulting structures would be. These models have provided plausible explanations for a variety of important sentence processing effects, including the increased difficulty of center embedded vs. right branching structures, lexical ambiguity effects (Vosse & Kempen, 2000), predictive processing (Vosse & Kempen, 2009), length effects in garden paths (Tabor & Hutchins, 2004), local

³ Villata et al. (2018) discuss a different extension of ACT-R that can also account for this effect. We consider this in the general discussion.

⁴ See Supplement 1 for simple proofs of the convergence in activations for two possible implementations of the race idea.

⁵ See footnote four of Nicenboim and Vasishth (2018) for an easy-to-use Shiny implementation (Chang, Cheng, Allaire, Xie, & McPherson, 2019) of a race model for sentence processing that can be used to illustrate this point.

coherence effects (Smith & Tabor, 2018), and surprisal effects (Cho et al., 2018). Recently, Villata et al. (2018) argued that this approach can also explain encoding interference. One contribution of this paper is to implement an SOSP model of encoding interference to make that argument concrete.

The basic unit of linguistic structure in SOSP is the syntactic treelet. These are lexically anchored chunks of syntactic and semantic information that determine not only the properties of a word but also how it can interact with other words. Each treelet has a head attachment site with syntactic (e.g., \pm PLURAL, \pm FEMININE) and semantic features (e.g., \pm ANIMATE, \pm BOAT) that reflect the properties of that word. Treelets can also have zero or more dependent attachment sites with features that reflect the word's selectional preferences for each of its dependents.

Treelets assemble into larger structures by forming links between attachment sites. For simplicity, we assume here that the types of dependent attachment sites and the structures they create are governed by a simple dependency grammar (Hays, 1964; Gaifman, 1965). Attachment links have a continuous-valued strength that grows according to how well the features match at each end of the link. Links can only form between head and dependent attachment sites on separate treelets, but there are no other restrictions placed on which links can exist. This means that the SOSP parser can build attachment configurations that would be banned under a strict application of a symbolic grammar, similar to some previous dynamical parsing proposals (e.g., Kempen & Vosse, 1989). For example, if given the sequence of words *the red ball*, an SOSP parser could build a structure in which the adjective *red* attaches as the determiner of *ball*. However, the feature match for such a configuration is low, and so it would almost always be out-competed by structures with a better feature match (e.g., *the* attaching as the determiner of *ball*). It is this property of SOSP that allows it to naturally explain local coherence effects (Konieczny, 2005; Paape & Vasishth, 2015; Tabor, Galantucci, & Richardson, 2004) as situations in which ungrammatical structures compete with grammatical ones.

A surprising shortcoming of previous self-organizing models is that they generally fare poorly in predicting processing times (although Cho et al., 2018; Tabor & Hutchins, 2004, are exceptions to that pattern). For example, Vosse and Kempen (2000)'s model predicts when parsing is likely to fail, but they do not report any systematic relationship between the likelihood of parse success or failure and reading times. Processing times are a major source of data in psycholinguistics (e.g., self-paced reading and eye-tracking while reading), so it is surprising that so many self-organizing models, which give time a central role in processing, do so poorly in this regard. The SOSP model that we report was designed to make quantitative processing time predictions in a transparent way. To see this, we briefly introduce the framework and show how the equations that govern sentence processing naturally predict processing time effects.

2.1. The SOSP framework

All structure building in SOSP is driven by feature match between dependency-linked attachment sites. This affects how strongly words interact, which structures are built, and how quickly processing proceeds. As shown below, SOSP transparently predicts that well-formed structures—those with good feature match—are built faster and are more likely to be built than ill-formed structures.

Feature match is formalized in SOSP according to Eq. (1) (Smith & Tabor, 2018):

$$h_i = \prod_{l \in \text{links}} \left(1 - \frac{\text{dist}_{\text{Hamming}}(\mathbf{f}_{l,\text{head}}, \mathbf{f}_{l,\text{dependent}})}{n_{\text{feat}}} \right) \quad (1)$$

For each configuration of links and features i , the harmony h_i (a continuous measure of well-formedness; Smolensky, 1986) is given by the product across all of the links l in the structure of 1 minus the Hamming distance (number of mismatched binary features) between the feature vectors on the head ($\mathbf{f}_{l,\text{head}}$) and dependent ($\mathbf{f}_{l,\text{dependent}}$) attachment sites, normalized by the number of features n_{feat} . Thus, the harmony ranges between 0 (very ill-formed) to 1 (perfectly grammatical). This equation is used to calculate the harmony of all possible well- and ill-formed symbolic structures—that is to say all the configurations in which features and links are either fully activated (strength of 1) or fully turned off (strength of 0). These states of the system easily map to symbolic structures. However, the strengths of the features and links are allowed to range freely in \mathbb{R}^m , where m is the number of dimensions of the system.

To calculate the harmony of states intermediate between these points, we interpolate using radial basis functions, similar to some pattern-completion neural networks (Ciocoitu, 1996; Ciocoitu, 2009; Han, Sayeh, & Zhang, 1989; Muezzinoglu & Zurada, 2006):

$$\phi_i(\mathbf{x}) = \exp\left(-\frac{(\mathbf{x} - \mathbf{c}_i)^T(\mathbf{x} - \mathbf{c}_i)}{\gamma}\right) \quad (2)$$

Each link and feature configuration i has its own radial basis function defined by ϕ_i . These Gaussian-shaped⁶ functions create a harmony peak at the location \mathbf{c}_i , where the links and features are either fully on or fully off. The width of these peaks is determined by γ , which is fixed for all i . The ϕ_i make it so that harmony generally decreases if the system's state \mathbf{x} moves away from the \mathbf{c}_i . Each ϕ_i represents a single structure, so to encode multiple link and feature configurations in memory, we take a weighted sum of all ϕ_i , creating a (high-dimensional) hilly harmony landscape (see Fig. 2 for a two-dimensional example):

⁶ Although the shape of these radial basis functions is Gaussian, they do not have a direct probabilistic interpretation.

$$H(\mathbf{x}) = \sum_i^n h_i \phi_i(\mathbf{x}) \quad (3)$$

The height of each harmony peak i near a structure \mathbf{c}_i is given by its harmony h_i , with more well-formed structures having higher hilltops. The hilly harmony landscape defined in Eq. 3 represents a competent language user's knowledge of which link and feature configurations are possible and how well formed they are.

To do incremental language comprehension in this system, we hypothesize that an input word sets the system's state \mathbf{x} by turning on features of the perceived word. From this point, the system tries to locally maximize the harmony of its state by noisily following the gradient of the harmony landscape uphill. Formally, the state \mathbf{x} changes in time via the following stochastic differential equation:

$$\begin{aligned} d\mathbf{x} &= \nabla_{\mathbf{x}} H(\mathbf{x}) dt + \sqrt{2D} d\mathbf{W} \\ &= -\frac{2}{\gamma} \sum_i^n h_i (\mathbf{x} - \mathbf{c}_i) \phi_i(\mathbf{x}) dt + \sqrt{2D} d\mathbf{W}, \end{aligned} \quad (4)$$

where $\nabla_{\mathbf{x}} H$ is the gradient of the harmony function H . The magnitude of the Gaussian noise process $d\mathbf{W}$ is given by D . Once the system is sufficiently close to one of the \mathbf{c}_i where links and features are fully active or inactive, the system inputs the next word, displacing its state (but leaving currently active features and links untouched) and starting the process anew. From the new point, the system follows the gradient uphill toward another \mathbf{c}_i . Thus, parsing is accomplished via noisy gradient ascent in the hilly harmony landscape with jumps determined by the features of words in the input. Formally, this makes SOSP a stochastic hybrid dynamical system (di Bernardo, Budd, Champneys, & Kowalczyk, 2008; Goebel, Sanfelice, & Teel, 2009). The attractors of the system dynamics are at the \mathbf{c}_i as long as certain conditions on γ , the h_i , and the noise magnitude D hold (Han et al., 1989; Smith, 2018).

With these equations, we can make predictions about processing times as soon as we know the harmonies h_i of link and feature configurations (Smith & Tabor, 2018). We illustrate this by first considering a situation in which only a single link and feature configuration \mathbf{c} can be built, meaning there is only a single harmony hill. In this case, Eq. (4) simplifies to

$$d\mathbf{x} = -\frac{2h}{\gamma} (\mathbf{x} - \mathbf{c}) \phi(\mathbf{x}) dt + \sqrt{2D} d\mathbf{W} \quad (5)$$

This shows that the rate of change in the state variables \mathbf{x} , i.e., the speed of processing, is proportional to the harmony h of the peak. Thus, a well-formed structure will be faster to build than an ill-formed one.

Now consider a system with two or more competing peaks. If the system starts between two peaks, it begins moving depending on the shape of the harmony landscape and the noise. Eventually, the system will get sufficiently close⁷ to one peak, and the time it takes to reach the top will still be proportional to its local harmony, as the influence of other peaks drops away exponentially.⁸ Thus, over many trials, the average processing time will be proportional to the harmonies of the chosen peaks weighted by how often they were chosen. This is a main source of processing time differences between experimental conditions in SOSP: Different experimental conditions correspond to different harmony landscapes, which can produce different average processing times if the harmonies of the peaks involved in each condition differ. Another cause of processing time slowdowns is this: When the system is between two peaks, it can sometimes move slowly along the boundary separating the basins of attraction for each point, called the *separatrix*. The gradient of the global harmony function is small in this area, so the system moves slowly while it is hamstrung between different structures. Thus, the shape of the harmony function is what determines processing times via the relative harmony heights of competing structures and the flattening of the landscape between them. We now apply this approach to the encoding interference materials introduced above.

2.2. An SOSP model of encoding interference

We now present an implemented SOSP parser to illustrate how the framework functions and to derive reading time predictions for sentences like examples (3-a) and (3-b). We focus on processing times at the verb, because this is where the spreading activation race model and SOSP make diverging predictions. At the verb, a subject has to be retrieved (in the spreading activation race model) or one of the nouns has to attach as the subject of the verb (in SOSP). Therefore, we focus on simulating SOSP's settling times after it has already processed *the canoe by the cabins/kayaks* and input the verb. Taking a snapshot in the middle of word-by-word processing like this facilitates understanding and visualizing the dynamics.

Our SOSP model of encoding interference included two competing structures (Figs. 1 and 2), the harmonies of which were varied between experimental conditions. One structure was the fully grammatical N1-to-verb structure—where the first noun *canoe* correctly attaches as the subject of the verb *was*, see Fig. 1A—and the other an imperfect N2-to-verb structure—where the second noun *cabin(s)/kayak(s)* attaches as the subject (Fig. 1B). To keep the simulations simple, we only simulated the strengths of the N1-to-verb and N2-to-

⁷ We measure proximity to a peak using the Chebyshev distance: $d(\mathbf{x}, \mathbf{c}_i) = \max_j (|x_j - c_{i,j}|)$. Once the system was within a distance of 0.1, parsing stopped, and the structure the system reached was recorded as being the structure built.

⁸ Note that moving towards one parse necessarily means moving away from other, competing parses, making SOSP a competition based model (McRae et al., 1998; MacDonald, Pearlmutter, & Seidenberg, 1994). This contrasts with the spreading activation race model, where an increase in the activation of one structure has no effect on the activation of other structures.

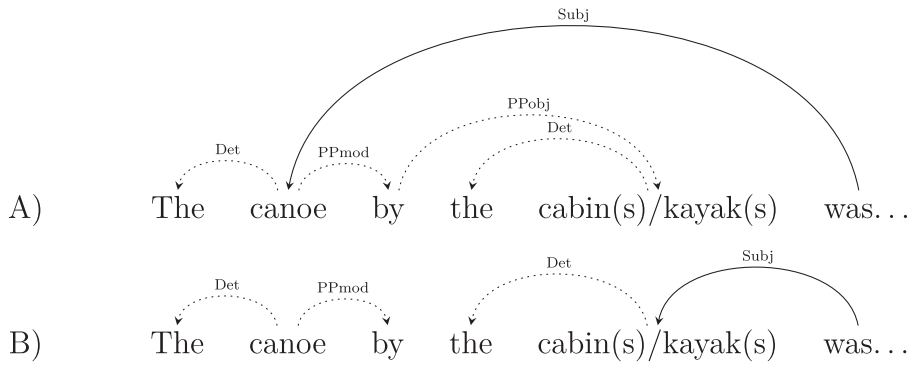


Fig. 1. Dependency structures for the encoding interference materials. The arrows point from the head/governor to its dependent. Solid lines show the competing dependency structures in the implemented SOSF model. The dotted lines show other dependencies in the subject NP that were not simulated. (A) shows the fully grammatical ($h_0 = 1.0$) structure, and (B) shows the implemented imperfect ($h_1 < 1.0$) way of attaching the second noun as the verb’s subject. Note that the N1 substructure and the N2/verb substructure in (B) are not linked to form a single structure. This configuration is assumed to be somewhat ill-formed, and its harmony is penalized (see main text for details). Abbreviations: Subj = subject, Det = determiner, PPmod = prepositional phrase modifier, PPobj = object of a preposition.

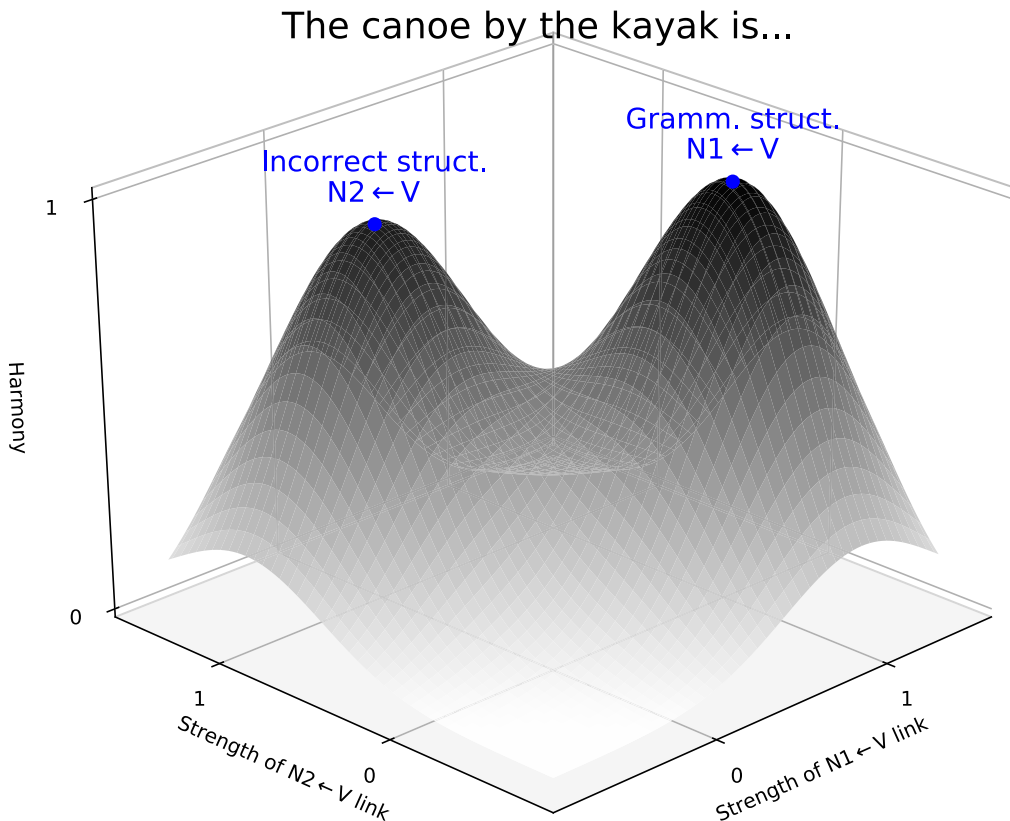


Fig. 2. Example harmony landscape for the implemented SOSF model. Axes in the horizontal plane code the strengths of the two simulated attachment links. Vertical axis is the harmony of the structures. The left hill corresponds to the imperfect structure in which the second noun, *kayak* in this case, attaches as the subject dependent of the verb ($N2 \leftarrow V$). The structure on the right corresponds to the full-harmony structure in which *canoe* attaches as the verb’s subject ($N1 \leftarrow V$). The harmony heights are given in Table 2.

verb attachment links (shown in solid lines in Fig. 1) and assumed that other attachment links would form consistently between conditions. This setup models the establishment of an attachment link between the verb’s subject attachment site and one of the nouns’ head attachment sites after reading the verb. A sample harmony landscape is shown in Fig. 2.

To calculate the local harmonies (h_i), we assumed there were three semantic features coding properties of being a boat and one number feature (see Table 1). This setup reflects our assumption that semantic representations of words are rich and high-dimensional,

Table 1
Feature values for words used in the simulated materials.

Word	Boat-1	Boat-2	Boat-3	Plural
<i>canoe</i>	1	1	1	0
<i>cabin</i>	0	0	0	0
<i>cabins</i>	0	0	0	1
<i>kayak</i>	1	1	1	0
<i>kayaks</i>	1	1	1	1
<i>is</i>	1	1	1	0
<i>are</i>	1	1	1	1

while syntactic representations are low-dimensional. It also puts the model in the range of parameters that Smith and Tabor (2018) show should lead to competition-based slowdowns. We also assumed that after reading *canoe*, the parser expects a verb that takes a subject with boat features. So, the verb was assumed to always have all three boat features active. The h_i were calculated using Eq. (1)⁹ and we assumed that all links other than the ones between the two nouns and the verb were perfect feature matches, so only feature mismatches on those links contributed to lower harmony. The final assumption going into the model was that linking the N2 to the verb leaves the N1 without a head/governor, which was penalized by multiplying its h_i by a free parameter set to 0.8, which provided reasonable results. The harmonies of the attractors are shown in Table 2. The initial condition had both link strengths set to zero, i.e., equidistant from each attractor. Two additional free parameters, the width parameter γ and the noise magnitude D , were set to 0.4 and 0.001, respectively. The model was run 2000 times in each of the eight conditions created by crossing the factors of semantic similarity between N1 and N2 (similar or dissimilar), N2 number (singular or plural), and verb number (singular or plural). Note that the first noun (N1) is always singular, so conditions with a plural verb correspond to ungrammatical sentences in the human experiments below. Code and data from these simulations are available on the OSF repository for this paper: <https://osf.io/hjrkkn/>.

2.3. Results and discussion

Table 2 shows the percentage of times the model settled into each attractor in each condition. In the semantically dissimilar conditions, the system always attached the N1 as the subject of the verb. In the semantically similar conditions, the N2 was able to attach as the subject, as the N2-verb attractors had higher harmonies in these conditions. The N2-verb attachment in these conditions never mismatched on the boat features, so it could only be penalized if the number markings on the N2 and the verb differed. The N2-verb structure was built most frequently in the semantically similar, plural-N2, plural-verb condition. This is because this structure has higher harmony than the N1-verb structure, which mismatches on the boat features.¹⁰

More important for the comparison with the predictions of the spreading activation race model is the predicted settling times at the verb, which are shown in Fig. 3 with error bars ranging from the first quartile to the third. A few patterns are worth noting, starting first with the grammatical conditions (left facet). First, settling times are longer in the semantically similar conditions than in the dissimilar conditions. This is because the N2 is a good feature match for the pre-activated boat features on the verb. The N2-verb structure has lower harmony than the N1-verb structure, though (except for the semantically dissimilar, plural N2, plural verb condition), so when it is built, it takes longer than the correct structure. The relatively high harmony of the N2-verb structure also flattens the harmony landscape between the structures somewhat, decreasing the magnitude of the gradient and slowing processing. Importantly, these slowdowns are the opposite of what the spreading activation race model predicts.

In addition, the grammatical singular N2 conditions settle more slowly than the plural N2 conditions. The singular N2 is a relatively good feature match for the verb's singular feature, slowing processing as it competes with the N1. This is similar to what ACT-R predicts (in cases of retrieval interference only; Jäger et al., 2017; Nicenboim, Vasishth, Engelmann, & Suckow, 2018) via the fan effect, where activation is distributed between all memory chunks that match retrieval cues. Finally, the effect of this competition is larger in the semantically similar conditions with singular verbs than in the dissimilar conditions. This is because *canoe* and *kayak* both have a good semantic feature match with the verb, so the competition is especially strong when both also match the verb in number.

The picture in the ungrammatical sentences (right facet) is somewhat different, although the mechanisms at work are the same. First, processing times overall are slower. There is also still a large effect of semantic similarity, with slower processing for semantically similar N2s. The effect of N2 number is flipped, though. This makes sense, as now a plural N2 has a good feature match with the verb, which makes competition with the N1 (which mismatches on number but matches on boat-features) stronger. Finally, there is again a

⁹ With four total features, missing on one feature decrements the harmony by 0.25. However, we felt it was important that even very bad structures should affect processing, as assumed in the general theory of SOSP. To this end, we decremented the harmony by 0.24 (instead of 0.25) for each missed feature so that no structure had a harmony of 0.0. Numerical optimization on the harmony surface showed that this did not create a peak for the low-harmony structure, but it allows the structure to have some minimal influence on processing.

¹⁰ These simulations model attaching one of the nouns as the subject of a verb provided to a participant, so they do not pertain directly to agreement production. However, we could set up a model in which the system is provided with N1 and N2 features and has to settle on either a singular or plural verb. The harmony landscapes would be very similar to the ones used in the present simulations, and pilot simulations suggest such a model reproduces the Barker et al. (2001) results quite well, i.e., more agreement attraction for *the canoe by the kayaks* than for *the canoe by the cabins*.

Table 2

Harmonies of the N1-verb and N2-verb structures and percentages of runs that settled to the two attractors. Recall that the N1 was always singular, so any configuration with a plural verb is ungrammatical.

Semantic similarity	Condition		Harmony		Percent built	
	N2 number	Verb number	N1-V	N2-V	N1-V	N2-V
Dissimilar	Singular	Singular	1.000	0.224	100%	0%
Dissimilar	Singular	Plural	0.760	0.032	100%	0%
Dissimilar	Plural	Singular	1.000	0.032	100%	0%
Dissimilar	Plural	Plural	0.760	0.224	100%	0%
Similar	Singular	Singular	1.000	0.800	98.5%	1.5%
Similar	Singular	Plural	0.760	0.608	97.15%	2.85%
Similar	Plural	Singular	1.000	0.608	100%	0%
Similar	Plural	Plural	0.760	0.800	31.05%	68.95%

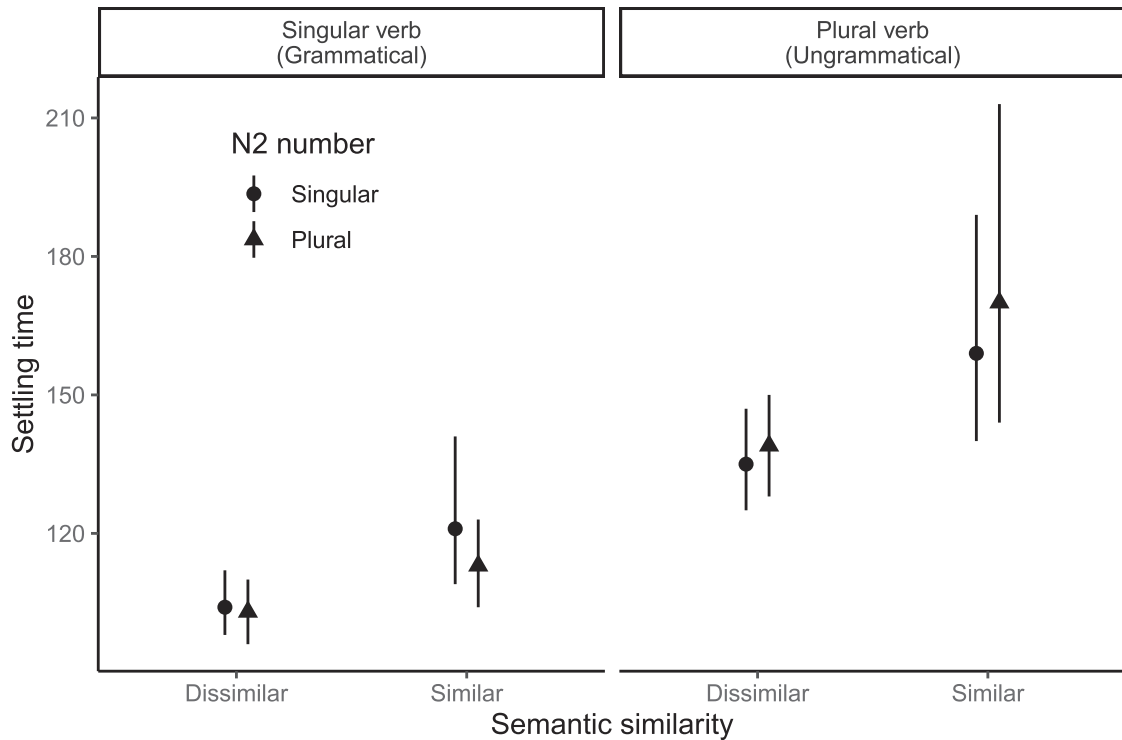


Fig. 3. Median settling times (arbitrary time units) for the SOSP model of reading times at the verb. Error bars range from the first to the third quartiles of the settling time distribution. The left facet shows the grammatical singular verb conditions, and the right facet the ungrammatical plural verb conditions.

larger effect of N2 number in the semantically similar conditions than in the dissimilar conditions for the same reason as in the singular verb case: The better feature match of the semantically similar plural N2 exacerbates the competition with the N1 more than with the semantically dissimilar plural N2.

To summarize, the SOSP model makes the following predictions about reading times at the verb in the self-paced reading experiments below. First, semantic similarity and verb number are predicted to have effects such that the semantically similar conditions and the conditions where the N2 matches the verb number are read more slowly. This is consistent with previous encoding interference results, e.g., Villata et al. (2018), but it is the opposite of what the spreading activation race model predicts. The SOSP model also predicts interactions between semantic similarity, N2 number and verb number. Processing times should be slowed when the verb and the N2 match in number, but this interacts with semantic similarity and verb number, with especially slow processing when the N2 is semantically similar to the N1 and the sentence has an ungrammatical plural verb. It is not clear whether the spreading activation race model makes any interactive predictions here. The lack of clear interactive predictions of the spreading activation race model, combined with the fact that sentence processing studies often involve high within- and between-participant variability (Jäger et al., 2017), we emphasize the differing main effect predictions of the models. While we did gather a larger-than-average sample size for both experiments (approximately 100 participants each), detecting three-way interactions is difficult with real data. Nonetheless, Experiments 1 and 2 were designed to test the full range of diverging predictions of SOSP and the spreading activation race model (with

regard to the effect of semantic similarity) and to test for evidence of the interactions that SOSP predicts.

3. Experiment 1

The effect we are most interested in is the effect of semantic similarity at the verb, as this is the locus of divergence between SOSP and the spreading activation race model. SOSP predicts effects here because the N1 and N2 compete to attach to the verb, with differences between conditions affecting the harmony landscape and causing differences in processing times. Specifically, reading times should be slower when the N1 and N2 are semantically similar compared to when they are dissimilar. In the spreading activation race model, reading the verb triggers a search for a subject in order to complete the subject-verb dependency. The different conditions change features on the nouns, affecting activation spreading and thus retrieval times. In contrast to SOSP, this model predicts speedups due to statistical facilitation when the nouns are similar. Because the full design involves attempting to estimate a three-way interaction from human reading time data (which are known to be noisy; Jäger et al., 2017, Appendix B) and including ungrammatical sentences as test stimuli can sometimes affect how participants approach the experiment (Franck et al., 2015), we decided to run an experiment with only grammatical items (i.e., with singular verbs) first. This should reduce variability due to shifting participant strategies and tests whether the human data follow the predictions of SOSP or the spreading activation race model more closely.

3.1. Method

3.1.1. Participants

One hundred and twenty-two participants were recruited from the University of Connecticut Department of Psychological Sciences participant pool; they took part for course credit. Two were removed for reporting a diagnosis of a speech or language problem, and ten were removed for failing to cooperate (e.g., admitting to “blowing through” the experiment to finish quickly) or errors in the experiment software. Thus, data from 110 participants were included in the analyses.

3.1.2. Materials

Thirty-six test sentences were constructed. Each sentence consisted of a subject NP of the form *the N1 Prep the N2*, followed by an adverb modifying the proposition (S-adverbs, Potsdam, 1998), *was*, an adjective or past participle, and finally a prepositional phrase. Only grammatical sentences were used in this experiment (see (4)). Following Barker et al. (2001), we manipulated the semantic similarity of the N2 to the N1 and the number marking on the N2 (singular or plural; N2 number). The semantically similar N2s were chosen to be similar in size, shape, and function to the N1, while the semantically dissimilar N2s were chosen simply to fit into a plausible situation with the N1 without being similar on those criteria. The N1 and N2 were both inanimate. We reasoned that a content verb (like *sailed*) might bias which noun participants attach as the verb or introduce retrieval cues that might favor one noun over the other; indeed, Thornton and MacDonald (2003) found that similarly plausible N1s and N2s caused a slowdown in reading times at the content verbs they used. Such an effect can be explained via retrieval interference, so we used the past tense form of *be* followed by a past participle, reasoning that the auxiliary verb would not induce a memory search where the similarity between the N1 and N2 would be irrelevant. The adverb was included owing to Wagers et al. (2009)’s finding that a plural-marked N2 causes slower reading times in a spillover region than a singular noun. The inclusion of the adverb was meant to catch any such spillover from the plural N2s that might obscure other effects at the verb. Sample items are given in (4).

- (4)
- a. *Dissimilar, N2 singular:*
The canoe by the cabin likely was damaged in the heavy storm.
 - b. *Dissimilar, N2 plural:*
The canoe by the cabins likely was damaged in the heavy storm.
 - c. *Similar, N2 singular:*
The canoe by the kayak likely was damaged in the heavy storm.
 - d. *Similar, N2 plural:*
The canoe by the kayaks likely was damaged in the heavy storm.

After each sentence, participants answered a two-alternative forced choice comprehension question with the choices presented below the question. The comprehension questions for the test sentences all asked what the subject was, in effect. The correct answer was always the N1 for test items (see (5)). This type of comprehension question allows us to test whether participants were more confused about the subject when the N1 and N2 were similar, providing at least suggestive evidence about which noun might have attached as the subject online. The questions were kept as short as possible while still making sense in the context of the test sentence.

- (5)
- What was damaged?
1. the canoe
2. the cabin

The test sentences were interleaved with 108 filler sentences (a three-to-one filler-to-test sentence ratio). Comprehension questions for the fillers were designed so that there was an equal number of questions for which the first noun and the second noun was correct across the whole experiment. All sentences were distributed in a latin square design with four lists. All items are provided in the [supplementary materials as Supplement 2](#).

3.1.3. Procedure

After giving informed consent, participants were seated at a computer in a private booth. Participants read the instructions, which, after describing how the sentences and comprehension questions would be presented, told them to read at their normal, natural pace. Participants were instructed to answer the comprehension questions as quickly and accurately as possible based on what they had read in the preceding sentence. Four practice sentences were provided for participants to get acquainted with the paradigm. The experimenter remained in the testing room during the practice sentences to answer any questions and then left when the actual experimental sentences began.

We used a non-cumulative, moving-window self-paced reading paradigm (Just, Carpenter, & Woolley, 1982) as implemented on the experiment presentation platform IbexFarm (created by Alex Drummond; <http://spellout.net/ibexfarm/>). The letters in each word of a sentence were replaced with underscores until a word was revealed. After the word was revealed and read, its letters were replaced again with underscores. Participants advanced to each successive word by pressing the spacebar, and they answered the comprehension questions using the “1” or “2” buttons on the keyboard. The answers were presented on separate lines (as in (5)), with the order of the answers randomized for each sentence.

3.1.4. Analyses

We analyzed the comprehension question accuracy, comprehension question response times, and the word-by-word self-paced reading times using Bayesian (generalized) linear mixed effects models using the *brms* package in R (Bürkner, 2017). Bayesian analyses allow us to estimate the posterior probability distribution of effects of interest, often succeeding in fitting complicated models where frequentist mixed effects models would fail to converge (see, e.g., Vasishth, Nicenboim, Beckman, Li, & Kong, 2018). Here, we report the mean of the posterior along with 95% credible intervals (CrI), that is, the range of estimated parameter values that contains the mean with 95% probability (given the data and the statistical model). As Cumming (2014) and Kruschke and Liddell (2017) argue, reporting effect sizes and uncertainty intervals is both more informative than simply reporting the outcome of a significance test and also avoids the temptation to think of results as important if they are significant. It instead puts the focus on determining the size and direction of an effect, along with its uncertainty. This also facilitates future meta-analyses and a more cumulative approach to psycholinguistics. Code and data for this experiment and the next are available on the OSF repository for this paper: <https://osf.io/hjrk/>. In addition, because the crucial predictions of the activation-based race model and SOSP differ in sign, we also report the proportion of posterior estimates that lies below zero, denoted $p(\beta < 0)$ (Makowski, Ben-Shachar, Chen, & Lüdtke, 2019). This quantifies how probable it is that an effect goes in a particular direction.

Bayesian analyses require the researcher to choose prior distributions for the parameters of interest. Here, we chose mildly informative priors that reflect domain expertise—e.g., self-paced reading times in English are typically around 400 ms—while not artificially constraining the posterior (Schad, Betancourt, & Vasishth, 2019). The priors are listed in the Appendix.

The fixed effects for all analyses included semantic similarity, N2 number, and their interaction. Factors were coded using deviation coding, with plural and semantically dissimilar coded as +0.5 and singular and similar coded as -0.5. We used the full random effects structure justified by the design (Barr, Levy, Scheepers, & Tily, 2013). Any question response times or reading times less than 50ms or greater than 10s were excluded from analysis (affecting about 1% of the data).

3.2. Results

3.2.1. Comprehension question accuracy

The raw accuracies are shown in Table 3. Participants were more likely to answer the question correctly in the semantically dissimilar condition than in the semantically similar condition ($b = 0.357$, 95% CrI [0.053, 0.639], $p(\beta < 0) = 0.013$). The effect of N2 number was centered on zero ($b = -0.084$, 95% CrI [-0.325, 0.173], $p(\beta < 0) = 0.754$). There was an interaction such that the difference in accuracy between singular and plural N2s was smaller when the nouns were semantically similar ($b = -0.402$, 95% CrI [-0.869, 0.068], $p(\beta < 0) = 0.954$).

3.2.2. Comprehension question response times

Only correct responses were analyzed, and response times were log-transformed before analysis. Participants answered questions more quickly in the semantically dissimilar condition than the semantically similar condition ($b = -0.039$, 95% CrI [-0.070, -0.008], $p(\beta < 0) = 0.992$). Participants were also faster in the plural N2 condition than in the singular N2 condition ($b = -0.021$, 95% CrI [-0.044, 0.002], $p(\beta < 0) = 0.962$). The interaction showed a larger difference in response times between singular and plural N2s for the semantically similar conditions than for the dissimilar conditions ($b = 0.039$, 95% CrI [-0.011, 0.090], $p(\beta < 0) = 0.059$).

3.2.3. Reading times

The analysis of reading times proceeded as follows. First, after removing trials in which participants answered the comprehension question incorrectly, very long (>10 s) and very short (<50 ms) reading times were removed (about 11% combined). Then, reading times were log-transformed; results are reported on this scale. The raw reading times for the verb region, along with the rescaled SOSP

Table 3
Raw percent accuracy in each condition in Experiment 1.

	Singular N2	Plural N2
Semantically dissimilar	91%	87%
Semantically similar	84%	85%

model predictions,¹¹ are shown in Fig. 4.

Verb The clearest result at the verb was that participants read the verb faster in the semantically dissimilar condition than in the similar condition ($b = -0.029$, 95% CrI $[-0.057, -0.002]$, $p(\beta < 0) = 0.984$). There was weak evidence of slower reading for plural N2s than singular N2s ($b = 0.017$, 95% CrI $[-0.009, 0.043]$, $p(\beta < 0) = 0.104$) and still weaker evidence for the interaction ($b = -0.026$, 95% CrI $[-0.088, 0.034]$, $p(\beta < 0) = 0.800$).

Spillover To test for spillover processing from the verb, we also analyzed reading times at the word immediately following the verb (a past participle like *damaged* in our materials). There was no evidence for effects of semantic similarity ($b = -0.003$, 95% CrI $[-0.029, 0.023]$, $p(\beta < 0) = 0.411$) or N2 number ($b = 0.004$, 95% CrI $[-0.025, 0.034]$, $p(\beta < 0) = 0.580$). But there was an interaction ($b = 0.058$, 95% CrI $[0.004, 0.113]$, $p(\beta < 0) = 0.018$): Reading times for plural N2s were longer than singular N2s in the semantically dissimilar condition, but the opposite pattern held in the semantically similar condition.

3.3. Discussion

This experiment was designed to test the diverging predictions of the spreading activation race model and SOSP in reading times at the verb in grammatical sentences. The spreading activation race model predicts faster reading times in the semantically similar conditions compared to the dissimilar conditions, whereas the implemented SOSP model produces slower reading times for the semantically similar conditions. In line with the predictions of SOSP, we observed a slowdown in reading times at the verb for the semantically similar N2s. This suggests that competition between words in a sentence that are semantically similar has an inhibitory effect and not a facilitatory effect, as would be expected under the spreading activation race model.

In the reading times at the verb, the SOSP model predicted an interaction between N2 number and semantic similarity such that the slowdown in the singular N2 conditions would be exaggerated in the the semantically similar conditions. (The spreading activation race model did not make clear predictions of an interaction.) We found no strong evidence for slower reading times in the singular N2 conditions ($p(\beta < 0) = 0.104$) or for the interaction in the human data. In fact, the human data for the interaction go in the opposite direction numerically ($p(\beta < 0) = 0.800$, see Fig. 4). However, despite the relatively large sample size, the credible intervals for both effects are quite large, so we simply cannot draw any strong conclusions with regard to this SOSP prediction.

We also observed an interaction in the spillover region following the verb. In the semantically dissimilar conditions, participants read the singular N2 condition faster than the plural N2 condition. While it makes intuitive sense that encountering a singular verb after recently seeing a plural noun might disrupt processing, it is not clear why this pattern should only obtain when the N2 is semantically dissimilar to the N1. In the semantically similar conditions, though, the pattern was reversed, with longer reading times in the singular N2 case. This pattern is consistent with what the SOSP model predicted at the verb and is similar to the pattern reported in Nicenboim et al. (2018). Nicenboim et al. (2018) used a design similar to the present experiment: All of their sentences were grammatical, all of the nouns preceding the verb of interest were semantically related, and the number on the two nouns intervening between the grammatical subject and the verb was manipulated. Thus, this effect seems to be a simple case of similarity-based interference. In our experiment, though, cue-based retrieval cannot be the explanation, as the similarity between the N1 and the N2 was not related to retrieval features. This is a somewhat puzzling pattern of results, and since SOSP cannot currently predict spillover effects, we choose to remain agnostic about the cause of the interaction in the spillover region and focus on the results at the verb, which largely support SOSP's prediction of similarity-based slowdowns. We discuss spillover effects more broadly in the General Discussion.

The results of the comprehension question accuracy and response times are also consistent with SOSP. In line with previous findings (e.g., Villata et al., 2018, who used number & gender features), participants were faster and more likely to answer the question correctly if N1 and N2 were dissimilar, suggesting that the dissimilarity facilitated building the correct structure. To account for this, Villata et al. (2018) and Villata and Franck (2019) suggest an extension to SOSP for predicting comprehension question performance. They assume that answering a question about a sentence leaves the words in the sentence and their lexical features activated, but turns off the attachment links between them. The attachment links are then allowed to reform under the usual dynamics. Participants answer the question based on which link structure forms, and their response times depend on the harmony of the structure that is reconstructed, just as in online processing. In sentences with much feature overlap between items, this predicts slower question answering times and lower accuracies, both of which we observed. This is because competition between nouns with similar features can lead to slowdowns and the production of less-than-perfect structures, just as in online processing.

¹¹ The timing predictions from the SOSP model are in arbitrary time units. To rescale these to the range of the human reading times, we fit a linear regression model of the human reading times at the verb (by condition) as a function of the median model settling times. We then transformed the model's settling times and first and third quartiles using the fitted intercept and slope from the linear regression model: Predicted model settling time in milliseconds = $343.658 + 1.133 \times \text{model settling time}$.

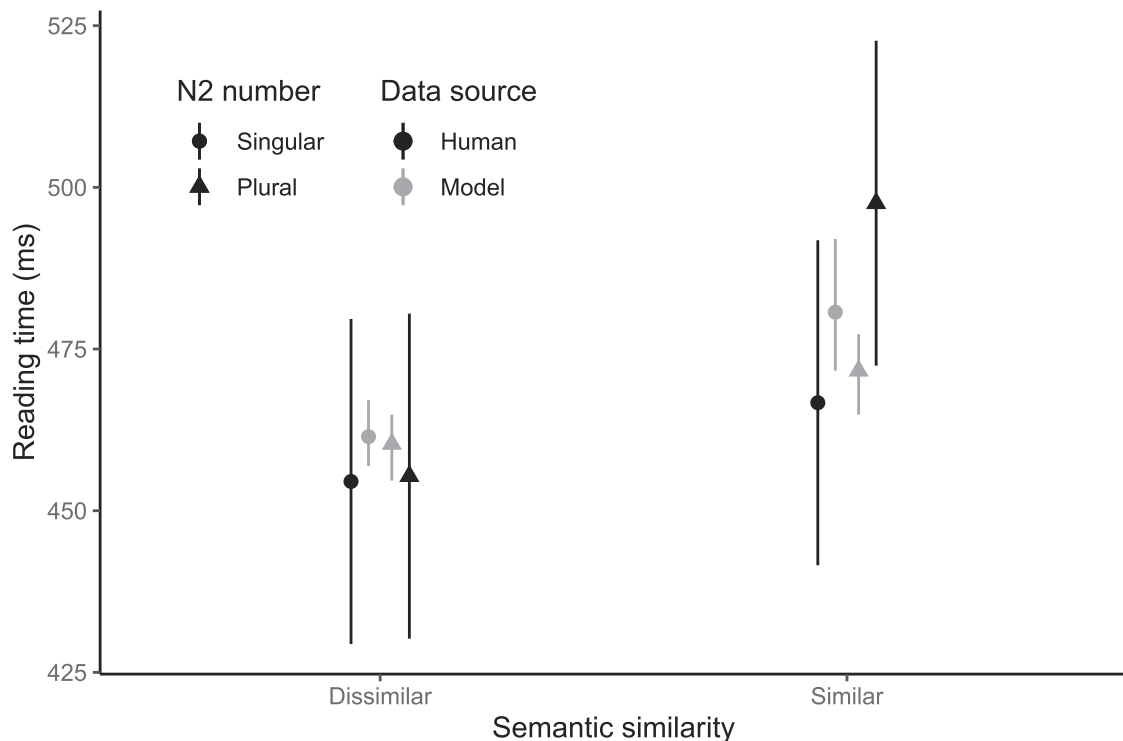


Fig. 4. Mean reading times at the verb (in milliseconds) with 95% confidence intervals (calculated from the data) from Experiment 1 (in black) and the rescaled SOSP model data, showing the median with error bars ranging from the first to the third quartile (in grey).

We also observed an interaction between N2 number and semantic similarity in the question accuracies and response times. These results are predicted under the Villata et al. link reassembly mechanism, if we assume that structures are reassembled at approximately the same frequencies and speeds as they are during online processing. In that case, the interactions in the question answering responses line up quite well with SOSP's predictions, especially with regard to question answering times, where the spreading activation race model makes opposite predictions to SOSP.

4. Experiment 2

Overall, Experiment 1 provides support for SOSP's prediction of similarity-induced slowdowns at the verb in the form of a main effect of semantic similarity. It included only grammatical test items, but SOSP and the spreading activation race model also make predictions about ungrammatical sentences. As shown above in Fig. 3, SOSP predicts slowdowns in ungrammatical (plural-verb) sentences for both semantically similar and plural N2s, with a larger difference between singular and plural N2s when the N2 is also semantically similar to the N1. The spreading activation race model, on the other hand, predicts faster processing when the N2 is plural and semantically similar to the N1. Thus, Experiment 2 included the grammatical items from Experiment 1 along with ungrammatical versions of them, allowing us to test the full range of predictions from both models.

4.1. Method

4.1.1. Participants

A total of 109 participants from the University of Connecticut participant pool took part in the study for course credit. Three participants were removed because they reported being diagnosed with a speech or language problem; four were removed for having average comprehension question accuracies at chance (50%); and twelve were removed for either for not cooperating with instructions (e.g., reading all sentences aloud despite our instructions to read silently) or for reporting that they noticed a pattern in the sentences (e.g., that N1 was the correct answer to many comprehension questions). Ninety participants were entered into the analyses.

4.1.2. Materials

In addition to the grammatical items in (4) above (with singular verbs), Experiment 2 included ungrammatical versions of those four conditions with plural verbs, e.g., *The canoe by the cabin likely were damaged in the heavy storm.*

The number of test items was increased to 40 from 36, so that we could evenly distribute the conditions across eight lists for the 2 x 2 x 2 design. The comprehension questions had the same format as Experiment 1. We included 120 filler items, eighty of which had

comprehension questions for which the correct answer was the second noun of the sentence. To keep the number of ungrammatical items in the experiment in line with previous studies (e.g., Wagers et al., 2009), no ungrammatical fillers were included, resulting in 14.3% ungrammatical items for each participant. The fillers and test items were distributed across eight lists in a Latin square design. All items are provided in the [supplementary materials as Supplement 2](#).

4.1.3. Procedure

The procedure was identical to that of Experiment 1.

4.1.4. Analyses

The analyses followed the same procedures as Experiment 1. Singular verb conditions were coded as -0.5 and plural verb conditions as $+0.5$; all three factors and their two- and three-way interactions were included in both the fixed and random effects.

4.2. Results

4.2.1. Comprehension question accuracy

The raw accuracies by condition are shown in [Table 4](#). There was evidence that semantically dissimilar nouns produced better performance than semantically similar ones ($b = 0.281$, 95%CrI $[-0.025, 0.591]$, $p(\beta < 0) = 0.035$). Plural N2s resulted in lower comprehension accuracy than singular N2s ($b = -0.202$, 95% CrI $[-0.399, -0.004]$, $p(\beta < 0) = 0.978$). There was evidence for the two-way interaction between semantic similarity and verb number ($b = -0.380$, 95% CrI $[-0.824, 0.088]$, $p(\beta < 0) = 0.945$). To better understand this interaction, we fit an additional model that included the effect of semantic similarity nested within the levels of verb number (similar to simple main effects in ANOVA analyses; [Schad, Vasishth, Hohenstein, & Kliegl, 2019](#)). This analysis showed that the interaction between semantic similarity and verb number was driven by the grammatical singular-verb conditions, where accuracy was increased for semantically dissimilar nouns ($b = 0.477$, 95% CrI $[0.080, 0.895]$, $p(\beta < 0) = 0.011$). There was no effect of semantic similarity in the ungrammatical plural-verb conditions ($b = 0.071$, 95% CrI $[-0.306, 0.445]$, $p(\beta < 0) = 0.379$). The interaction between N2 number and semantic similarity ($b = -0.029$, 95%CrI $[-0.477, 0.396]$, $p(\beta < 0) = 0.541$), the interaction between N2 number and verb number ($b = -0.011$, 95%CrI $[-0.431, 0.393]$, $p(\beta < 0) = 0.522$), and the three-way interaction were all approximately centered on zero ($b = 0.466$, 95%CrI $[-0.336, 1.220]$, $p(\beta < 0) = 0.128$).

4.2.2. Comprehension question response times

Only times from correct trials were used. The only clear effect in this analysis was that of semantic similarity, which showed that response times were faster in the semantically dissimilar conditions ($b = -0.047$, 95% CrI $[-0.077, -0.016]$, $p(\beta < 0) = 0.999$). The other effects were relatively weak: N2 number ($b = -0.019$, 95% CrI $[-0.047, 0.008]$, $p(\beta < 0) = 0.917$), verb number ($b = 0.010$, 95% CrI $[-0.026, 0.048]$, $p(\beta < 0) = 0.299$), the semantic similarity-N2 number interaction ($b = 0.037$, 95% CrI $[-0.017, 0.093]$, $p(\beta < 0) = 0.087$), the semantic similarity-verb number interaction ($b = -0.003$, 95% CrI $[-0.051, 0.055]$, $p(\beta < 0) = 0.452$), the N2 number-verb number interaction ($b = -0.027$, 95% CrI $[-0.085, 0.030]$, $p(\beta < 0) = 0.825$), and the three-way interaction ($b = -0.002$, 95% CrI $[-0.109, 0.105]$, $p(\beta < 0) = 0.508$).

4.2.3. Reading times

As in Experiment 1, only reading times between 50 ms and 10 s were included in the analysis. Reading times were log transformed as above before being entered into the analysis. The raw reading times at the verb are plotted in [Fig. 5](#) along with the rescaled model predictions.¹²

Verb At the verb, the only clear effect was that of verb number: ungrammatical plural verbs were read more slowly than grammatical singular verbs ($b = 0.043$, 95%CrI $[0.011, 0.074]$, $p(\beta < 0) = 0.003$). The other effects were weak: N2 number ($b = -0.012$, 95%CrI $[-0.039, 0.014]$, $p(\beta < 0) = 0.817$), semantic similarity ($b = -0.018$, 95%CrI $[-0.049, 0.013]$, $p(\beta < 0) = 0.871$), the N2 number-semantic similarity interaction ($b = 0.005$, 95%CrI $[-0.047, 0.056]$, $p(\beta < 0) = 0.419$), the N2 number-verb number interaction ($b = -0.023$, 95%CrI $[-0.082, 0.035]$, $p(\beta < 0) = 0.792$), the semantic similarity-verb number interaction ($b = -0.038$, 95%CrI $[-0.092, 0.018]$, $p(\beta < 0) = 0.912$), and the three-way interaction ($b = -0.014$, 95%CrI $[-0.112, 0.088]$, $p(\beta < 0) = 0.608$).

However, we were interested in whether the effect of semantic similarity differed between singular and plural verbs, so we fit another model nesting the effect of semantic similarity within verb number. This analysis showed that there was a speed up for dissimilar nouns, but only in the ungrammatical, plural-verb sentences ($b = -0.037$, 95%CrI $[-0.083, 0.008]$, $p(\beta < 0) = 0.947$) and not the grammatical, singular-verb sentences ($b = 0.001$, 95%CrI $[-0.039, 0.043]$, $p(\beta < 0) = 0.474$).

Spillover Here, there were three effects of interest. The first is a slowdown when the preceding verb was ungrammatical (plural: $b = 0.085$, 95%CrI $[0.052, 0.119]$, $p(\beta < 0) \approx 0$). The second was a speedup for plural N2 conditions ($b = -0.033$, 95%CrI $[-0.061, -0.003]$, $p(\beta < 0) = 0.985$). These were qualified the commonly observed interaction between N2 number and verb number (e.g., [Wagers et al., 2009](#); [Dillon et al., 2013](#)); the cost for an ungrammatical plural verb is reduced when the N2 is plural ($b = -0.065$, 95% CrI $[-0.121, -0.009]$, $p(\beta < 0) = 0.987$). The effect of semantic similarity ($b = 0.000$, 95%CrI $[-0.028, 0.028]$, $p(\beta < 0) = 0.492$),

¹² As in Experiment 1, we regressed the human reading times at the verb on the median model settling times. The transformed model settling times and quartiles were calculated according as: Predicted model settling time in milliseconds = $360.358 + 0.755 \times \text{model settling time}$.

Table 4
Raw percent accuracy in each condition in Experiment 2.

	Singular verb		Plural verb	
	Singular N2	Plural N2	Singular N2	Plural N2
Semantically dissimilar	88%	85%	84%	83%
Semantically similar	81%	80%	84%	80%

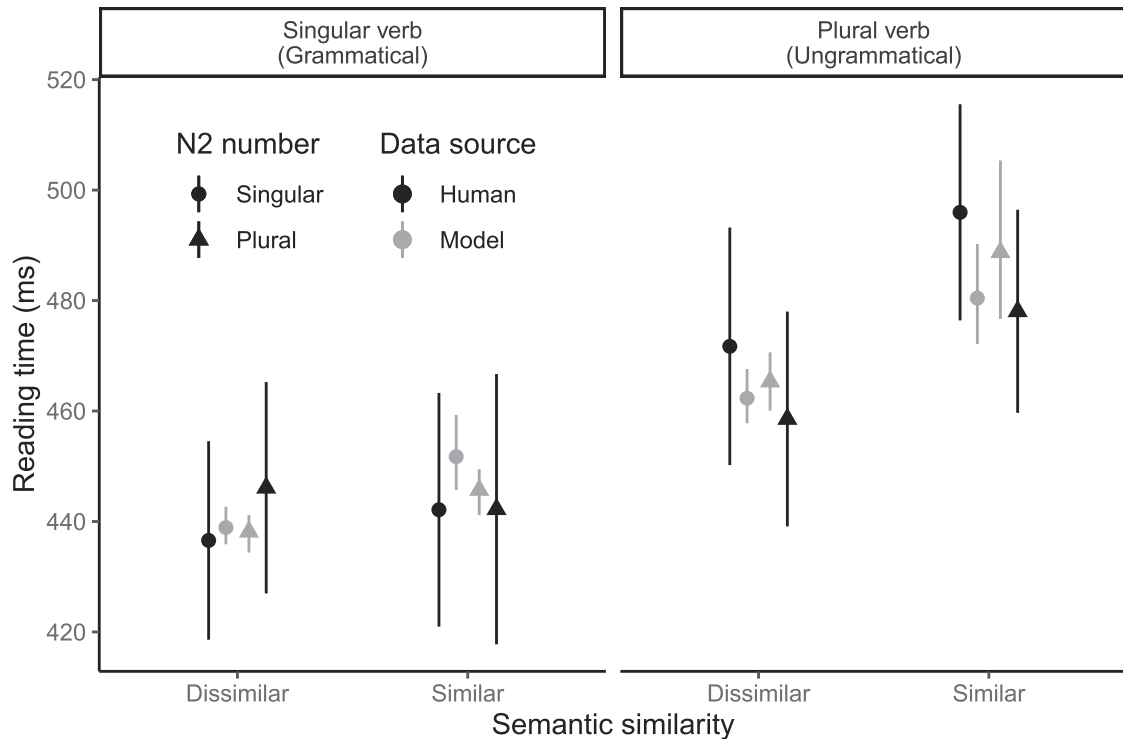


Fig. 5. Mean reading times at the verb (in milliseconds) with 95% confidence intervals (calculated from the data) from Experiment 2 (in black) and the rescaled SOSP model data, showing the median with error bars ranging from the first to the third quartile (in grey). The left facet shows the grammatical (singular verb) conditions and the right facet the ungrammatical (plural verb) conditions.

the N2 number-semantic similarity interaction ($b = -0.007$, 95%CrI $[-0.069, 0.054]$, $p(\beta < 0) = 0.593$), semantic similarity-verb number interaction ($b = 0.005$, 95%CrI $[-0.052, 0.061]$, $p(\beta < 0) = 0.432$), and the three-way interaction ($b = 0.036$, 95%CrI $[-0.086, 0.159]$, $p(\beta < 0) = 0.269$) were all approximately centered on zero in the spillover region.

4.3. Discussion

The results of Experiment 2 provide tentative support for some of the predictions of the implemented SOSP model over those of the spreading activation race extension of ACT-R. There are two effects to comment on from the reading time data. Most importantly for this study, there was an effect of semantic similarity in the verb region of ungrammatical sentences such that the dissimilar conditions were read faster. This is consistent with SOSP's predictions and inconsistent those of the spreading activation race model, which predicted faster processing. However, it is puzzling why we did not observe this effect in the grammatical sentences like we did in Experiment 1. It might be that repeatedly reading ungrammatical sentences led participants to develop a strategy of speeding through sentences until they found a cue—an ungrammatical verb—that signaled that extra consideration was necessary. Grammatical sentences contain no such cue, so participants might not give the sentence enough consideration to allow a semantic similarity effect to surface. Such an effect would be consistent with a growing body of literature showing important task effects on online sentence processing (Franck et al., 2015; Hammerly, Staub, & Dillon, 2019; Logačev & Vasishth, 2015; Swets, Desmet, Clifton, & Ferreira, 2008), and it suggests that our models of sentence processing need to specify how offline information can affect online processing.

The second reading time effect in Experiment 2 was a main effect of verb number. Ungrammatical, plural-verb sentences showed slower reading times in the spillover region. This makes sense, as participants should be perturbed by ungrammatical sentences (e.g., Parker, 2019; Dillon et al., 2013). Another alternative is that this effect might simply be an effect of the markedness of plurals in general (Wagers et al., 2009). We also note that we replicated the well-known interaction between verb number and N2 number.

Findings of a reduction in the ungrammaticality cost in the plural N2 conditions are common in the subject-verb agreement literature (so-called “illusions of grammaticality;” [Pearlmuter et al., 1999](#); [Wagers et al., 2009](#); [Lago et al., 2015](#); [Jäger et al., 2017](#), among others). This suggests that, while some of the effects in Experiment 2 were weak, our participants at least performed as expected on a classic manipulation in the field.

In the comprehension question accuracies for grammatical sentences, we replicated the result from Experiment 1: Participants were more likely to answer questions correctly in the semantically dissimilar conditions than in the similar conditions in grammatical (singular verb) sentences, but not in ungrammatical sentences. This was especially pronounced when the N2 was singular. This, along with faster question answering for semantically dissimilar items (both grammatical and ungrammatical), is consistent with SOSP. We also observed lower accuracy for plural N2s, which is not clearly predicted by either SOSP or the spreading activation race model.

Overall then, we must interpret the results of Experiment 2 somewhat tentatively. We know of no existing theory that would predict to the exact pattern of results observed, especially with regard to the similarity-induced reading slowdown in ungrammatical sentences only. It is possible that this unexpected result will turn out to be replicable in future experiments, in which case we will need to develop new theories to account for it. But the results of Experiment 2, such as they are, correspond more closely to the predictions of SOSP than to those of the spreading activation race model.

5. General discussion

The two experiments presented here were designed to pit the predictions of the spreading activation race model against those of the implemented SOSP model. In self-paced reading, the spreading activation race model predicts that we should observe faster reading times at the verb when it is preceded by two semantically related nouns. This is because the two nouns spread activation to each other, egging each other on in the race to the activation threshold for retrieval. The SOSP model predicts the opposite effect: Semantic similarity between the nouns should lead to competition-induced slowdowns in reading times at the verb. Experiment 1 clearly supports the prediction of SOSP in this regard. Participants read the verb more slowly when the preverbal nouns were semantically related. The ungrammatical conditions of Experiment 2 also support SOSP, although Experiment 2 did not replicate the similarity-based slowdown in the grammatical items. Overall, the reading time data are clearly more consistent with the predictions of SOSP than those of the spreading activation race model (which predicted results in the opposite direction to those observed). Caution is warranted, though, given the complex pattern of results in Experiment 2 and the lack of experimental evidence for the N2 number by semantic similarity interactions that SOSP predicted.

The reading time results from the two experiments were somewhat inconsistent, but the comprehension question data were more consistent. The comprehension question data can be explained quite well by the implemented SOSP model if we assume the link reassembly account described in [Villata et al. \(2018\)](#) and [Villata and Franck \(2019\)](#) and sketched above. [Villata et al. \(2018\)](#) propose that, in order to answer comprehension questions, all of the word features in a sentence are left active, but the links are set to zero strength. From this state, the system settles again to some attractor, and participants answer the question based on the structure that was reassembled. The comprehension question accuracies in Experiment 1 and the grammatical items of Experiment 2 supported this idea: Participants were more accurate when the nouns were dissimilar. In both experiments, the comprehension question response times are also consistent with SOSP, with participants responding faster to questions when the nouns were dissimilar. To speculate about what the spreading activation race model might predict, let us assume that question answering works similarly to online processing. In that case, having similar nouns in the sentence should lead to faster question answering times (due to egging-on through spreading activation) but lower accuracy (due to the N2 winning races more frequently) in the semantically similar conditions. This could account for the observed lower accuracies in those conditions but not the observed response times. Overall, then, the human question accuracy and latency data are more consistent with SOSP than with the spreading activation race model, assuming computational implementations of these verbal models (link reassembly and spreading activation race model for comprehension questions) produce predictions consistent with our speculations.

Why do we observe this consistency between experiments only in the offline measures? One explanation is good enough processing ([Ferreira, Bailey, & Ferraro, 2002](#); [Ferreira & Patson, 2007](#)). Online, participants might only form representations that are detailed enough to extract some basic meaning from the sentence. But when asked a comprehension question, they are forced to construct a more detailed analysis of the sentence, for example using the link reassembly account of [Villata et al. \(2018\)](#) and [Villata and Franck \(2019\)](#). If participants always use link reassembly after the sentence but only sometimes allow attachment links to form while reading the sentence, then this might explain the consistent comprehension question data and inconsistent online reading time data.

While SOSP receives some support from the human data, it is important to acknowledge the limitations of our simulations. In the ungrammatical plural verb conditions, the model produced slower processing for plural N2s. This is the opposite of the results of our Experiment 2 in the spillover region and of [Jäger et al. \(2017\)](#)'s meta-analysis for subject verb agreement. Pooling the data from dozens of experiments, they found evidence for speeded reading times when the N2 is plural in ungrammatical sentences ($N1[sg.]-N2[pl.]-V[pl.] < N1[sg.]-N2[sg.]-V[pl.]$), the classic illusion of grammaticality. As discussed in the introduction, this is a natural prediction of ACT-R, and by extension, the spreading activation race model. Extensive explorations of SOSP's parameter space are underway to determine the full range of the model's predictions ([Roberts & Pashler, 2000](#)) and to see whether there are any parameter settings for which SOSP predicts this well-established effect.

A further limitation of the current implementation of SOSP is that it cannot predict spillover effects. Spillover effects are observed very often and are used to draw conclusions about processing in preceding sentence regions. But there is no leading theory of why we should observe spillover effects at all. [Cho, Szudlarek, and Tabor \(2016\)](#) suggest that certain dynamical language processors can be perturbed by unexpected linguistic material and that it can take multiple subsequent words to recover from such a perturbation. While

this is an appealing account, its applicability to SOSP, in the form we present here, seems limited. Unexpected or ill-fitting material can take extra time to integrate into the existing structure, but SOSP processes each word completely, integrating it into the existing structure, before moving to the next word. Spillover effects are important and widespread, though, so the system should be developed further to explain them. A “commitment policy” that systematically controls how firmly a model commits to a structure before inputting the next word (Cho & Smolensky, 2016; Cho et al., 2017) might be one way to account for spillover effects. Another possibility would be to separate structure building from the decision to move to the next word into separate processes. This allows the next word to be input before processing is complete on the previous one (Mitchell, 1984), as has been done in Brasoveanu and Dotlačil (2019)’s Python implementation of an ACT-R parser with eye-movement control.

Another approach to spillover effects is discussed in Smith and Levy (2013). They argue that a highly incremental parser that uses preceding words (or sub-word units) to predict upcoming words naturally predicts spillover effects. The current implementation of SOSP is not compatible with this, as it simulates processing at a single word after already fully integrating information from previous words. However, a more fully developed, incremental version of SOSP might be able to predict spillover effects. For incremental parsing, a fuller SOSP model would include peaks in the harmony landscape that correspond to the partial parses available as each word in the input is turned on. Parsing functions just as in the model implemented here: the system moves noisily uphill from a starting position determined by the word it just read toward a nearby harmony peak. When a new word is read, the state of the system is displaced to a new part of the harmony landscape and the process repeats. High-harmony partial parses create a tall peak in the harmony landscape, but as the slopes of the peak taper away, they have the effect of flattening the harmony landscape nearby. If, when a new word is read, it is displaced close to the high-harmony peaks associated with the partial parses available at the last word, then the processing of the new word will be slowed relative to a condition in which the partial parses at the previous word had lower harmony values. Thorough testing of the incremental system is needed to verify these speculations, but if borne out, they might provide a natural mechanism for explaining spillover effects via self-organization.

Turning now to other approaches, we note that SOSP is not alone in being able to account for the general pattern of results. Hofmeister and Vasishth (2014), building on a feature overwriting approach by Nairne (1990), argue that features of words can be overwritten by features of subsequent words at the time they are encoded into memory. Thus, when the parser tries to retrieve the first noun, it is less distinct from other words and therefore more difficult to retrieve. If ease of identification is inversely related to retrieval time, this approach also predicts the similarity-based slowdowns we observed in Experiments 1 and 2.

However, we see a couple of issues with this approach. First, feature overwriting during sentence processing seems to imply much more labile representations than are typically assumed, even under SOSP. The noisy channel theory of Levy (2008) and Futrell, Gibson, and Levy (2020) does allow representations of words to change after reading via added random noise, but it does not say that the noise that affects one word should be correlated with features on another word, as would be required under the Nairne (1990) mechanism. Similarly, SOSP, as currently implemented, allows words’ features to change during processing if the system settles to a harmony peak that corresponds to different features being active. This is unlikely because the system would have to traverse relatively large distances in state space purely under the influence of noise. Even if it did occur, there is again no feature-specific influence happening between words. Patson and Husband (2016)’s noun phrase misinterpretation data might be an example of such feature overwriting, but, to our knowledge, there is little empirical data available supporting large-scale feature overwriting, which is not predicted by most current theories anyway. Thus, while feature overwriting might be able to explain our results, it might go too far in its predictions.

Secondly, the feature overwriting proposal requires an additional mechanism beyond what is needed for assembling syntactic structure in order to capture the results, namely, some way for features to change that is separate from perception and parsing. Having such an additional mechanism might be worth the added complexity if, for example, it could explain the unexpected finding of slowdowns for semantic similarity only in ungrammatical sentences. However, since neither SOSP nor feature overwriting clearly predict this effect, parsimony favors SOSP.

Finally, Villata et al. (2018) consider a different type of extension to cue-based retrieval: activation leveling. Activation leveling extends the fan effect in ACT-R—where activation is split among all chunks sharing a retrieval cue—to other features not necessarily relevant for retrieval. If two chunks have similar features, their activations will tend to become more equal: The activation of the more highly activated chunk will be lowered, and the activation of the less activated chunk will be raised. This predicts slower processing when features are shared between multiple words, just like SOSP and against the prediction of the spreading activation race model. However, as Villata et al. (2018) argue, SOSP’s ability to account for the observed interference effects using only mechanisms that are otherwise needed to build parses gives it the benefit of parsimony over any extension to cue-based retrieval that requires separate mechanisms for encoding and retrieval effects.

This is indeed SOSP’s great strength: there is only structure building. Thus, what memory-based accounts consider to be two separate processes—encoding and retrieval—can actually be thought of as two manifestations of a single structure-building process in SOSP. The work presented here is a step toward determining how far this theoretical parsimony can get us in explaining how people understand sentences.

Acknowledgements

This project was supported in part by NSF IGERT grant DGE-1144399.

Appendix A. Prior specifications for all analyses

We chose priors so that, before fitting to the data, the statistical models would produce outputs that accord with what we know

about typical effects in sentence processing, following the recommendations in [Schad et al. \(2019\)](#). Due to the similar structure of the experiments, the priors were kept constant for both experiments. Model convergence was checked by visually inspecting the mixing of the sampling chains.

The full data sets for both experiments, as well as the analysis scripts, are available at <https://osf.io/hjrkn/>.

Comprehension question accuracy

- Intercept $\sim N(0, 5)$
- Regression coefficients $\sim N(0, 1)$
- Random effects covariance matrix $\sim LKJ(2)$
- Random effects standard deviation $\sim \text{Truncated } N(0, 1)$

Comprehension question response time

- Intercept $\sim N(7, 1)$
- Regression coefficients $\sim N(0, 1)$
- Residual standard deviation $\sim \text{Truncated } N(0, 1)$
- Random effects covariance matrix $\sim LKJ(2)$
- Random effects standard deviation $\sim \text{Truncated } N(0, 1)$

Reading time

- Intercept $\sim N(6, 1)$
- Regression coefficients $\sim N(0, 0.5)$
- Residual standard deviation $\sim \text{Truncated } N(0, 1)$
- Random effects covariance matrix $\sim LKJ(2)$
- Random effects standard deviation $\sim \text{Truncated } N(0, 1)$

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.cogpsych.2020.101356>.

References

- Adani, F., Van der Lely, H. K., Forgiarini, M., & Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua*, *120*(9), 2148–2166.
- Anderson, J. R. (1990). *The adaptive character of thought*. Lawrence Erlbaum Associates Inc.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. <https://doi.org/10.1037/0033-295X.111.4.1036>.
- Barker, J., Nicol, J., & Garrett, M. F. (2001). Semantic factors in the production of number agreement. *Journal of Psycholinguistic Research*, *30*(1), 91–114.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.
- Brasoveanu, A. & Dotlačil, J. (2019). *Computing dynamic meanings: Building integrated competence-performance theories for formal semantics*. (Downloaded 25.03.2019).
- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>.
- Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2019). *shiny: Web application framework for r* [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=shiny> (R package version 1.3.2).
- Cho, P. W., Goldrick, M., Lewis, R. L., & Smolensky, P. (2018). Dynamic encoding of structural uncertainty in gradient symbols. In *Proceedings of the 8th workshop on cognitive modeling and computational linguistics*.
- Cho, P. W., Goldrick, M., & Smolensky, P. (2017). Incremental parsing in a continuous dynamical system: Sentence processing in Gradient Symbolic Computation. *Linguistics Vanguard*, *3*(1).
- Cho, P. W., & Smolensky, P. (2016). Bifurcation analysis of a gradient symbolic computation model of incremental processing. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 1487–1492).
- Cho, P. W., Szudlarek, E., & Tabor, W. (2016). Discovery of a recursive principle: An artificial grammar investigation of human learning of a counting recursion language. *Frontiers in Psychology*, *7*, 867. <https://doi.org/10.3389/fpsyg.2016.00867>.
- Ciocoiu, I. B. (1996). Analog decoding using a gradient-type neural network. *IEEE Transactions on Neural Networks*, *7*(4), 1034–1038.
- Ciocoiu, I. B. (2009). Invariant pattern recognition using analog recurrent associative memories. *Neurocomputing*, *73*, 119–126. <https://doi.org/10.1016/j.neucom.2009.02.024>.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(7), 7–29.
- di Bernardo, M., Budd, C. J., Champneys, A. R., & Kowalczyk, P. (2008). *Piecewise-smooth dynamical systems: Theory and applications*. Springer.
- 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.
- Engelmann, F., Jäger, L. A., & Vasishth, S. (2019). The effect of prominence and cue association in retrieval processes: A computational account. *Cognitive Science*, *43* (e12800).
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15.
- Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83.

- Franck, J., Colonna, S., & Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in Psychology*, 6, 349.
- Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, 44(3), e12814. <https://doi.org/10.1111/cogs.12814>.
- Gaifman, H. (1965). Dependency systems and phrase-structure systems. *Information and Control*, 8, 304–337.
- Goebel, R., Sanfelice, R. G., & Teel, A. R. (2009). Hybrid dynamical systems. *IEEE Control Systems Magazine*. <https://doi.org/10.1109/MCS.2008.931718>.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(6), 1411–1423.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51, 97–114.
- Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(6), 1304–1321. <https://doi.org/10.1037/0278-7393.32.6.1304>.
- Green, M. J., & Mitchell, D. C. (2006). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, 55, 1–17. <https://doi.org/10.1016/j.jml.2006.03.003>.
- Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, 110(70–104). <https://doi.org/10.1016/j.cogpsych.2019.01.001>.
- Han, J. Y., Sayeh, M. R., & Zhang, J. (1989). Convergence and limit points of neural network and its application to pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5), 1217–1222.
- Hays, D. G. (1964). Dependency theory: A formalism and some observations. *Language*, 40(4), 511–525.
- Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology*, 3(292). <https://doi.org/10.3389/fpsyg.2012.00292>.
- Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, 26(3), 376–405.
- Hofmeister, P., & Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01237>.
- 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, 316–339. <https://doi.org/10.1016/j.jml.2017.01.004>.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111(2), 228–238.
- Kempen, G., & Vosse, T. (1989). Incremental syntactic tree formation in human sentence processing: A cognitive architecture based on activation decay and simulated annealing. *Connection Science*, 1(3), 273–290.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 380–602.
- Konieczny, L. (2005). The psychological reality of local coherences in sentence processing. In *Proceedings of the 27th annual conference of the cognitive science society* (pp. 1178–1183).
- Kruschke, J. K., & Liddell, T. M. (2017). The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-016-1221-4>.
- Kush, D., Johns, C. L., & Van Dyke, J. A. (2015). Identifying the role of phonology in sentence-level reading. *Journal of Memory and Language*, 79–80, 18–29.
- Lago, S., Shalom, D. E., Sigman, M., Lau, E., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, 82, 133–149. <https://doi.org/10.1016/j.jml.2015.02.002>.
- Laurinavichyute, A., Jäger, L. A., Akinina, Y., Roß, J., & Dragoy, O. (2017). Retrieval and encoding interference: Cross-linguistic evidence from anaphor processing. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00965>.
- Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the 2008 conference on empirical methods in natural language processing* (pp. 234–243). Association for Computational Linguistics.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375–419.
- Logacev, P., & Vasishth, S. (2015). A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension. *Cognitive Science*, 40(2), 1–33. <https://doi.org/10.1111/cogs.12228>.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703.
- Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdtke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Frontiers in Psychology*, 10, 2767. <https://doi.org/10.3389/fpsyg.2019.02767>.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67–91.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283–312.
- Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processing in reading. In D. Kieras, & M. A. Just (Eds.), *New methods in reading comprehension research*. Hillsdale, New Jersey: Erlbaum.
- Muezzinoglu, M. K., & Zurada, J. M. (2006). RBF-based neurodynamic nearest neighbor classification in real pattern space. *Pattern Recognition*, 39, 747–760.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18(3), 251–269.
- 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.
- Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science*, 42(S4), 1075–1100. <https://doi.org/10.1111/cogs.12589>.
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55, 601–626. <https://doi.org/10.1016/j.jml.2006.08.009>.
- Paape, D., & Vasishth, S. (2015). Local coherence and preemptive digging-in effects in German. *Language and Speech*, 1–17. <https://doi.org/10.1177/0023830915608410>.
- Parker, D. (2019). Cue combinatorics in memory retrieval for anaphora. *Cognitive Science*, 43(e13715), 1–30. <https://doi.org/10.1111/cogs.12715>.
- Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology*, 69(5), 950–971. <https://doi.org/10.1080/17470218.2014.992445>.
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41, 427–456.
- Potsdam, E. (1998). A syntax for adverbs. In *Proceedings of the twenty-seventh western conference on linguistics* (pp. 397–411).
- Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24(5), 574–590.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107(2), 358–367.
- Rouder, J. N., Province, J. M., Morey, R., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, 80(2), 491–513. <https://doi.org/10.1007/s11336-013-9396-3>.
- Schad, D., Betancourt, M., & Vasishth, S. (2019). *Toward a principled bayesian workflow in cognitive science*. Retrieved from <https://www.osf.io/b2vx9>.
- Schad, D., Vasishth, S., Hohenstein, S., & Kliegl, R. (2019). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110.
- Smith, G. (2018). *A theory of timing effects in a self-organizing model of sentence processing (Unpublished doctoral dissertation)*. University of Connecticut.
- Smith, G., Franck, J., & Tabor, W. (2018). A self-organizing approach to subject-verb number agreement. *Cognitive Science*, 42(S4), 1043–1074. <https://doi.org/10.1111/cogs.12591>.
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128, 302–319. <https://doi.org/10.1016/j.cognition.2013.02.013>.

- Smith, G., & Tabor, W. (2018). Toward a theory of timing effects in self-organized sentence processing. In I. Juvina, J. Hout, & C. Myers (Eds.), *Proceedings of the 16th international conference on cognitive modeling* (pp. 138–143).
- Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D.E. Rumelhart, J.L. McClelland & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (vol. I: Foundations, pp. 194–281). MIT Press.
- Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, 23(4), 295–322.
- Stevenson, S. (1994). *A competitive attachment model for resolving syntactic ambiguities in natural language parsing* (Unpublished doctoral dissertation). University of Maryland.
- Stevenson, S., & Merlo, P. (1997). Lexical structure and parsing complexity. *Language and Cognitive Processes*, 12(2–3), 349–399. <https://doi.org/10.1080/016909697386880>.
- Swets, B., Desmet, T., Clifton, C. J., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, 36(1), 201–216. <https://doi.org/10.3758/MC.36.1.201>.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50(4), 355–370.
- Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 431–450.
- Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, 48, 740–759.
- van der Velde, F., & de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences*, 29, 37–108.
- van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 621–648). Elsevier.
- van Maanen, L., & van Rijn, H. (2007). An accumulator model of semantic interference. *Cognitive Systems Research*, 8, 174–181.
- van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review*, 16(6), 987–999. <https://doi.org/10.3758/PBR.16.6.987>.
- van Maanen, L., van Rijn, H., & Taatgen, N. A. (2012). RACE/A: An incremental architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science*, 36, 62–101.
- Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, 71, 147–161. <https://doi.org/10.1016/j.wocn.2018.07.008>.
- Villata, S., & Franck, J. (2019). Similarity-based interference in agreement comprehension and production: Evidence from object agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000718>.
- Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, 9(2), 1–16.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75, 105–143.
- Vosse, T., & Kempen, G. (2009). The Unification Space implemented as a localist neural net: Predictions and error-tolerance in a constraint-based parser. *Cognitive Neurodynamics*, 3(4), 331–346.
- Wagers, M., Lau, E., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206–237. <https://doi.org/10.1016/j.jml.2009.04.002>.