11-30-2018

A Theory of Timing Effects in a Self-Organizing Model of Sentence Processing

Garrett Smith

University of Connecticut - Storrs, garrett.smith@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/dissertations

Recommended Citation

A Theory of Timing Effects in a Self-Organizing Model of Sentence Processing
Garrett Smith, Ph.D.
University of Connecticut, 2018

Many leading theories of human sentence processing assume that language comprehension and production take place under the strict control of a symbolic grammar. For example, in sentence comprehension, reading or hearing a word triggers the application of a grammar rule that incorporates the word into the existing sentence structure so that the resulting structure is consistent with the all of the rules of the grammar. These theories have had wide success in explaining important timing effects, e.g., predicting speed-ups or slowdowns while reading a sentence. A number of phenomena have been identified, though, that challenge these grammar-controlled theories and motivate the development of an alternative theory. In *local coherence* effects, people seem to entertain syntactic structures that are compatible with a subset of the words in a sentence but ungrammatical in the context of the rest of the sentence. *Agreement attraction* occurs when the verb of a sentence agrees in number with a noun other than the subject, in violation of the rules of a grammar. The existence of these phenomena, which grammar-controlled theories struggle to account for, motivates *self-organizing sentence processing-treelet harmony* (SOSP-TH), the focus of this dissertation. Instead of being strictly controlled by a symbolic grammar, lexically anchored syntactic treelets in SOSP-TH self-organize into larger structures via local interactions that try to maximize the well-formedness (harmony) of the resulting structure. Importantly, SOSP-TH includes less-than-perfect syntactic structures, which allows it to account for local coherence and agreement attraction effects as a
natural by-product of its strongly bottom-up syntactic processing. In contrast to many previous self-organizing models, the mathematical formulation of SOSP-TH allows us to make precise predictions about processing times, which we test in three experiments on interference effects in subject-verb agreement. Overall, the model provides a good fit to the human data and provides a parsimonious explanation for the semantic interference effects tested in the experiments. SOSP-TH does face some challenges in accounting for certain data points, but those, combined with a set of new predictions, make it a promising theory for future research at the intersection of theoretical linguistics, dynamical systems modeling, and experimental psycholinguistics.
A Theory of Timing Effects in a Self-Organizing Model of Sentence Processing

Garrett Smith

B.A., Texas Tech University, 2010
M.A., Humboldt-Universität zu Berlin, 2014
M.S., University of Connecticut, 2016

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2018
Completing a dissertation is a lot of work. Actually writing the document took a (very harried) month or so, but that was the culmination of years of support, instruction, guidance, commiseration, encouragement, and patience from many people. All of them have my thanks, but I want to highlight a few especially important supporters without whom I could not have finished this project.

First, I want to thank Dr. Whitney Tabor, my advisor. He is exceedingly generous with his time and invested countless hours working with me, teaching me, and thinking through everything from statistical analyses to international politics. Whit combines the magnanimity of taking every idea seriously (no matter how half-baked) with an uncompromising need to interrogate and test every underlying assumption (no matter how venerated). This made being his student challenging but exceedingly rewarding.

I also want to thank my committee members, Dr. Gerry Altmann, Dr. Julie Franck, Dr. Till Frank, and Dr. James Magnuson for many helpful discussions and comments as I developed the dissertation. Working with this intellectually diverse and stimulating group of scholars over the last few years has really enriched my training.

I never could have gotten anywhere near a Ph.D. if it weren’t for my family. When my parents first explained higher education with all its different possible degrees, I told them I wanted one of each, from an associate’s degree to a Ph.D. I don’t know how seriously my parents took me at the time, but they and my sister have never been anything less fully supportive of me and whatever crazy, unemployable skills I’ve wanted to learn next. They helped me find the path that took me to this point.
Finally, I want to thank Pam, who has encouraged me, been a sounding board—
I’ve talked to her about SOSP almost as much as Whit—and made the last few years,
trying as they were, endlessly fun and full of love.
Contents

1 Introduction 1
  1.1 Two approaches to sentence processing 1
  1.2 Two challenges for grammar-controlled theories 5
    1.2.1 Local coherence effects 5
    1.2.2 Interference in subject-verb number agreement 12
  1.3 Roadmap 16

2 Smith and Tabor (2018) 19

3 SOSP-TH details 27
  3.1 Introduction 27
  3.2 Linguistic representations 28
  3.3 Design and implementation choices 31
  3.4 Mathematical properties 34
  3.5 Example models 40
    3.5.1 Classic agreement attraction 40
    3.5.2 An incremental local coherence model 45
3.6 Discussion .......................................................... 53

4 Pseudopartitives .................................................. 57
4.1 Introduction ....................................................... 57
4.2 An SOSP-TH model of verb choice ..................... 62
4.3 Experiment 1 .................................................... 70
   4.3.1 Method .................................................... 70
   4.3.2 Results .................................................... 72
   4.3.3 Discussion ................................................ 75
4.4 General Discussion ........................................... 77

5 Encoding interference .......................................... 81
5.1 Introduction ..................................................... 81
5.2 An SOSP-TH approach to encoding interference ........ 87
   5.2.1 Model design ........................................... 89
   5.2.2 Results and discussion ................................ 90
5.3 Experiment 2A .................................................. 93
   5.3.1 Method .................................................... 94
   5.3.2 Results .................................................... 98
   5.3.3 Discussion ................................................ 101
5.4 Experiment 2B .................................................. 103
   5.4.1 Method .................................................... 104
   5.4.2 Results .................................................... 106
   5.4.3 Discussion ................................................ 111
5.5 General Discussion .................................................. 112

6 General discussion ................................................. 115

6.1 Summary and assessment of results ......................... 115
  6.1.1 Chapter 1: Two approaches to sentence processing .... 115
  6.1.2 Chapter 2: Introducing SOSP-TH ......................... 117
  6.1.3 Chapter 3: From equations to full-fledged models ...... 119
  6.1.4 Chapter 4: Pseudopartitives ............................... 121
  6.1.5 Chapter 5: Encoding interference ....................... 122

6.2 Future directions .................................................. 124
  6.2.1 Accounting for human timing data ...................... 124
  6.2.2 Extending the model ....................................... 130

6.3 Conclusion ......................................................... 135

Appendix: Experimental materials ............................. 136
  A.1 Experiment 1: Pseudopartitives ......................... 137
  A.2 Experiments 2A and 2B: Encoding interference ....... 140

References ............................................................. 163
Chapter 1

Introduction

1.1 Two approaches to sentence processing

Many leading theories of human sentence processing (Frazier & Fodor, 1978; Futrell & Levy, 2017; Gibson, 2006; Hale, 2001, 2003, 2011; Jurafsky, 1996; Levy, 2008a; Lewis & Vasishth, 2005; McElree, Foraker, & Dyer, 2003; Rasmussen & Schuler, 2017) assume that, during language comprehension and production, fragments of linguistic structure are assembled under the control of a symbolic grammar. For example, in language comprehension, a word in the input triggers symbolic grammar rules that tell the mind how to incorporate that word into a representation of what has come before and what is likely to come next. Grammar-controlled theories of this sort have been very successful at explaining many sentence processing phenomena, including garden paths, the subject-/object-relative asymmetry, the difficulty of multiple center-embeddings, similarity-based interference, and frequency effects. The success of these
theories lends support to the idea that the mind is strongly guided by symbolic rules during language processing.

However, two important classes of sentence processing effects challenge grammar-controlled theories: local coherence effects (Cai, Sturt, & Pickering, 2012; Konieczny, 2005; Konieczny, Müller, Hachmann, Schwarzkopf, & Wolfer, 2009; Levy, Bicknell, Slattery, & Rayner, 2009; Paape & Vasishth, 2015; Tabor, Galantucci, & Richardson, 2004) and interference in subject-verb number agreement (e.g., Barker, Nicol, & Garrett, 2001; Bock & Miller, 1991; Pearlmutter, Garnsey, & Bock, 1999; Smith, Franck, & Tabor, 2018; Villata, Tabor, & Franck, 2018). In both cases, people sometimes behave as if their minds are flouting symbolic rules by entertaining linguistic structures that are not derivable from a plausible grammar. When grammar-controlled theories attempt to explain these effects, they either rely on phenomenon-specific processing modules, weakening the generality of the models (e.g., Eberhard, Cutting, & Bock, 2005), or if the theories remain general-purpose (Levy, 2008b; Lewis & Vasishth, 2005), they fail to account for the full range of observed grammar-flouting effects.

An alternative approach to sentence processing, *self-organized sentence processing* (SOSP), handles these difficult cases as a natural consequence of basic structure-building mechanisms (beim Graben, Pinotsis, Saddy, & Potthast, 2008; Cho, Goldrick, Lewis, & Smolensky, 2018; Cho, Goldrick, & Smolensky, 2017; Cho & Smolensky, 2016; Gerth & beim Graben, 2009; Kempen & Vosse, 1989; Smith et al., 2018; Smith & Tabor, 2018; Stevenson, 1994a; Stevenson & Merlo, 1997; Tabor & Hutchins,
In self-organizing theories, small pieces of linguistic structure, usually lexically anchored syntactic treelets (Fodor, 1998, 2017), are not controlled or manipulated by any kind of grammatical overseer. Instead, they interact with each other spontaneously, using only local information to competitively form continuous-valued attachment links that grow and decay depending on the syntactic and semantic feature match between the linked attachment sites. In this way, these self-organizing models parallel other cases of self-organization where local interactions between small subunits result in the emergence of large-scale structure under certain circumstances, e.g., the configuration of atoms in molecules and protein folding (Wales, 2003), the orientation of spins in magnetic materials (Wu, 1982), and biological structure formation (Haken, 1983; Turing, 1952). Self-organizing theories are interactive theories in that information at any level of analysis (lexical, syntactic, semantic, pragmatic) is allowed to influence structure building (Altmann & Steedman, 1988; Marslen-Wilson, 1975; Taraban & McClelland, 1988). This puts them in the same class as other interactive models of language processing (e.g., TRACE and the interactive activation model of word and letter perception; McClelland & Elman, 1986; McClelland & Rumelhart, 1981).

The instantiation of SOSP presented in Chapters 2 and 3, called SOSP-TH (“treelet harmony” to distinguish it from other SOSP models), starts with a global harmony function that assigns a continuous-valued well-formedness score to every possible treelet and link configuration available given a lexicon and corpus of word sequences. The framework is flexible, though; it can be used to implement the

---

1 I group all of these together under the term SOSP even though the various models differ in many regards.
processing of any grammar whose structures can be represented as numerical vectors. The assumptions and design choices discussed in Chapters 2 and 3 were made to capture important effects in sentence processing. The results of these choices justify the use of this framework while highlighting areas where different assumptions may be necessary for other phenomena.

During processing, SOSP-TH noisily follows the gradient of the harmony function to build a (partial) parse that locally maximizes the feature match between linked treelets. The attractors of the system (points to which the system returns after small perturbations; Strogatz, 1994) include both high-harmony, fully grammatical structures, as well as lower-harmony structures with semantic or syntactic feature clashes. By including ungrammatical states as attractors, the system resembles some previous SOSP models (Kempen & Vosse, 1989; Smith et al., 2018; Tabor & Hutchins, 2004; Vosse & Kempen, 2000, 2009) but not others (beim Graben et al., 2008; Cho et al., 2017; Gerth & beim Graben, 2009; Stevenson, 1994a, 1994b; Stevenson & Merlo, 1997). Selecting a low-harmony structure embodies the assumption that the mind optimizes structures for internal consistency, but it may not always reach a globally optimal state. Including both grammatical and ungrammatical structures includes rational parsing strategies (e.g., Hale, 2011; Levy, 2008a) as special cases, but it also allows the system to explore a broader range of linguistic structures than just those allowed by a symbolic grammar. This property of SOSP-TH will prove crucial in explaining sentence processing phenomena where grammar-controlled theories fail.

Like grammar-controlled theories, SOSP has been shown to account for many sentence processing phenomena, including phenomena that grammar-controlled theories
have difficulty with. SOSP models, usually defined using difference or differential equations, give time a central place in how linguistic structures are built: The mechanisms for assembling structures are defined in terms of how the structures change through time. But with few exceptions (specifically Cho et al., 2018; Smith & Tabor, 2018; Tabor & Hutchins, 2004) they are surprisingly bad at making predictions about how long particular structures take to process. Since timing data (from self-paced reading or eye-tracking while reading) are one of the main sources of data about online sentence processing, this is an embarrassing lacuna for self-organizing theories. The purpose of this dissertation is 1) to develop a generally applicable, mathematical theory of sentence processing from which we can derive precise timing predictions for any given linguistic structure and 2) to demonstrate that it can capture important data points from existing and new experiments.

The remainder of this chapter motivates investigating SOSP models by discussing sentence processing phenomena that are difficult to account for under grammar-controlled theories and concludes with a brief outline of the other chapters.

1.2 Two challenges for grammar-controlled theories

1.2.1 Local coherence effects

Local coherence effects were first reported in Tabor et al. (2004). Using self-paced reading, Tabor et al. compared the sentences in (1):
They found a slowdown at the participle (*tossed or thrown*) in the reduced relative clauses in (1-a)-(1-b) compared to the non-reduced (1-c)-(1-d), but the crucial finding was an additional slowdown for (1-a) over (1-b). *Tossed* is temporarily ambiguous between the grammatical past participle and the locally coherent but globally un-grammatical main verb interpretations. This result suggests competition between the grammatical and ungrammatical parses, causing slowed reading times. Further empirical support for local coherence effects comes from Konieczny (2005), Konieczny et al. (2009), Cai et al. (2012), Paape and Vasishth (2015), Levy et al. (2009), and Bicknell, Levy, and Demberg (2009).

One might argue that this effect is due to confusion at the lexical level about the part of speech of *tossed* (see Kukona, Cho, Magnuson, & Tabor, 2014; Marslen-Wilson, 1975, for similar effects). Assuming this would allow limited self-organization in the lexicon, supporting the traditional modularity arguments (Fodor, 1985; Marslen-Wilson, 1975). This approach predicts a main effect of ambiguity, with slower processing for the reduced and non-reduced conditions with *tossed* compared to *thrown*. Tabor et al. (2004) did not observe such a main effect, suggesting that it is truly syntactic local coherence that is driving reading times and not merely part-of-speech confusion.
Tabor et al. argue that these effects are best explained with self-organization. Incoming words attempt to form attachment links based on how well local features on linked attachment sites match without any “overseer” monitoring the global well-formedness. Because *tossed* is ambiguous, it can attach to nearby treelets in multiple ways such that the local features match well: *Player* is a good feature match for the subject attachment site on *tossed* as a main verb, and *tossed* is a good feature match to be a reduced relative modifier of *player*. Because these parses both have high harmony, the competition between them takes a long time to resolve. This contrasts with the *thrown* condition, where the only high-harmony parse available has *thrown* attaching as the head of a relative clause modifying *player*, so there is little slowdown due to competition from other parses.

As Tabor et al. (2004) discuss, many plausible grammar-controlled theories fail to plausibly account for this effect. Levy (2008a)’s surprisal account fails to predict the effect, as Levy himself concedes (Levy, 2008a, p. 1167). Unless a parser can at least temporarily entertain *the player tossed the Frisbee* as a coherent substring, the effect cannot be explained\(^2\). Some grammar-controlled theories allow just this: Frazier and Fodor (1978), Crocker and Brants (2000), Hale (2011), Bicknell et al. (2009), and Bicknell and Levy (2009) all involve chunking subsequences of the input and assigning them a structure before incorporating them into a representation of the whole sentence. If the structure assigned to the chunks cannot be grammatically incorporated into the existing structure, then the chunk has to be reanalyzed, which slows processing. These approaches could explain the results of Tabor et al. (2004) if

\(^2\)Note that strongly grammar-constrained SOSP models like Stevenson (1994a) are not likely to produce the standard local coherence effect either.
the parser assigns *the player tossed the frisbee* the structure of a main clause and then has to revise that analysis to incorporate that chunk into the rest of the sentence.

These other approaches suffer from a number of drawbacks, though. While the surprisal-based model of Bicknell and Levy (2009) is a promising start to a Bayesian approach to local coherence effects, their model is quite sensitive to the size of the chunks. It only partially reproduces the pattern from Tabor et al. (2004) with three-word chunks and shows quite different patterns for one- and two-word chunks. As Tabor et al. (2004) discuss, it is important to specify how much syntax the chunking mechanism is allowed to assign for a parsing model to be complete in order for a model to be fully specified. Crocker and Brants (2000) treat chunks as unstructured \( n \)-grams, Bicknell et al. (2009) assigns only part of speech tags, and Frazier and Fodor (1978) and Hale (2011) assign hierarchical structures to substrings. Of these, only Hale (2011) has been applied directly to the Tabor et al. (2004) results; without an implementation of the others, it is not clear whether they can reproduce the result. Hale (2011)’s model can derive the interaction; doing so relies on using the particular search heuristic Hale chose (A* search) that is needed in addition to the actual parsing mechanism (a left-corner parser). Thus, while this approach seems reasonable in principle, actual implementations of it face a number of issues.

Another grammar-controlled approach to the Tabor et al. (2004) finding is the noisy-channel extension to the surprisal theory (Futrell & Levy, 2017; Levy, 2008b, 2011; Levy et al., 2009), which has been argued to account for local coherence effects. In the noisy channel approach, the processing time at a word is proportional to its surprisal. But instead of using a veridical representation of the input, the noisy
channel assumes a noisy, uncertain representation for the input: It introduces a probability distribution over likely words given the actual input. The surprisal at each word is then calculated as the negative logarithm of the probability of the word given the noisy representation of the preceding words. The noise in the input is modeled by insertions, deletions, and substitutions of previously perceived words. This type of noise leads to additional structural alternatives being entertained in addition to the ones available on the basis of the noiseless input. Processing costs are incurred when the parser must make a large change to its previous beliefs about the structure of the sentence, for example, by shifting probability away from one word in the preceding input to another one. In the context of eye movements during reading, the noisy channel model predicts that the participants should slow down and increase regressions out of regions containing a word that induces a large shift in the probability distribution over previous words.

Levy et al. (2009) and (Levy, 2008b) argue that the local coherence effects are not due to the parser temporarily activating ungrammatical structures, which is not possible in a grammar-controlled theory like this. Rather, they are due to the parser revising its beliefs about the previous input to make it consistent with some grammatical structure (even if this structure is not consistent with the actual input). To test these predictions, Levy et al. (2009) studied sentences like (2).

(2)  a. The coach smiled at the player tossed the frisbee.
    b. The coach smiled at the player thrown the frisbee.
    c. The coach smiled toward the player tossed the frisbee.
    d. The coach smiled toward the player thrown the frisbee.
In a corpus analysis, Levy (2008b) showed that the reduced relative parse of *tossed* in (2-a) is less frequent and alternative parses involving edits to the previous words are relatively more frequent (e.g., *the coach smiled* and/ *as* the player tossed, *the coach smiled at the player* who tossed). This analysis showed that it is less costly to change beliefs over the previous input than to build the low-frequency, but correct parse, which induces the processing slowdown. Indeed, compared to (2-b), (2-c), or (2-d), (2-a) incurred a significant processing cost in the form of longer go-past reading times, more regressions out, and more regressions to the preposition (*at* or *toward*) (Levy et al., 2009). The (2-b) condition does not incur this cost because *thrown* cannot serve grammatically as a past-tense verb, so there is no possibility for the parser to entertain an alternative parse that would require a revision of belief over previous words. (2-c) and (2-d) do not incur these costs because *toward* does not have any grammatical perceptual neighbors that could receive redistributed probability mass. Under this account, therefore, local coherence effects are caused by revising beliefs about previous material, not competition between grammatical and ungrammatical parses.

In (2), it is clear that a self-organizing approach as proposed by Tabor et al. (2004) would predict a main effect of ambiguity (*tossed* versus *thrown*) but not the observed interaction between that and the *at/toward* manipulation. This is because there is still a local coherence with *toward*, so it should compete with the globally coherent parse either way. In the first-pass reading times (an early measure of processing difficulty), Levy et al. (2009) do indeed report a main effect of ambiguity and no significant interaction. The main effect and lack of interaction in this early measure
suggests that there is at least an early local coherence effect in both conditions after all. It might be that this early effect triggers the parser to then go back and check to see if it indeed had the correct input, as Levy et al. argue, but in any case, this main effect is not predicted by the noisy channel.

Moreover, as Kukona et al. (2014) note, the noisy channel account relies on orthographic and phonological similarity between previously read words and words with a higher probability given the new input. This account fails for an additional type of local coherence effect from visual-world eye-tracking that Kukona et al. (2014) report. Here, participants heard sentences like, *The boy will eat the white cake*, while viewing a screen with pictures of white cake, brown cake, a white car, and a fourth distractor item (e.g., a brown car). Kukona et al. found that participants showed a higher proportion of looks to the white car than to the brown car (although both of these received fewer looks that either cake). Even though the context of eating should minimize looks to either car, hearing *white* caused participants to entertain the contextually inappropriate completion of *white car*.³ In this case, revising the input to make *white car* more likely would involve major changes to the input, e.g., by changing *eat* to *play with*, which involves both an insertion of new word and editing an existing word to a phonologically distant one. Such a large degree of input-editing is not likely to induce a lower cost than just building a structure on the veridical input, which would not explain the Kukona et al. (2014) results. This result, then, seems out of reach of the noisy channel approach and more naturally explained by

³Looks to the white car do not entail that participants are entertaining it as the direct object, but it still suggests that participants are being influenced by the bottom-up information from *white* despite the *eat* context.
bottom-up local coherence as predicted by self-organizing theories.

Thus, grammar-controlled theories, while they successfully explain many sentence processing phenomena, have trouble with local coherence effects.

### 1.2.2 Interference in subject-verb number agreement

Subject-verb number agreement has served as a paradigm case for studying interference in long distance dependencies. One of the most commonly studied interference effects, agreement attraction, arises when a word in a sentence agrees in number\(^4\) with a word other than its typical agreement controller (e.g., Barker et al., 2001; Bock & Miller, 1991; Eberhard et al., 2005; Pearlmutter et al., 1999; Wagers, Lau, & Phillips, 2009). An oft-studied example is when people produce a sentence with a verb with a different number marking from its canonical controller, e.g., *the key to the cabinets are on the table* instead of *the key to the cabinets is on the table* (Bock & Miller, 1991). A strictly grammar-controlled theory cannot explain this well-replicated finding, as no plausible symbolic grammar of English would include structures in which the subject and verb disagree in number. This phenomenon provides compelling evidence that people not only temporarily entertain ungrammatical structures, they sometimes produce them.

There are two prominent extensions to grammar-controlled theories to explain agreement attraction. One, most often applied to production, is the Marking and Morphing model of Eberhard et al. (2005) and Bock, Eberhard, Cutting, Meyer, and Schriefers (2001). Marking and Morphing posits a two-step process for determining agreement attraction has also been studied using gender and case (e.g., Badecker & Kuminiak, 2007; Lorimor, Jackson, & Foote, 2015; Vigliocco & Franck, 1999, 2001)
verb number marking. First, in the Marking stage, semantic properties of the subject noun phrase (NP; the key to the cabinets above) contribute a continuous value to the probability of choosing a plural verb. Subsequently, in the Morphing step, morphosyntactic features of the nouns combine additively with the semantic contribution to determine the final probability of choosing a plural verb. While this mechanism does a good job of accounting for many agreement attraction effects (Eberhard et al., 2005), it constitutes an additional processing mechanism beyond what is independently needed to actually build the structure and is not motivated by other empirical phenomena.

SOSP-TH can explain this effect without recourse to additional processing mechanisms because it includes low-harmony structures like the key to the cabinet(s) are as attractors of the system dynamics. As discussed in Chapters 2 and 3, these lower-harmony structure are less likely to be built, but because the system can make a relatively harmonious structure by attaching cabinets as the subject of are, the model will produce more plural verbs after cabinets than cabinet, as observed in, e.g., Bock and Miller (1991).

A more parsimonious grammar-controlled theory, most commonly applied to comprehension, is the ACT-R model (adaptive control of thought-rational; Anderson et al., 2004) of Lewis and Vasishth (2005). ACT-R takes similarity-based interference in cue-based memory retrieval as a central cause of processing difficulty in sentence processing. In ACT-R, when the parser gets to the verb, it must search in memory for words that match cues provided by the verb. For example, consider the ungrammatical sequences in (3):
(3)  
a. The key to the cabinet are…  
b. The key to the cabinets are…

When the parser arrives at the verb, it uses cues like +NOUN, +PLURAL, and +SUBJECT to search in memory for good feature matches. ACT-R predicts that (3-b) should be faster to process than (3-a) (Jäger, Engelmann, & Vasishth, 2017). This is because both nouns in (3-b) are partial feature matches for the verb’s retrieval cues, which leads to a race to reach a retrieval activation threshold. As demonstrated in Logačev and Vasishth (2015), such a race will produce faster average processing times than cases when the nouns are less well matched. This is an effect of the noise: when both are relatively fast, the only way for one to win is for the noise in its activation to bump it over the finish line first, causing, on average, a speedup. In (3-a), though, there is only one partial match in memory (key), which takes a certain (relatively long) amount of time to retrieve. This prediction is supported by the data summarized in Jäger et al. (2017)’s Bayesian meta-analysis.

On the other hand, ACT-R fails to predict observed effects in grammatical sentences like:

(4)  
a. The key to the cabinet is…  
b. The key to the cabinets is…

The Jäger et al. meta-analysis finds weak evidence that the verb in sentences like (4-a) is processed more quickly than (4-b), whereas ACT-R predicts the opposite effect due to key and cabinet both being partial matches.
In Chapter 5, we will see that SOSP-TH can account for the timing data in grammatical sentences, but it does not capture the pattern in ungrammatical sentences, the opposite of what ACT-R does. SOSP-TH’s predictions seem sensitive to parameter settings, so further simulations will be necessary to determine if this pattern is out of the reach of SOSP-TH under the simplest assumptions.

Another case of subject-verb agreement where grammar-controlled theories have trouble involves encoding interference. Encoding interference is when features of a word that are present when the word is perceived and encoded into memory affect sentence processing, even though they are not relevant for retrieval. As discussed above, ACT-R posits that processing difficulty during language comprehension is due to similarity-based interference during retrieval of words from memory. However, not all interference effects can be explained as retrieval interference. A number of studies have provided evidence that features not relevant for retrieval can interfere with language processing. For example, Gordon, Hendrick, and Johnson (2001) found that the typical processing difficulty in object-extracted structures was attenuated when the two NPs involved (the extracted matrix subject and the embedded subject) were of different types (definite description vs. pronoun or proper name). NP type is not helpful for determining what the correct retrieval target is, making this an instance of encoding interference. Similarly, Hofmeister (2011) found significant reading time slowdowns at the retrieval site when the retrieval target was less syntactically and semantically distinct from competitors. In production, Gennari, Mirković, and MacDonald (2012) found that the semantic similarity between two animate nouns correlated with the rate at which one of them was left out of an answer.
in a relative clause question-answering paradigm. Gennari et al. interpret this as an effect of increased competition between the semantically similar items. Finally, Villata et al. (2018) tested for encoding interference in subject-verb agreement in grammatical English and Italian sentences. They found evidence of interference in comprehension question accuracy, where participants were significantly more likely to correctly answer the question in the encoding feature mismatch conditions than in the encoding feature match condition. There was also a parallel effect in reading times at the verb, with faster reading times in the mismatch condition compared to the match condition. Together, these studies provide evidence that when two items have similar encoding features, they tend to be harder to process than when their encoding features are distinct. ACT-R cannot explain these findings because it simply lacks a way of allowing encoding features to affect processing. As discussed in Chapter 5, SOSP-TH handles these cases by allowing features on noun treelets to affect features on verb treelets, which in turn interferes with how the nouns compete to attach to that verb.

1.3 Roadmap

Here, we have surveyed two classes of phenomena that challenge grammar-controlled theories of sentence processing. Some cases of local coherence effects are explainable in terms of uncertainty about the input (Levy, 2008b; Levy et al., 2009), but other cases (Kukona et al., 2014) can only be explained if the parser can entertain contextually inappropriate analyses. Explaining interference effects in subject-verb number
agreement with grammar-controlled theories required adding additional mechanisms in addition to what is needed to assemble words together into larger structures. SOSP approaches, on the other hand, derive these effects using only structure building. Given the previous success with other sentence processing effects (Cho et al., 2018, 2017; Kempen & Vosse, 1989; Tabor & Hutchins, 2004; Vosse & Kempen, 2000, 2009), capturing the additional data points of local coherence and subject-verb agreement interference warrants further testing of the model and its predictions.

As mentioned above, SOSP models have had very limited success at predicting timing data. Smith and Tabor (2018), included with permission here as Chapter 2, presents a solution to this by providing the formal details of the SOSP-TH model. The major novelty of the model is that the system dynamics and predictions about word-by-word processing times and parse formation rates are derived directly from a harmony function. After providing more detail on the implementation of the model, we show in Chapter 3 that it can account for local coherence effects (Tabor et al., 2004) in comprehension (with some limitations) and the classic agreement attraction production effects discussed above.

Chapter 4 focuses on a case of semantic interference in subject-verb number agreement presented in Smith et al. (2018). Smith et al. (2018) showed that an SOSP parser that uses a different mathematical framework can predict rates of singular and plural number agreement with a class of pseudopartitive subject noun phrases (e.g., a box of oranges, a stack of sandwiches, a lot of newspapers) based on semantic features on the first noun. However, Smith et al. did not analyze the participants’ response times for making their agreement choice, so we analyze that data (Experiment 1) and
present an updated SOSP-TH model that captures the somewhat surprising response
time patterns from the human data.

Chapter 5 reports two new experiments on encoding interference in subject-verb
number agreement. We build on the results of Barker et al. (2001), who showed
increased rates of incorrect plural verb productions for subject NPs like *the canoe
by the sailboats* compared to *the canoe by the cabins*, suggesting that the semantic
similarity between *canoe* and *sailboats* led to increased confusion about the proper
controller of verb number. We explain how SOSP-TH reproduces this effect and
present an implemented SOSP-TH parser that predicts slower self-paced reading
times in the *sailboats* condition than in the *cabins* condition. This is contrasted with
an extension to ACT-R that allows activation spreading between memory chunks
based on encoding similarity (van Maanen & van Rijn, 2007; van Maanen, van Rijn,
& Borst, 2009; van Maanen, van Rijn, & Taatgen, 2012). In contrast to SOSP-TH,
this extension predicts a speedup in reading times at the verb, while still predicting
the Barker et al. production data. We test the diverging predictions in two self-paced
reading experiments (Experiments 2A and 2B), the results of which are largely in
accordance with the results of the SOSP-TH model.

The dissertation concludes with a general discussion, where we note the limitations
of the SOSP framework, discuss it in relation to competing theories, and enumerate
a number of future directions for continued research in this area.
Chapter 2

Smith and Tabor (2018)
To Whom it May Concern:

On behalf of the MathPsych/ICCM 2018 Organizing Committee, I grant permission for Garrett Smith to include the article which submitted to the conference in his dissertation and I assure that there are no permission issues nor will the inclusion of the paper result in any violations of copyright.

Sincerely,

Joseph Houpt
Conference Organizer
MathPsych/ICCM 2018
Toward a Theory of Timing Effects in Self-Organized Sentence Processing

Garrett Smith (garrett.smith@uconn.edu)
Department of Psychological Sciences, 406 Babidge Road, Unit 1020
Storrs, CT 06269 USA

Whitney Tabor (whitney.tabor@uconn.edu)
Department of Psychological Sciences, 406 Babidge Road, Unit 1020
Storrs, CT 06269 USA

Abstract

Many theories of sentence processing are based on the idea that a discrete, symbolic grammar defines all of the structures relevant for parsing, effectively supervising the parser as it selects from those structures the one that best fits the input. However, local coherence effects, where people’s parsing behavior suggests they are entertaining locally viable but globally impossible structures, suggest that this may not always be the case. We introduce a self-organized sentence processing (SOSP) model of local coherence effects and use it to demonstrate how predictions about timing effects (a major source of psycholinguistic data and a shortcoming of many previous dynamical parsers) can be derived directly from a harmony (well-formedness) function covering both grammatical and ungrammatical structures. This framework allows us to simulate the processing of any set of lexical features and attachment links, making it widely applicable to psycholinguistic phenomena.

Keywords: sentence processing, local coherence effects, dynamical systems models, self-organization

Introduction

The current, most fully-developed models of online sentence processing adopt an assumption which may be called grammar supervision. With grammar supervision, a symbolic grammar specifies the universe of structures possible for language comprehension and production, and the parser only considers those grammatical structures. An example is surprisal theory (Hale, 2001; Levy, 2008), in which the parser distributes probability over all grammatical structures compatible with the current input at each word. The processing time for each word is proportional to how much change in the probability distribution is needed after incorporating a new word (the Kullback-Leibler divergence between prior and posterior distributions estimated from a large corpus). This kind of theory has been massively successful in modeling reading times in both experimentally designed stimuli and natural corpora (Levy, 2008; N. J. Smith & Levy, 2013).

However, empirical studies over the past several decades have identified a number of phenomena that challenge the grammar-supervision hypothesis. We focus on local coherence effects (Ex. (1); Bicknell, Levy, & Demberg, 2009; Konieczny, Müller, Hachmann, Schwarzkopf, & Wolfer, 2009; Kukona, Cho, Magnuson, & Tabor, 2014; Levy, Bicknell, Slattery, & Rayner, 2009; Paape & Vasishth, 2015; Tabor, Galantucci, & Richardson, 2004). Early-arriving words make it so that, if the grammar were supervising, only one parse would be possible, but when later words are perceived, people show evidence of entertaining a second, conflicting parse motivated by the later-arriving words. For example, the reduced forms in of Ex. (1) (i.e., without who was) showed slowed reading at tossed/thrown relative to the unreduced form, but this effect was significantly larger for (1-a) than for (1-b) (Tabor et al., 2004).

(1) a. The coach smiled at the player (who was) tossed the Frisbee by the opposing team.
   b. The coach smiled at the player (who was) thrown the Frisbee by the opposing team.

We can make sense of this result if we assume that the words the player tossed . . . (but not thrown) cause the parser to construct an active clause with the player as its subject, even though English grammar mandates that, in this context, tossed be a passive verb heading a reduced relative clause modifying the player. This process is inconsistent with grammar-supervision theories, but it is naturally predicted if parsing is governed by principles of self-organization.1

Self-organized sentence processing (SOSP; Kempen & Vosse, 1989; Stevenson, 1994; Tabor & Hutchins, 2004; van der Velde & de Kamps, 2006; Vosse & Kempen, 2000, 2009; Cho et al., 2017; G. Smith, Franck, & Tabor, 2018; Gerth & beim Graben, 2009)) is an approach to modeling sentence processing which does not assume grammar supervision. Instead, in analogy to many physical chemical and biological processes (see, e.g., Haken, 1983), parses self-organize (without any controller or external supervision) via continuous, local, bottom-up interaction among small pieces of syntactic tree structure (treelets) activated by the words that have been perceived or are being produced. In SOSP, feedback interactions among the treelets generally drive the formation of structure consistent with the grammar, but when two or more incompatible structures receive bottom-up support, the system can stabilize in an ungrammatical state of conflict, causing processing difficulty. Such models have produced plausible accounts of center embedding vs. right branching, garden path effects, lexical ambiguity processing (Vosse & Kempen, 2000), length effects (Tabor & Hutchins, 2004), and agreement attraction (G. Smith et al., 2018), among others.

---

1Levy et al. (2009) argue that surprisal can account for Tabor et al. (2004) with a noisy channel assumption—words may be misperceived (e.g., at was actually and in Ex. (1-a)). Cho, Goldrick, and Smolensky (2017) present a similar approach in a dynamical model. However, not all local coherence effects are plausibly amenable to this explanation (Kukona et al., 2014; Paape & Vasishth, 2015).
Oddly, there are relatively few SOSP results on timing data, even though timing data are the most common kind of psycholinguistic data, and even though self-organization is generally understood via dynamical systems theory, the mathematics of variables interacting in time. Our main contribution here is a novel SOSP framework that addresses this shortcoming by making the relationship between well-formedness and processing times transparent. Influenced by Cho et al. (2017), Smolensky (1986), and Haken (1983), we define a harmony function (also known as a potential or energy function) that specifies the global well-formedness of system states (configurations of features on attachment sites and attachment links, Fig. 1). We employ a systematic method of deriving the harmony function from lexical features in parsed sentences, creating a hilly landscape with peaks corresponding to both fully grammatical structures and conflict states (Fig. 2). The sentence processing dynamics noisily push the system uphill on this landscape to find local harmony maxima. This leads to a theory of timing effects in which, all other things being equal, a higher-harmony parse is built faster than a lower-harmony one. This is because higher peaks have steeper gradients, causing the system to move faster toward the peak. In ambiguous sentences, the system stochastically selects among different peaks, and its path will be more curved if competing peaks are more equally well-formed. Therefore, average processing times over many trials depend on which peaks are selected and how curved the trajectories are.

Below, we present our SOSP framework (called SOSP-TH (“treelet harmony”) to distinguish it from other SOSP models), show how it makes timing predictions, report an implemented SOSP-TH model of local coherence, and finally discuss SOSP-TH in relation to other psycholinguistic theories.

The SOSP-TH framework

In SOSP-TH, linguistic structures are built out of lexically anchored syntactic treelets that connect with each other via graded attachment links (Fig. 1). We assume for simplicity a dependency grammar formalism (e.g., McDonald et al., 2013), so the only attachment sites are ones linking a word as the dependent of another word (head attachment sites) and ones linking other words as dependents (dependent attachment sites). The head and dependent attachment sites are feature vectors encoding syntactic and semantic properties of a word and its expected dependents, respectively. Some features can change (e.g., the determiner the gets its number marking from its licensor), while others are fixed in the lexicon. The only constraints on link formation are that 1) no links can form within a single treelet (e.g., a determiner dependent site on a noun cannot link to the head of that same noun) and 2) links can only form between head attachment sites and dependent attachment sites, i.e., no head-head or dependent-dependent links. All other links, grammatical and ungrammatical, are allowed to form. Finally, a special
Features and links that are fully “on” and “off” are coded as 1 and 0, respectively. In order to allow multiple tokens of the same treelet in one sentence (e.g., _the in the dog saw the cat_), all of a treelet’s dimensions are repeated for every position in a sentence. Thus, there is a set of dimensions corresponding to _the_ as the first word of a sentence, a different set of dimensions for _the_ as the second word, etc. Links (additional dimensions of the system) are between sentence-position-specific instances of treelets.

Not all attachment links make equally well-formed structures, though. Structures in which all linked feature vectors are perfectly matched receive the maximum harmony of 1. Any feature mismatch lowers the harmony for that structure. In this way, SOSP implements a graded notion of well-formedness. We quantify the local harmony for a treelet _i_’s configuration of features and links, using Eq. 1:

\[
h_i = \prod_{l \in \text{links}} \left( 1 - \frac{\text{dist}(f_{\text{head}}^l, f_{\text{dependent}}^l)}{n_{\text{feat}}} \right)
\]

where \( n_{\text{feat}} \) is the number of partial and full parses (harmony peaks) we wish to encode:

\[
H(\mathbf{x}) = \sum_{i} h_i \phi_i(\mathbf{x})
\]

The local harmony of a structure is the product of one minus the normalized Hamming distances between the head feature vectors \( f_{\text{head}}^l \) and dependent feature vectors \( f_{\text{dependent}}^l \) for each link _l_. \( n_{\text{feat}} \) is the number of elements in the feature vectors. This definition of local harmony is valid for any combination of features and links, even those that strongly violate rules of a symbolic grammar, e.g., the fragmentary, locally coherent structure [ _at_ ][ _Root_ [ _S_ _[Subj [Det the] player ]] tossed_]. In the simulations below, we will see that including these lower-harmony structures in the mental representation of possible structures plays a key role in explaining observed timing effects. Eq. 1 allows us to calculate the harmony of any linguistic configuration, but on their own, the \( h_i \)'s do not tell us how to choose a structure given the input. To that end, we define a global harmony function and derive the dynamics from it.

**Defining the harmony landscape and dynamics**

We can define where the peaks in our harmony function are by using a sum of radial basis functions (RBFs) \( \phi_i \) (Han, Sayeh, & Zhang, 1989; Muezzinoglu & Zurada, 2006):

\[
\phi_i(\mathbf{x}) = \exp \left( -\frac{(\mathbf{x} - \mathbf{c}_i)^T(\mathbf{x} - \mathbf{c}_i)}{\gamma} \right)
\]

Here, \( \mathbf{x} \) (a column vector) is the \( d \)-dimensional state of the system encoding values of all features and links in \( \mathbb{R}^d \), each \( \mathbf{c}_i \) is the location of the \( i \)th (partial) parse (encoding desired feature values and link strengths), \( \mathbf{x} \) denotes the vector transpose \(^T\), and \( \gamma \) (a free parameter) sets the width of the RBFs.

We then define the harmony function \( H(\mathbf{x}) \) as the sum of \( n \) RBFs, where \( n \) is the number of partial and full parses (harmony peaks) we wish to encode:

\[
H(\mathbf{x}) = \sum_{i} h_i \phi_i(\mathbf{x})
\]

where the \( h_i \) give the local harmony of a (partial) parse, computed using Eq. 1. This equation creates a hilly harmony landscape analogous to Fig. 2, assigning harmony values both to the \( \mathbf{c}_i \) and to all states intermediate between them.

In SOSP-TH, treelets are interacting subsystems that attempt to assemble themselves through local interactions that locally maximize harmony. Since the gradient of a scalar-valued function like \( H(\mathbf{x}) \) points in the direction of steepest ascent, we make the system change in time so that it follows this gradient uphill in a noisy way:

\[
\frac{d\mathbf{x}}{dt} = \nabla_{\mathbf{x}} H(\mathbf{x}) = -\gamma \sum_{i} h_i (\mathbf{x} - \mathbf{c}_i) \phi_i(\mathbf{x}) + \sqrt{2D} dW
\]

(\( D \) scales the magnitude of the Gaussian noise process \( dW \). For \( D = 0 \), gradient dynamical systems like this simply settle from an initial condition to an attractor (points to which the system will return after a small perturbation; Strogatz, 1994). For \( D > 0 \), the noise helps determine which attractor the system converges on.

Any parsed corpus can be represented as a set of vectors (the \( \mathbf{c}_i \)) of lexical features at particular sentence positions and links between attachment sites, making SOSP-TH a general theory of sentence processing. Note that once the \( \mathbf{c}_i \) are specified, the harmony landscape does not change, unlike in the Gradient Symbolic Computation framework (Cho & Smolensky, 2016; Cho et al., 2017; Cho, Goldrick, Lewis, & Smolensky, 2018), in which the harmony function changes with the input. Since the parsing dynamics are derived directly from the harmony function, the SOSP-TH parser is derived directly from a parsed corpus of sentences. We now show how we can derive processing time predictions from these equations.

**Predicting processing times**

To derive predictions about processing times, we first consider the simplest possible case, a one-dimensional system with a single harmony peak at \( x = 0 \). The harmony function is \( H(x) = h \phi(x) = h \exp \left( -\frac{x^2}{\gamma} \right) \) and the dynamics are given by \( \dot{x} = -\frac{2h}{\gamma} x \phi(x) \). From this equation, we can already see that the higher the harmony of the attractor, the faster system moves toward it: Well-formed structures are faster to build than ill-formed structures.

In general, though, an SOSP-TH parser will have many dimensions coding multiple features and link strengths, and

\[\text{There are other ways to show how settling times in a single trial depend on the harmony of the parse that forms. One is to consider the time } dt \text{ it takes to travel an infinitesimal distance } dx, \text{ } dt = dx / \gamma, \text{ since time equals distance divided by velocity. Integrating both sides shows the settling time } t \approx (2h)^{-1}. \text{ A third option, linear stability analysis (Strogatz, 1994) provides a similar result.}\]
there will be many attractors corresponding to different structural alternatives. To see that higher harmony still means faster processing, we can approximate Eq. 3 near an attractor \( i \) by neglecting all terms \( j \neq i \) in the sum in Eq. 3, as the effect of all other attractors drops off exponentially: 
\[
\dot{x} \approx -\frac{2h_i}{\gamma}(x - c_i)\phi_i(x)
\]
It is clear that the same relation between settling time and harmony holds. However, the effects of other attractors are, in general, not completely negligible. Fig. 3 shows how the presence of a relatively high-harmony competitor can bow trajectories away from an attractor by warping the harmony landscape, even though the system is not in the basin of attraction of the competitor.

Thus, the overall theory of timing effects in SOSP-TH is this: Within a basin of attraction of a structure, the settling time scales approximately inversely proportional to the harmony of that parse, modulo the noise and the bowing. Over repeated trials, noise will bump the system toward attractors of different harmony heights, so the average settling time at a word is the average of the settling times to each selected attractor weighted by how often the attractor is selected. We now illustrate this in a simple model of local coherence.

An SOSP-TH model of local coherence effects

A full model of the incremental processing of the sentences in (1) would involve incrementally turning on features of words in their sentence positions, letting the system settle to an attractor associated with a partial parse, and repeating until the sentence ends (see Fig. 1). We can model the main local coherence finding from Tabor et al. (2004) in a focused way by assuming that the parser has already read up to The coach smiled at the player tossed/thrown... and that it must now choose how to attach player and tossed/thrown. We need only two dimensions, one for the grammatical player-tossed link and one for the locally coherent tossed-Root link. There is thus an attractor at \([1, 0]\) (local harmony \(h_0 = 1.0\)) and one at \([0, 1]\), which will have different sub-maximal harmonies \((h_i)\) depending on whether tossed or thrown has been read (see Fig. 3). Player is a good feature match to be the subject of tossed, and tossed can function as a main verb attaching to the root node, so the attractor at \([0, 1]\) is penalized only for leaving the coach smiled at unattached to the rest of the structure. For thrown, though, \([0, 1]\) is additionally penalized because thrown cannot function as a main verb, so its features do not match Root’s main-verb dependent features. We start the system at \([0, 0]\), not biased toward either attractor.

SOSP-TH predicts that the noise should bump the system toward the grammatical parse in most cases because its high harmony dominates the harmony landscape. When the noise does push the state toward the locally coherent attractor, it will approach it more slowly in the thrown condition than in the tossed condition because of thrown’s especially low harmony. But because this happens so rarely, the average time will be dominated by fast approaches to the grammatical attractor. The locally coherent parse for tossed will be selected more often due to its higher harmony, so it will increase the average settling time more than thrown. There is also more trajectory bowing for tossed, which also slows processing (Fig 3). Thus, a relatively high-harmony competitor for the grammatical parse will, on average, cause a competition-based slowdown.

We simulated both conditions 2000 times using Euler forward discretization with a time step of 0.01, \(D = 0.001\), and \(\gamma = 0.25\). The system ran until it got within a small radius of an attractor. The local harmony \(h_0\) of the locally coherent attractor \([0, 1]\) was set to 0.8 in the tossed condition, and in the thrown condition to 0.5. As predicted, the system settled to the ungrammatical attractor in both cases, and it did so more frequently in the tossed condition (about 14% of runs) than in the thrown condition (<1% of runs). This increased the average settling time for tossed \((M = 159.073\) time steps, \(SD = 27.692\)) more than for thrown \((M = 149.794, SD = 24.698)\), modeling Tabor et al. (2004)’s effect.

These simulations show local coherence effects for one parameter setting, but Fig. 4 shows how the same pattern holds...
over a wide range of parameter settings. Where it does not hold, there is possibly empirical evidence for a phenomenon that corresponds to the model, different from local coherence. Fig. 4 shows mean settling times as a function of the harmony \( h_1 \) of the ungrammatical parse. We used \( \gamma = 0.25 \) here, but the pattern holds for a wide range of \( \gamma \) values. This figure shows that we will observe local coherence effects as long as \( 0 < h_{\text{thrown}} < h_{\text{tossed}} < 0.85 \). This predicts that local coherence effects should be widespread, a result supported by a large-scale eye-tracking corpus study (Bicknell et al., 2009).

For \( h_1 \) greater than about 0.85, the pattern changes: As the ungrammatical parse increases in harmony, the time its settling time approaches that of the grammatical parse, so it no longer pushes the overall average settling time up as much and the average settling time starts to drop (Fig. 4, bottom panel). The competition still causes a slowdown, but not as strongly as for somewhat lower-harmony competitors. Thus, the model predicts the strongest competition-induced slowdowns when the competing structure is of moderate harmony and smaller-magnitude slowdowns for both very low harmony competitors and (to a lesser extent) higher harmony competitors. This is, to our knowledge, unique among models of sentence processing. We speculate that this property of SOSP-TH might provide a new explanation for ambiguity advantage effects (e.g., Traxler, Pickering, & Clifton, 1998), where certain ambiguous relative clause and adjunct attachments are read more quickly than comparable unambiguous structures. If the harmonies of the two competing parses are close to 1.0 in the ambiguous condition but one is appreciably less than 1.0 in the unambiguous conditions, the competition-based SOSP-TH might be able to explain this puzzling effect that has been argued to rule out competition-based theories.

**Discussion**

In this paper, we presented a theory of timing effects in a self-organizing sentence processing (SOSP) framework, demonstrated how it can explain local coherence effects, and speculated on a possible new approach to ambiguity advantage effects. In our SOSP-TH framework, the amount of time it takes to build a structure depends on how well-formed the structure is, and the average structure-building time over many trials is the weighted average of settling times to each parse chosen.\(^6\)

The local coherence model highlights the crucial role that lower-harmony structures play in SOSP-TH: A relatively well-formed but ungrammatical competitor slows processing more than a very ill-formed competitor because the higher-harmony competitor is built more often. This account differs from the grammar-supervised noisy channel approach to local coherence (Levy et al., 2009), which explains some (but not all; Kukona et al., 2014; Paape & Vasishth, 2015) local coherence effects by allowing the parser to edit its input to preserve grammaticality. By comparison, ACT-R for sentence processing (Lewis & Vasishth, 2005) might be thought of as partially grammar-supervised: Ungrammatical structures can affect processing via noisy memory retrieval that sometimes retrieves incorrect structures, but the cues used for retrieval are set by the grammar, preventing it from explaining local coherence effects via incorrect retrieval. By allowing both grammatical and ungrammatical structures to always influence processing, SOSP-TH occupies a unique and parsimonious place among theories of sentence processing.

**Acknowledgments**

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

**References**


\(^6\) This is similar to recent cue-based retrieval approaches (e.g., Lewis & Vasishth, 2005) that model reading times with statistical hierarchical mixture models (e.g., Nicenboim & Vasishth, 2018).


Chapter 3

SOSP-TH details

3.1 Introduction

This chapter expands on a number of topics outlined in Chapter 2 (Smith & Tabor, 2018). We first provide more detail on the linguistic representations and the design choices that went into the implementation. We then discuss additional mathematical properties of the model, showing that the system is globally stable. Then, we present two SOSP-TH models, a simple, one-word model of classical agreement attraction and an incremental, word-by-word model of local coherence effects. These models highlight a number of important issues in implementing and extending SOSP-TH models. The chapter provides both additional details for the framework as described in Smith and Tabor (2018)/Chapter 2 and demonstrates its generality as a theory of timing and parse-formation effects in sentence processing.
3.2 Linguistic representations

The goal of an SOSP-TH parser is to build the most well-formed structure it can by locally maximizing structural harmony. The trajectories the system takes, which determine the predictions it makes about human language processing, therefore strongly depend on the types of linguistic representations used. However, any grammar that generates structures that are representable as points in a vector space can be used. This section discusses the motivations for results of using the linguistic representations we chose, but other choices might be necessary for processing other phenomena. SOSP-TH is flexible enough to accommodate them, though.

To begin building a parser, we first start with a specification of the grammar it should implement. As the present SOSP-TH parser is not a learning model, this grammar reflects the grammatical knowledge a competent speaker has accumulated and is kept fixed. We use a lexicalist dependency grammar that specifies dependency relations between governor words and zero or more dependent words, (required complements and optional modifiers; Gaifman, 1965; Hays, 1964; McDonald et al., 2013; Müller, 2015). We most closely follow the formalism of Nasr (1995), in which

\footnote{Hays (1964) argues that a generative dependency grammar needs a central element or root node to anchor the whole dependency structure of a sequence of words. A central element requires no governor. This is common in self-organizing models (e.g., the root nodes in Tabor and Hutchins (2004) and Vosse and Kempen (2000)). In SOSP-TH, we take a different approach. All finite verbs in the lexicon are specified as not requiring a governor, making them all possible central elements if they appear in a structure. If more than one finite verb appears in a sequence, which one becomes the central element depends on local competitions via local constraints on individual treelets with one of the verbs attaching as the dependent of another treelet. Thus, only one verb ends up without a governor, satisfying dependency grammar’s requirement that there be only one central element in a sentence. In, e.g., Vosse and Kempen (2000) and Tabor and Hutchins (2004), an explicit root node to which a central element can attach serves only to ensure that there is a single central element. In most cases, local competitions in SOSP-TH will have a single central element because such structures are most well-formed. But including a root node excludes many fragmentary parses.}
...smiled at the player tossed...

a lot of newspapers are...

...canoe by the kayaks is...

Figure 3.1: Example dependency grammar trees. Arrows point from the governor to its dependent. “Subj” is the subject dependent and the head of the subject NP; “PPmod” is a prepositional phrase modifier; “Ncomp” is nominal complement dependent; “Det” is a determiner dependent. Head and dependent feature vectors not shown. Top: The globally coherent parse of a locally coherent sentence. Middle: The structure of a pseudopartitive subject NP and verb (see Chapter 3). Bottom: Grammatical structure for classical agreement attraction and encoding interference effects (see examples below and in Chapter 4).

Each attachment site is a vector of features specifying properties of the word when it is the dependent of another word (head features) or the expected features of each of its dependents (dependent features). We refer to lexical items along with their head and dependent feature vectors as treelets. Sample dependency structures are shown in Figure 3.1.

Each treelet specifies which of its attachment sites must have a link. All words except for main verbs require a governor, i.e., a link from their head attachment site to the dependent attachment site of another word. In addition, words may have required that might be important in explaining timing effects, e.g., local coherence effects, which seem to require a fragmentary structure with tossed as its own central element in addition to smiled.
dependents, e.g., a direct object dependent for a transitive verb. For a sequence of words to be well-formed, each word must have all of its required attachment sites attached to exactly one word. Thus, for a sequence of \( w \) words, the maximum number of dependency links is \( w - 1 \). We call any structure with less than \( w - 1 \) links a \textit{fragmentary structure}, because words or phrases are left “floating,” broken off from the rest of the structure. Structures can also be incomplete or partial, that is, having at least one word with at least one required link missing. Note that partial parses can be, but need not be, fragmentary, and all fragmentary structures contain at least one partial structure.

This linguistic formalism was chosen for a number of reasons. First, the use of lexically anchored treelets made it clear what features to turn on when a word is perceived or produced. Also, such treelets obviate the need to project abstract (non-lexically anchored) structures. This would complicate the word-by-word processing algorithm by forcing us to decide which features they should have and how and when they should be turned on. Finally, the linguistic structures just described can easily be represented as points in a high-dimensional vector space, and so they can simply be plugged in to the equations of SOSP-TH to serve as centers.

As discussed below, though, the system is highly sensitive to how the linguistic representations are structured: Certain structures can actually prevent the system from capturing important effects. While the structures under consideration in this dissertation can be successfully modeled in the chosen dependency grammar framework, more work is needed to determine if it can be made to work for other structures. SOSP-TH can be used to process any structure that can be represented as a vector
of features, so it provides desirable flexibility in searching for the best combination of processing parameters and linguistic representations.

3.3 Design and implementation choices

Given this framework for linguistic structures, we can now set up an SOSP-TH model. The linguistic structures just described, including partial and fragmentary structures, will be represented as high-dimensional vectors that form the centers of the RBFs (radial basis functions). In addition to specifying the centers, we must also make a number of other choices.

First, the local harmony $h_i$ of each center is calculated according to Equation (1) in Chapter 2, unless the structure is fragmentary or partial. In those cases, the harmony calculated according to Equation (1) in Chapter 2 is multiplied by a penalty parameter $p$ (a free parameter) for each required attachment that is not present in the structure. For example, if a partial structure has two missing required links, the harmony that is calculated based on the links it does have is multiplied by $p$ twice to get the final $h_i$ for that center.$^2$

The other two free parameters, $\gamma$ and $D$ must also be set. The variance $D$ of the Gaussian white noise added to the system is set, after pilot simulations, to a level that allows the system to find multiple attractors, but small enough to ensure that it eventually settles into one attractor. For most simulations, $D$ was set to 0.005. The parameter $\gamma$ is the width parameter of the RBFs. The number of distinct

---

$^2$This could also be implemented by having otherwise unattached attachment sites link to a null treelet. Links to the null treelet would be specified to be of quite low harmony.
maxima of the global harmony function depends on width parameter \( \gamma \) and the local harmony \( h_i \) of the peaks (Muezzinoglu & Zurada, 2006). Very low harmony centers can be absorbed into neighboring peaks, especially when \( \gamma \) is high, so the number of attractors of the dynamics is often less than the number of centers that go into the equations. Similarly, setting \( \gamma \) to be too large causes even high-harmony centers to merge into a single harmony peak intermediate between two or more centers.\(^3\) Muezzinoglu and Zurada (2006) prove that setting \( \gamma \) to less than half of the Euclidean distance between the two nearest centers will ensure that the two nearest centers have separate harmony peaks. For more than two centers, this does not guarantee that all centers will be separate, so Muezzinoglu and Zurada recommend (on the basis of simulation studies) multiplying the \( \gamma \) calculated in this way by 0.8 to ensure that all equal-harmony centers have separate peaks. In the simulations for this dissertation, I set \( \gamma = 0.4 \), which follows Muezzinoglu and Zurada (2006)'s heuristic: Since all of our centers are at corners of the unit hypercube, the shortest distance between a pair of centers is 1.0, so \((1.0 / 2) * 0.8 = 0.4\). Since many of the \( h_i \) are less than 1.0, this does not guarantee that all centers will be attractors of the dynamics (i.e., local maxima of the global harmony function): Low-harmony centers may merge with other centers. They can still affect processing, though, by creating relatively flat spots in the harmony landscape that slow processing by deflecting trajectories away from more direct paths to an attractor (see Figure 3 in Chapter 2), an effect we

\(^3\)Setting \( \gamma \) to be large also displaces the actual attractors of the dynamics away from the centers (Ciocoiu, 1996; Muezzinoglu & Zurada, 2006). Thus, the locations of the actual attractors can be found using numerical optimization. For all of the simulations, I used the L-BFGS-B algorithm to do this using the SciPy implementation (Byrd, Lu, Nocedal, & Zhu, 1995; Jones, Oliphant, Peterson, & others, 2001–).
term soft deflection. Note that soft deflection can also occur when there are separate attractors. The presence of any center with non-zero harmony will always distort the harmony landscape and affect trajectories that pass near them.

Most of the simulations in this dissertation model reading or processing times at a single word: The model is initialized at some point that represents a configuration of treelets and links and is far from an attractor, and it settles from there under the gradient dynamics and the noise. The initial conditions used here are typically equidistant between two or more attractors. In the noiseless system, it takes an infinite amount of time for the system’s state to reach the attractor exactly, so we make the assumption that if the system is close enough to an attractor, it has built the structure corresponding to that attractor. In the simulations presented here, the system settles until it gets close enough to the attractor as measured by the Chebyshev distance. The Chebyshev distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ is the largest absolute difference on any dimension, i.e.:

$$\text{dist}_C(\mathbf{u}, \mathbf{v}) = \max_i |u_i - v_i|$$

Thus, the system processes a word until it is within some tolerance (a free parameter) of the attractor on all dimensions. We have also experimented with using the Euclidean distance to the attractor and a velocity stopping criteria (i.e., processing until the system slows down sufficiently, as dynamical systems slow as they approach attractors). The velocity stopping criterion seems particularly interesting for future explorations, as it does not require the system to have knowledge of the locations of its own attractors, an assumption that does not seem cognitively or neurobiologically
Table 3.1: Free parameters and typical values.

<table>
<thead>
<tr>
<th>Free parameter</th>
<th>Typical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF width $\gamma$</td>
<td>0.4</td>
</tr>
<tr>
<td>Noise magnitude $D$</td>
<td>0.005</td>
</tr>
<tr>
<td>Missing link penalty $p$</td>
<td>0.8</td>
</tr>
<tr>
<td>Attractor proximity tolerance</td>
<td>0.1</td>
</tr>
</tbody>
</table>

plausible. Nevertheless, these other systems seem to behave similarly to the Chebyshev stopping criterion, so we use it for simplicity. The tolerance and the other free parameters are listed in Table 3.1.

The SOSP-TH parser is implemented in the free, general purpose programming language, Python. In order to make the implementation fully transparent and replicable, only open-source libraries are used, and all code and data are hosted on GitHub (https://www.github.com/garrett-m-smith/sosp).

3.4 Mathematical properties

In the SOSP-TH approach to sentence processing, structure formation can be thought of in different ways. We often emphasize the view of self-organization, with treelets interacting with each other locally as they coalesce into larger structures, akin to protein folding or the formation of clusters of atoms (Wales, 2003). SOSP-TH’s approach can also be viewed as pattern recognition: Linguistic structures are stored as patterns (attractors) in an associative memory. Online sentence processing is then

---

4 Yet another option would be to use a dynamic stopping criterion as in Green and Mitchell (2006); McRae, Spivey-Knowlton, and Tanenhaus (1998), where processing continues until the system reaches some harmony threshold that decreases with time, allowing the system to settle for a worse structure if processing takes too long.
modeled as an iterative pattern recognition process using the associative memory: Each new word tends to move the system progressively towards a pattern associated with a structure for the whole sequence of words. The equations of SOSP-TH build on attractor neural networks, pattern recognition systems, and associative memory models (Athinarayanan, 1998; Ciocoiu, 1996, 2009; Cohen, 1992; Han, Sayeh, & Zhang, 1989; Li, Ramanathan, Ning, Shi, & Wen, 2015; Muezzinoglu & Zurada, 2006; Sayeh & Athinarayanan, 1994; Sayeh, Athinarayanan, & Zargham, 1994). SOSP-TH follows most closely the design of Muezzinoglu and Zurada (2006), which uses radial basis functions (RBFs) to define energy or harmony maxima at each stored pattern (the proposals of Han et al. (1989) and Ciocoiu (1996, 2009) are very similar, differing only in that they switch the sign of the harmony function so that patterns correspond to minima). For each pattern to be stored, the harmony function includes an RBF with a center corresponding to that pattern, making a peak at that location in state space. The height of the peak in this setup has a natural interpretation: the higher the peak, the higher the local harmony of the structure at that location, and, as we showed in Chapter 2, the faster the system approaches that peak. Pattern recognition consists of initializing the system at some point in its state space and allowing it to settle to a known pattern under the dynamics defined by the harmony function and the noise.

SOSP-TH is specified as a set of nonlinear stochastic\(^5\) differential equations. For such systems, there is no general way of finding analytical solutions to them, i.e.,

---
\(^5\)Without noise, the system can only ever take one trajectory given a particular initial condition. The noise is included to allow the system to take different trajectories and approach different attractors than it otherwise would given a particular initial condition.
explicit equations describing the exact trajectories a system will take given some initial condition (Strogatz, 1994). Thus, we strive for a more qualitative understanding of the system, asking what the long-term behaviors of the system are. We focus here on showing that the system is stable: For any initial condition in the state space, the system will converge to one of the points corresponding to the centers of the RBFs (in the absence of noise). This is important because it ensures that the system will only end up in states that are interpretable as more or less well-formed linguistic structures. This contrasts with, e.g., Hopfield neural networks (Hopfield, 1982), which, while mathematically well-understood and trivial to train (Hertz, Krogh, & Palmer, 1991), contain attractors that correspond neither to training patterns nor to theoretically interesting patterns like the low-harmony structures used in SOSP-TH\footnote{Hopfield networks trained using the Hebbian covariance rule with binary patterns contain spurious attractors that simply flip the bit on every dimension of a stored pattern. For example, if 10010 is an attractor, so is 01101 (Hertz et al., 1991; Hopfield, 1982; Hopfield, Feinstein, & Palmer, 1983). It is not clear that such attractors, interpreted as linguistic structures, would make any interesting predictions or account for existing data.}.

I now show that a noiseless SOSP-TH model will eventually converge to one of its attractors (and not oscillate or wander off to infinity, for example). Here, we repeat the equations of the system, following Muezzinoglu and Zurada (2006). The global harmony function is a sum of Gaussian RBFs with centers that correspond to partial, fragmentary, or complete parses. Each partial parse $i$ has its own RBF centered at $c_i$:

$$\phi_i(x) = \exp \left(-\frac{\|x - c_i\|^2}{\gamma}\right) = \exp \left(-\frac{(x - c_i)^T(x - c_i)}{\gamma}\right)$$

$\| \cdot \|^2$ denotes the squared L2-norm, which is equivalent to the squared Euclidean
norm or simply the dot product of the vector \( x - c_i \) with itself. As Han et al. (1989) note, any norm could be used here, but the L2-norm has a simple interpretation: The harmony of a configuration of links and features drops off exponentially with the distance from a center. The global harmony \( H(x) \) is just the sum over all \( n \) RBFs, each weighted by its local harmony \( h_i \):

\[
H(x) = \sum_{i}^{n} h_i \phi_i(x)
\]  

(3.1)

This \( H(x) \) defines a harmony landscape over the state space, associating with each point a harmony value. The dynamics of the system, i.e., the change of state over time, is given by the gradient (vector derivative) of \( H(x) \):

\[
\dot{x} = \nabla H(x) = -\frac{2}{\gamma} \sum_{i}^{n} h_i (x - c_i) \phi_i(x)
\]  

(3.2)

We can now ask whether the system is stable or not. To do that, we want to show that the harmony function \( H(x) \) is a Lyapunov function for the dynamical system in Eq. 3.2 (Guckenheimer & Holmes, 1983; Hirsch & Smale, 1974; Hirsch, Smale, & Devaney, 2004; Strogatz, 1994; Wiggins, 2003). A Lyapunov function is scalar-valued, continuously differentiable function \( V(x) \) defined in the neighborhood of a fixed point \( x^* \) (a point where \( \dot{x} = 0 \)) with the following properties:

1. \( V(x) > 0, \forall x \neq x^* \) and \( V(x^*) = 0 \), i.e., \( V \) is positive definite

2. \( \dot{V} = \frac{dV}{dt} < 0, \forall x \neq x^* : \) \( V(x) \) decreases along trajectories until the system reaches \( x^* \), where \( \dot{V}(x^*) = V(x^*) = 0 \).
Intuitively, we can think of a Lyapunov function as a generalized potential energy function. For example, in physical systems like a ball thrown vertically into the air, the ball’s movement minimizes its potential energy that arises due to gravity, causing it to fall back to the ground. Minimizing such a gravitational energy function completely defines the dynamics of the system. Thus, if we can find a Lyapunov function for some fixed point $x^*$, we can make concrete statements about trajectories near $x^*$ without explicitly solving the system. If conditions (1) and (2) hold, then the fixed point $x^*$ is asymptotically stable. A system is asymptotically stable if it starts out near enough to $x^*$, then it will tend to $x^*$ as $t \to \infty$ without straying beyond a fixed distance away from $x^*$ (Strogatz, 1994).\(^7\)

We define $G(x) = -H(x)$ to preserve the sign conventions in the above conditions. It is simple to show that criterion (2) is met for $G(x)$ above, i.e., that $G(x)$ decreases along trajectories of $\dot{x}$. By the chain rule,

$$\frac{d}{dt}G(x) = \nabla G(x) \cdot \dot{x} = \nabla G(x) \cdot (-\nabla G(x)) = -\|\nabla G(x)\|^2 \leq 0 \quad (3.3)$$

where $\| \cdot \|_2$ is the Euclidean norm as before. Thus, the system tends to decrease the negative harmony over parses in time, or, equivalently, increase the harmony, fulfilling condition (2).

Now, to complete the proof that a particular fixed point $x^*$ is asymptotically stable, we show that a Lyapunov function defined in a region around it is positive definite.

\(^7\)In any complex harmony or energy landscape, there will be many $x^*$ where $\dot{x} = 0$ (Maxwell, 1870). However, conditions (1) and (2) will not be met for them, so they are not asymptotically stable. Mostly, these will be unstable, so if the noise bumps the system away from one of them, it will not return.
(criterion (1)). For any fixed point \( x^* \), the we define the function \( \tilde{G}(x) = G(x) - G(x^*) \) (Wiggins, 2003). In the neighborhood around \( x^* \), the value of \( G(y) \) is higher for all \( y \neq x^* \) than for \( G(x^*) \). In general, harmony peaks are of different heights, so using \( \tilde{G}(x) \), with its subtraction term, as the Lyapunov function for \( x^* \) ensures that the Lyapunov function is zero at \( x^* \) and greater than zero everywhere else, satisfying condition (1). Since the criteria (1) and (2) above are met for any fixed point \( x^* \), we can conclude that all \( x^* \) are asymptotically stable.

This shows that individual fixed points \( x^* \) are asymptotically stable, but if all of the fixed points of a gradient dynamical system\(^8\) are isolated, we can conclude that the trajectory starting from any initial condition will converge to some \( x^* \), not oscillating or exhibiting any more exotic dynamical behavior (Hirsch et al., 2004, pp. 206–207). All of the fixed points \( x^* \) are isolated: If they were not isolated, they would have at least one zero eigenvalue (Strogatz, 1994), but because \( \tilde{G}(x) \) is positive definite, it can only have strictly positive eigenvalues, ruling out non-isolated fixed points. Thus, SOSP-TH systems will eventually stabilize on some well- or ill-formed linguistic structure.

For the stochastic system, the probability distribution over states of the system will be concentrated near the maxima of \( H(x) \) as the time \( t \to \infty \). We can show this using the stationary solution to the Fokker-Planck equation, a deterministic partial differential equation that describes the change in the probability distribution over states in time (Gardiner, 1985; Haken, 1983; Risken, 1989). For gradient dynamical

\(^8\)A gradient dynamical system is one for which the dynamics are given by the gradient of a scalar-valued function (Strogatz, 1994), i.e., \( \dot{x} = -\nabla V(x) \). Gradient dynamical systems can only have fixed point attractors; other dynamical behaviors like limit cycles, homo- and heteroclinic cycles, and chaos (Hirsch et al., 2004).
systems like SOSP-TH, the stationary probability distribution, i.e., the probability
distribution over states after the system has equilibrated in the long-time limit, is
given by Equation 3.4:

\[ P(x) = N e^{-\frac{H(x)}{D}} \]  \hspace{1cm} (3.4)

\(N\) is a normalization constant that ensures that the probability density \(P\) sums to
one, and \(D\) is the variance of the noise. As long as the noise \(D\) is small enough,
Equation 3.4 will have local probability maxima in the same places as \(H(x)\). If the
noise is too large, the probability peaks can merge, similar to what happens with
\(H(x)\) if \(\gamma\) is too large. This probability density function is only valid in the limit of
infinite time, and thus does not tell us the probability of forming a particular parse
given some initial condition.

### 3.5 Example models

Now that the SOSP-TH framework is fully described, we present two demonstrations
that the system can capture important sentence processing effects that we (in Chapter
1) and others have argued provide evidence for self-organizing theories: a single-word
model of agreement attraction and a word-by-word, incremental parsing model of
local coherence effects.

#### 3.5.1 Classic agreement attraction

As discussed in Chapter 1, classical agreement attraction occurs when a verb is
produced that agrees in number with a noun other than the subject of the clause,
e.g., the key to the cabinets are… (Bock & Miller, 1991). The correct subject, the first noun (N1), is singular, but the verb is plural, agreeing with the second noun (N2). This effect is often tested using a 2 (N1 number) x 2 (N2 number) design in which participants are provided with a subject NP to repeat the subject noun phrase and then complete the sentence. Typically, participants make agreement attraction errors in the N1[sg]-N2[pl] condition about 5-10% of the time and less frequently in the N1[pl]-N2[sg] condition (e.g., Barker et al., 2001; Bock & Miller, 1991; Vigliocco, Butterworth, & Garrett, 1996). As discussed in Chapter 1, this is an example of grammar-flouting behavior that should be easily handled by a self-organizing theory.

To test whether SOSP-TH can reproduce this typical pattern, we set up a simple model with eight attractors, one for each combination of singular and plural marking on the N1, the N2, and the to-be-produced verb. The system had three dimensions, with each one coding the number marking on one of the constituents (0 = singular, 1 = plural). Only one feature was considered: the number marking, so a number mismatch (according to Equation 1 of Chapter 2) would mean an $h_i$ of 0.0. To allow such cases to still influence processing, we made their local harmonies equal to 0.01 instead of 0.0. This approximates a more complete model in which the N2-headed parse has more features, making its feature match to the verb non-zero. The harmonies were chosen so that, if there was more than one way of having a given configuration of number markings, the one with the highest harmony was chosen under the assumption that the system would rarely choose a lower-harmony way of making a configuration. For example, the ungrammatical configuration N1[sg]-N2[pl]-V[pl] can be made by attaching either N1 or the N2 to the verb. Attaching the N2 as the subject has higher
Table 3.2: Centers and harmonies for the agreement attraction simulations.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Coordinates</th>
<th>Initial condition</th>
<th>$h_i$ (plural not marked)</th>
<th>$h_i$ (plural marked)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1[sg] N2[sg] V[sg]</td>
<td>(0, 0, 0)</td>
<td>(0, 0, 0.5)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>N1[sg] N2[sg] V[pl]</td>
<td>(0, 0, 0)</td>
<td>(0, 0, 0.5)</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>N1[sg] N2[pl] V[sg]</td>
<td>(0, 0, 0)</td>
<td>(0, 1, 0.5)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>N1[sg] N2[pl] V[pl]</td>
<td>(0, 0, 0)</td>
<td>(0, 1, 0.5)</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>N1[pl] N2[sg] V[sg]</td>
<td>(0, 0, 0)</td>
<td>(1, 0, 0.5)</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>N1[pl] N2[sg] V[pl]</td>
<td>(0, 0, 0)</td>
<td>(1, 0, 0.5)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>N1[pl] N2[pl] V[sg]</td>
<td>(0, 0, 0)</td>
<td>(1, 1, 0.5)</td>
<td>0.01</td>
<td>0.005</td>
</tr>
<tr>
<td>N1[pl] N2[pl] V[pl]</td>
<td>(0, 0, 0)</td>
<td>(1, 1, 0.5)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

harmony, as the missing link penalty was 0.8, while the other attachment had an $h_i$ of 0.01, so the former was used as the local harmony for that configuration. Each condition was started from a different initial condition that had the number markings on the N1 and N2 turned on and the verb’s number set halfway between its two possible markings. The centers, the initial conditions, and the $h_i$ are provided in Table 3.2. (The last column of Table 3.2 will be discussed shortly). In order for the model to explore more of the state space, the noise magnitude $D$ was increased from its default 0.005 to 0.0075. Each condition was run 2000 times.

With these settings, the system produced the configurations shown in Table 3.3 (in the “plural not marked” columns). The model produced approximately equal amounts of agreement attraction for both N1[sg]-N2[pl] and N1[pl]-N2[sg] ($\chi^2(1) = 0.134, p = .71$), a clear agreement attraction effect. It is able to produce this result because, in the mismatch conditions, there are relatively high-harmony alternatives that the system can sometimes choose in addition to the perfectly grammatical options, whereas in the match conditions, the grammatical attractor is the only
nearby option.

Under the current settings, these two conditions should produce the same distribution of correct and agreement-attraction parses simply due to the symmetry in the harmony relationships between the parses in both cases. However, this fact contrasts with the common pattern in human experiments where agreement attraction occurs mainly in the N1[sg]-N2[pl] condition. It has been argued that this asymmetry is due to singulars being the unmarked, default form and plurals being the marked number form (Badecker & Kuminiak, 2007; Bock & Miller, 1991; Eberhard, 1997; Farkas & de Swart, 2010; Franck, 2015). The marked form is somehow more salient to participants, making agreement attraction errors more likely in the N1[sg]-N2[pl] condition because the singular marking on the N1 can be overshadowed by the plural marking on the N2. Haspelmath (2006) argues that markedness in linguistic theory can, in most cases, be reduced to frequency asymmetries: Marked forms are simply less frequent. Indeed, in English, French, Dutch, Sanskrit, Latin, and Russian, singular nouns are much more frequent than plural nouns, approximately three singulars for every plural (Greenberg, 1966, cited in Tiersma, 1982; Gimenes, Brysbaert, & New, 2016). So, participants might be more likely to notice the plural forms because they are rarer.9

To account for the asymmetry in the human data using SOSP-TH, we considered

Another way of implementing markedness would be to leave the $h_i$ alone and use different values of $\gamma$ for the different conditions. If $\gamma$ were set to a larger value for the V[pl] attractors, this would have two effects. First, it would make it more likely that the system would choose the plural verb by making its basin of attraction larger than the V[sg] attractor’s. It would also slow approaches to plural-verb more slowly, possibly relating to findings such as Wagers et al. (2009) who found that plurals in general are processed more slowly than singulars. The theoretical motivation for this still needs development, though, so for now, we assume that markedness affects the local harmonies via frequency asymmetries and stick with the simple, but maybe ultimately incorrect, assumption of equal $\gamma$s for all centers.

---

9Another way of implementing markedness would be to leave the $h_i$ alone and use different values of $\gamma$ for the different conditions. If $\gamma$ were set to a larger value for the V[pl] attractors, this would have two effects. First, it would make it more likely that the system would choose the plural verb by making its basin of attraction larger than the V[sg] attractor’s. It would also slow approaches to plural-verb more slowly, possibly relating to findings such as Wagers et al. (2009) who found that plurals in general are processed more slowly than singulars. The theoretical motivation for this still needs development, though, so for now, we assume that markedness affects the local harmonies via frequency asymmetries and stick with the simple, but maybe ultimately incorrect, assumption of equal $\gamma$s for all centers.
a hypothetical learning version of the model (see the discussion in Chapter 6). We assume that the learning model develops head and dependent feature representations using large text corpora. One example of such an approach is to use modern neural networking techniques (Mikolov, Sutskever, Chen, Greg, & Dean, 2013) to learn dependency-specific feature vectors (Bansal, 2015; Levy & Goldberg, 2014; Zhao, Huang, Dai, Zhang, & Chen, 2014). These learning techniques make the system sensitive to the frequencies of structures in the corpus, and therefore the local harmony values will reflect the effects of frequency. Haskell, Thornton, and MacDonald (2010) found no occurrences of N1[pl]-N2[sg]-V[sg] in the Brown corpus, while N1[sg]-N2[pl]-V[pl] occurred in over twenty percent of cases. Thus, while N1[sg]-N2[pl]-V[pl] will have somewhat lower harmony (because singular N1s are most often paired with singular verbs), it will have a higher harmony than N1[pl]-N2[sg]-V[sg]. As a rough approximation of this learning-induced asymmetry, we penalized the $h_i$s of the N1[pl]-N2[sg]-V[sg] and N1[pl]-N2[pl]-V[sg] centers (which also never occurred in the Haskell et al. corpus study) by dividing their normal $h_i$s by two. This effectively punishes the lower-frequency structures, makes them slower to build, and less likely to form. When we repeated the simulations with the new $h_i$, the system produced the asymmetry observed in the human data (Table 3.3, “plural marked” column). A chi-squared test showed that the parse distributions were not independent of attraction condition ($\chi^2(1) = 29.212, p < .001$). This amount of penalty was chosen for illustration purposes, but any asymmetry in the harmonies will have the same type of effect.

Thus, SOSP-TH is capable of handling a well-known and often-replicated finding from sentence production. It does so by allowing ungrammatical structures to form via
Table 3.3: Results of the SOSP-TH agreement attraction simulations, with 2000 runs per condition. In the conditions where the N1 and N2 number markings differ, the incorrect parse was always the associated agreement attraction center.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Plural not marked</th>
<th>Plural marked</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td>N1[sg] N2[sg]</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>N1[sg] N2[pl]</td>
<td>99.2%</td>
<td>0.8%</td>
</tr>
<tr>
<td>N1[pl] N2[sg]</td>
<td>99.3%</td>
<td>0.7%</td>
</tr>
<tr>
<td>N1[pl] N2[pl]</td>
<td>100%</td>
<td>0%</td>
</tr>
</tbody>
</table>

bottom-up interactions between treelets. This model clearly illustrates the property of SOSP-TH that the rate at which a parse forms depends on its well-formedness. The rate of agreement attraction in the model is lower than what is typically observed in human experiments, and capturing the asymmetry between singular and plural N1s required additional assumptions. However, this model provides a simple demonstration that SOSP-TH models can provide parsimonious explanations for interesting sentence processing phenomena as long as the linguistic representations reflect important properties of the structures involved.

### 3.5.2 An incremental local coherence model

The agreement attraction model just presented, as well as all of the other models in this dissertation, model the processing that occurs at a single point in the sentence. These models are valuable because the majority of conclusions drawn from sentence processing timing data are typically drawn on the basis of effects in one or two regions in a sentence. Thus, these models offer a way of focusing on important effects while
keeping the models relatively simple and easy to understand.

However, a general theory of sentence processing should be able to predict word-by-word processing times for entire sentences. Competing theories like ACT-R (Engelmann, Jäger, & Vasishth, under review; Lewis & Vasishth, 2005) and surprisal (Hale, 2001; Levy, 2008a), for example, make reading time predictions for every word. ACT-R is especially developed in this regard, with some recent efforts focused explicitly on modeling word-by-word reading times for entire sentences (Brasoveanu & Dotlačil, 2018; Dotlačil, 2018). Thus, if SOSP-TH is to be a competitive theory of sentence processing, it should capture processing time data in individual regions of within the context of entire sentences.

Fortunately, the framework described in Chapter 2 and this chapter is extensible to allow for word-by-word processing. I now describe an incremental model of local coherence effects that serves as a proof of concept for incremental processing in SOSP-TH. Additionally, the model, in certain conditions, produces results compatible with the simple, one-word local coherence model discussed in Chapter 2, supporting the idea that the simpler models are useful approximations of the incremental system. The modifications of the system required to get local coherence effects highlight important theoretical and implementation choices that will be necessary to scale the system up to full word-by-word processing.

**Design.** SOSP-TH, described at its most general level, is a framework for creating a dynamical system with attractors at predetermined locations in the state space. What the attractors represent is up to the modeler. For the one-word models used elsewhere in this dissertation, the attractors typically correspond to entire parses, with most of
the treelet features and links left implicit in the local harmony calculation. For this incremental model, the attractors correspond to partial, fragmentary, and complete parses generated by the grammar. In contrast to the one-word models, they explicitly represent phonological forms, head features, dependent features, and link strengths as separate dimensions in the state space.

Like the TRACE model of spoken word recognition (McClelland & Elman, 1986), we use a slot-based encoding to represent different positions in the sentence, repeating each word’s dimensions in each slot.\textsuperscript{10} This assumption is unrealistic in that it is unlikely that this is how the brain actually processes temporal information. Fractal representations of hierarchical information spread over time may be more appropriate (Pollack, 1990; Tabor, 2000, 2003), however the success of models like TRACE suggest that this representation can be a very useful first approximation.

We have constructed an algorithm for setting up the system, locating the attractors, calculating their local harmonies, and parsing sequences of words. The system requires only a lexicon file that specifies for each word the phonological form, the head and dependent features, whether the word requires a governor, and which dependents (if any) are required. The lexicon includes a special empty treelet, which serves as a placeholder for words that have not yet been processed or as padding for sequences

\textsuperscript{10}An alternative choice would be to forgo slots and have just one set of dimensions per word (including phonological, syntactic, and semantic feature dimensions) in addition to link dimensions. Dependency grammars need not be constrained to encode linear order, so taking this step would move SOSP-TH closer to ACT-R, which has had great success in modeling reading times without explicitly representing linear order (Lewis & Vasishth, 2005). This would necessitate having multiple, separate representations of individual word types in order to allow the same word to be used more than once. It seems that any gradient dynamical system will need to have such duplication of word representations because the dimension of its state space must remain fixed for all possible sequences of words.
shorter than the pre-specified maximum sentence length.

The attractors of the system that are constructed consist of all complete, fragmentary, and partial parses possible given the lexicon and the maximum sentence length. As this system often requires hundreds of dimensions and thousands of attractors for even modest lexicon sizes, there is also an option of limiting the system to a corpus of word sequences most relevant to modeling the phenomenon under consideration, greatly reducing the number of dimensions and attractors required. This is only an approximation of the first approach, which we think more accurately reflects the mental structure involved in parsing, but as long as the corpus includes all of the word sequences relevant to studying a particular phenomenon, the results should be very similar, since the system will not go anywhere near most of the attractors not relevant for parsing a particular sequence.

A word is introduced by turning on the dimensions in the state vector that correspond to the word’s phonological form and treelet features at the relevant position in the sentence. If a word is ambiguous, the treelet features of all of its senses are averaged to make the initial state. When a new word is introduced, the values on dimensions corresponding to previous words and links are not changed, but the values of treelet features of yet-to-be-processed words are set to zero. In this way, structures are built up word by word, progressing from one word, to multi-word partial structures (which are penalized if required links are missing), to complete parses (usually) once all of the words in the sequence have been processed. The remaining properties of SOSP-TH are left unchanged: After introducing each word, the system settles until it is sufficiently close to an attractor (corresponding to partial,
fragmentary, or complete structures), and the next word is input.

**Simulation details and results.** To simulate the local coherence effects in Tabor et al. (2004) while still keeping the system to a manageable size, we simulated the subset of words in (1) printed in boldface (leaving out the):

(1) a. The coach smiled **at the player tossed** the frisbee.
   b. The coach smiled **at the player thrown** the frisbee.

The system set up all of the attractors and calculated their local harmonies according to Equation (1) in Chapter 2. The noise magnitude $D$ was set to 0.0001, and the system was run 100 times each of the two conditions of (1). **Tossed** was ambiguous between a main verb, which can take **player** as its subject dependent, and a participle, which **player** can take as a relative clause modifier.

The system always assembles the correct structure in the **thrown** condition, but it can produce two different structures in addition to the correct parse. Without modifying the system, the system produces correct parses in 6 out of 100 trials in the **tossed** condition, a surprisingly low proportion. When it does not produce a correct parse, it produces a fragmentary structure, where **tossed** is analyzed as an active verb, but it does not form a link with **player**, which remains attached as the dependent of **at**. It does not make a similar fragmentary parse with **tossed** analyzed as a participle because that would have lower harmony; **tossed** as a participle requires a governor, whereas the active verb reading does not. This is somewhat

---

11In large systems like this, the harmony surface seems to get very flat due to the summing together of the tails of hundreds of RBFs, making the system especially sensitive to noise.
akin to results like Marslen-Wilson (1975) or Kukona et al. (2014), where lexical ambiguity leads participants to temporarily entertain a locally coherent interpretation. This system produces processing times consistent with Tabor et al. (2004) (\textit{thrown}: $M = 29.28, SD = 0.779$; \textit{tossed}: $M = 55.15, SD = 4.234$). The crux of the Tabor et al. (2004) result, though, is that it seems to require \textit{syntactic} local coherence, in which the system actually builds a syntactic structure incompatible with the rest of the sentence instead of just being confused about part-of-speech labels.

It is therefore important to show that our self-organizing model can produce syntactic local coherence, which is, after all, one of the main data points motivating self-organizing models in the first place. The syntactic local coherence, in which \textit{player} detaches from \textit{at} and reattaches as \textit{tossed}’s subject is always an attractor of the system, but it seems that the geometry of the state space makes it so that the fragmentary parse just discussed is in between the initial condition after reading \textit{player} and the syntactic local coherence, so it always falls into the lexical attractor. Indeed, in runs with no noise, the system always chooses the fragmentary for \textit{tossed}. In order to allow the system to find the syntactic local coherence, we set the harmony of the fragmentary parse to zero so that it is no longer an attractor and cannot softly deflect the system. In addition, pilot simulations showed that the system still failed to produce syntactic local coherence if the links between previously processed words were not reset to zero at each word. These two changes constitute a drastic intervention into the system, but it does allow the system to build the syntactic local coherence in 18 out of 100 trials. The average settling time at \textit{tossed} is higher in than at \textit{thrown} ($M = 62.71, SD = 7.711$ and $M = 29.29, SD = 0.769$, respectively), producing the
effect from Tabor et al. (2004). As expected based on the results of Chapter 2, the lower-harmony syntactic local coherence attractor shows longer settling times ($M = 67.056, SD = 6.593$), although settling times are slow for the correct parse as well ($M = 61.756, SD = 7.644$), suggesting that the harmony landscape is rather flat in the region around these two parses. Overall, these results, especially in conjunction with the one-word local coherence model in Chapter 2, provide initial support to arguments that local coherence effects support self-organizing theories like SOSP-TH.

The incremental model required two significant adjustments to capture the effect, and these highlight considerations for future work. First, the fact that we had to eliminate a competing parse (the fragmentary parse) in order to observe the desired effect shows that we need to better understand how linguistic structures shape the harmony landscape. One approach to this, disconnectivity graphs, has been developed in physics and chemistry (Wales, 2003; Wales, Miller, & Walsh, 1998). A disconnectivity graph plots the harmony (or energy) each attractor of the system and the paths between attractors over energy barriers. Such plots would help us understand how the harmony topology makes the system more likely to choose certain parses over others. It may turn out that different treetlet features or even a very different grammatical formalism are necessary to allow the harmony landscape to have the properties required to reproduce human behavior.

\[12\] This assumes that a more complete version would reproduce the additional result of Tabor et al. (2004), the faster processing for the non-reduced forms ... smiled at the player who was tossed/thrown. We anticipate that a learning model, as discussed above for agreement attraction, would easily deliver this result because participles are relatively infrequent as relative clause dependents (compared to active verbs, e.g., ... player who caught the frisbee...). Thus, we expect a learning model will capture the full interaction in Tabor et al. (2004).
Removing the fragmentary parse also makes it clear that we need a better theory of ungrammatical structures. The big chance we are taking with SOSP-TH is including ungrammatical structures and showing that they affect the processing dynamics in such a way as to provide a parsimonious explanation of interesting timing effects. Here, we have taken a very unrestricted approach to ungrammaticality: Any word can link with any attachment site on any other word or fail to attach anywhere. This laissez-faire approach might be too unrestricted, though. Generally, theories of grammar exclude fragmentary structures from the languages they generate; it might be good to follow suit here and require that all structures have the maximum number of links allowed given the number of words that have been processed. As we continue work on the incremental system, this will be an important area for more exploration.

The second adjustment we had to make to the incremental model, the switching off of previously established links whenever a new word is introduced, highlights the fact that the system, as originally set up, tends to stick to structures it has already established. In other words, new incoming treelets have a difficult time disrupting existing structure. This suggests that the default settings might easily model garden path effects (e.g., Bever, 1970; Frazier & Rayner, 1982), but it is exactly the reason why it has trouble with local coherence effects. Additional research is needed to establish what other, more theoretically motivated changes to the system will allow it to fully capture the human data.
3.6 Discussion

Chapters 2 and 3 have presented the SOSP-TH framework for modeling human sentence processing. It consists of a gradient dynamical system in which the attractors correspond to linguistic structures of varying degrees of well-formedness. The main theoretical result of Chapter 2 was to show that there are three main influences on processing time: First, the amount of time it takes to build a particular structure depends largely on the local harmony of that structure. Second, the average processing time over many runs is the average of the settling times to each attractor chosen, weighted by how often each parse was chosen. Finally, the curvature of the harmony landscape caused by nearby centers can also affect processing times through soft deflection. Chapter 2 also presented the first published self-organizing model of local coherence effects, one of the effects argued to require self-organizing theories like SOSP-TH.

The present chapter provided additional details about the linguistic representations used, provided a proof of the system’s stable long-term behavior, and presented a one-word model of classic agreement attraction and an incremental model of local coherence. The agreement attraction model produced results consistent with well-replicated effects in human language production and brought attention to the need to carefully encode linguistic representations in order to capture psycholinguistic effects. The incremental local coherence model, while providing an important proof of concept, required some adjustments to our initial assumptions about the incremental model in order to fully capture local coherence effects. It highlighted the need to more fully understand how linguistic structures shape the harmony landscape and to
carefully consider the effect that incoming words have on existing structure.

SOSP-TH is quite similar to the Gradient Symbolic Computation (GSC) framework of Cho and Smolensky (2016), Cho et al. (2017), and Cho et al. (2018). GSC defines attractors based on neural embeddings of filler-role representations (Smolensky, 1986, 1990), and its processing dynamics are also driven by the gradient of a harmony function. The main difference between SOSP-TH and GSC is the fact that SOSP-TH allows ungrammatical structures to be attractors of the system, while GSC is constrained to only allow fully grammatical structures.\(^{13}\) GSC accounts for garden path and local coherence effects by having the system settle on grammatical structures that are incompatible with the input (Cho et al., 2017; Cho & Smolensky, 2016). For garden paths, the system narrows in on a particular structure compatible with the ambiguous first part of the input, but when disambiguating input comes later, it is unable to move away from the incorrect attractor that is only partially compatible with the input. For local coherence effects, the opposite happens. The system does not narrow down the set of structures fast enough and ends up in an attractor that is compatible with only the latter part of the input, but not the entire string. In both cases, only grammatical structures are constructed, though. The same contrast holds between SOSP-TH and many other self-organizing theories (Kempen

\(^{13}\)GSC also allows blend states, which are located between fully-specified grammatical structures (Cho & Smolensky, 2016). The point of the blend states is to allow the system to remain uncommitted to a final structure before it has seen the entire input string. SOSP-TH solves this problem by making unblended partial parses attractors. GSC can also stabilize on a grammatical structure that is not consistent with the input. These “hallucinatory” parses have lower harmony than ones consistent with the input, but they are no worse feature matches. SOSP-TH also contains such attractors, but we have not observed the system visiting them. As discussed, GSC relies on these states for modeling garden path and local coherence effects (Cho et al., 2017; Cho & Smolensky, 2016).
& Vosse, 1989; Stevenson, 1994a; Stevenson & Merlo, 1997; van der Velde & de Kamps, 2006), with the exception of SOPARSE (Tabor & Hutchins, 2004): Most other self-organizing models ban ungrammatical structures. The simulations here and in subsequent chapters provide strong evidence that allowing less-than-perfect structures to form provides a powerful and parsimonious way of explaining many human sentence processing effects.

As discussed above, SOSP-TH, as a framework for language processing, is compatible with any grammar formalism that can encode structures as numerical vectors. It provides an intuitive approach to handling less-than-perfect structures as well, although it is also compatible with linguistic formalisms that ban all imperfect-harmony structures. Choices about the linguistics structures can certainly affect the framework’s ability to account for human data (as illustrated by both the agreement attraction and incremental local coherence models), but the processing dynamics are independent of the choice of grammar formalism: Regardless of what the attractors represent, this gradient dynamical system will approach one of its attractors in the long term. This raises the question of what contribution SOSP-TH brings to the theory of sentence processing, given its flexibility. One clear contribution is its transparent account of processing times, which is independent of the grammar used: Ill-formed structures can be built (as long as the grammar assigns them nonzero harmonies), but it will take a long time. Moreover, even if a perfectly harmonious structure is constructed, soft deflection by nearby structures can affect processing. This can be thought of as a type of similarity-based interference, where structures in dense neighborhoods should be slower to process than structures in sparser neighborhoods.
However, the relative independence of grammar formalism and processing dynamics in SOSP-TH might lead one to ask whether SOSP-TH is simply a theory of processing that tells us little about what the representations should be. The modeling results in this chapter suggest that linguistic theory and the processing dynamics of SOSP-TH are tightly bound, at least if SOSP-TH is to be a theory of human sentence processing. Yes, SOSP-TH will find some attractor corresponding to some linguistic structure, and if we just want it to build structures, e.g., for use as an automatic dependency parser, we are free to choose the grammar formalism to maximize efficiency on that task. But if our goal is to understand how humans process sentences, then, as we have seen, the set of attractors in the dynamics and their harmonies have to be properly constrained by the grammar, otherwise the system will fail to reproduce well-established processing effects. Thus, the constraint of explaining human behavior strongly limits the range of choices of linguistic formalism, making SOSP-TH not only a theory of processing but also a constraint on linguistic theory.
Chapter 4

Pseudopartititives

4.1 Introduction

Since the first study of agreement attraction (Bock & Miller, 1991), a number of influences on the agreement process have been identified (see Franck, 2015, for a review). An important class of effects are semantic in nature. Notional plurality—the degree to which a subject noun phrase (NP) refers to more than one thing—has been shown to reliably predict how often participants produce a plural verb (Eberhard, 1999; Humphreys & Bock, 2005; Vigliocco, Butterworth, & Garrett, 1996, among others). For example, Vigliocco, Hartsuiker, Jarema, and Kolk (1996) compared subject NPs like *de kooi met de gorillas* (*the cage with the gorillas*), with subject NPs like *de handtekening op de cheques* (*the signature on the checks*). The sentences like the former were rated as more notionally singular (referring to one thing, the single cage), and sentences like the latter were rated as more notionally plural (referring to
more than one thing, the one repeated signature). They found that the notionally plural subject NPs were associated with higher rates of incorrect plural verb agreement than the notionally singular subject NPs. This suggests that semantic properties of subject NPs can affect the syntactic process of number agreement with the verb.

To explain findings of this sort, Bock et al. (2001) and Eberhard et al. (2005) developed the Marking and Morphing model of subject-verb agreement in sentence production. In the Marking stage, a continuous-valued semantic number rating is associated with the whole subject NP. It is only during the Marking stage that semantic properties of the words, like notional plurality, can affect the agreement process. After that, in the Marking stage, morphosyntactic properties of the words in the subject NP are combined with the semantic number marking to produce a probability of producing a plural verb. These mechanisms, combined with additional mechanisms for actually putting together syntactic structure, allow Marking and Morphing to account for a wide range of semantic (including notional plurality) and morphosyntactic effects in agreement processing.

However, this approach faces some challenges. First, the notional plurality ratings are not independently motivated theoretical constructs. Instead, they are measured via norming studies based on participants’ subjective ratings and then entered into the model, making them an unanalyzed black box in the theory. Second, the way these ratings are combined with morphosyntactic information is not independently motivated. Eberhard et al. (2005) simply stipulate that semantic and morphosyntactic cues are combined linearly, with free parameters related to the weighting of morphosyntactic cues fitted to verb agreement data. Finally, the Marking and Morphing
steps are separate from structure-building. While they do provide an explanation for notional plurality effects, it comes at the cost of a more complicated model compared to one that could capture the data without having additional processing steps.

In Smith et al. (2018), we aimed to contribute to the research on semantic effects in agreement attraction in three ways. First, we showed that a different class of subject NPs (pseudopartitives as in (1)) are a structure to which notional plurality using the standard norming procedure. Marking and Morphing should therefore apply to these materials. Pseudopartitives are a class of NP with the form $N1 \textit{ of } N2$, where the $N1$ denotes and amount or quantity of the stuff denoted by the $N2$ (Brems, 2003; Deevy, 1999; Koptjevskaja-Tamm, 2001; Rutkowski, 2007; Selkirk, 1977; Stickney, 2009). We divided pseudopartitives into three classes (plus plural quantifiers for comparison):

(1)  
   a. \textit{Containers}: a box of oranges  
   b. \textit{Collections}: a stack of sandwiches  
   c. \textit{Measure Phrases}: a lot of newspapers  
   d. \textit{Quantifiers}: fewer mosquitos

Linguistic analyses suggest that there are two syntactic structures available for pseudopartitives, one headed by the $N1$ and one headed by the $N2$ (Figure 4.1; Deevy, 1999; Selkirk, 1977; Stickney, 2009). The head of the subject NP controls the verb’s number marking. In our materials (see below), the $N1$ was always singular, and the $N2$ was always plural, so the $N1$-headed structure should get a singular verb, and the $N2$-headed structure should get a plural verb. The notional plurality

59
norming showed that participants rated each consecutive level of pseudopartitive as more notionally plural than the previous level, i.e., Containers < Collections < Measure Phrases and Quantifiers. Marking and Morphing therefore predicts that we should observe increasing rates of plural agreement in the same direction. The verb agreement choice experiment reported in Smith et al. (2018) and summarized below bears out this prediction, suggesting that semantic effects on agreement are at play in pseudopartitives.

However, this still leaves notional plurality as a black box. It is desirable to replace such a black box with an explicit, independently motivated mechanism. As discussed in Chapters 2 and 3, the assembly of syntactic structures in self-organizing theories is guided by feature match. We therefore ran a series of norming studies to uncover suitable features. The results revealed a hierarchy of semantic features that differentiated the classes of pseudopartitives in (1). The presence of the features is related to the concreteness of the N1, which has been argued to be associated with subject-hood (Keenan, 1976), while the absence of the features is associated with quantifier-hood, which would make the N1 less suited to be the subject. One
feature was related to whether the N1 denoted a container (±Container), one was whether the N1 constrained the physical configuration of the objects denoted by the N2 (±SpatialConfig), and one was whether the N1 is prohibited from being paired with an abstract N2 (like ideas, ±AbstractN2Prohib). The hierarchy of these features tracked the progression of notional plurality ratings (see Table 4.1), so the second contribution of Smith et al. (2018) was to show that the black box of notional plurality could be unpacked into a set of independently motivated semantic features. Finally, we presented an SOSP model that employed the semantic features to guide the formation of an N1- or N2-headed structure and reproduced the pattern of verb choice results observed in the verb selection experiment. In contrast to Marking and Morphing, it did so without relying on mechanisms beyond those used for basic structure building and without relying on theoretical black boxes.

In Smith et al. (2018), we focused on the verb choice results in the human data and their SOSP model, and we did not report the verb choice response times. The main goal of this dissertation is to show that SOSP-TH is a general-purpose, widely applicable model of sentence processing times. Thus, the present chapter analyzes the response time data from Smith et al. (2018) and compares those results with the predictions of a new SOSP-TH model. Why did we not report the response times in Smith et al. (2018)? The main reason was limitations of the SOSP model.

---

1Example items from the semantic feature norming are:

(i) a. Container-hood: We added so many [strawberries/shirts] that the [dish of strawberries/pile of shirts] overflowed.
   b. Spatial configuration: She moved a tall [pile/bunch] of shirts into the garage.
   c. She defined a [pile/bunch] of concepts.
we used for that work. That model was based on the Lotka-Volterra competition equations (Frank, 2014; Goel, Maitra, & Montroll, 1971; Lotka, 1920), which did a good job of modeling the verb choice data. Pilot simulations with that framework suggested that it would be possible to generate response time predictions from that model, but in order for those equations to produce reasonable behavior, we had to restrict the types of interactions between words that were allowed, violating the spirit of self-organization and making extending the equations to other structures difficult. The SOSP-TH model described in the next section is an instantiation of the general-purpose theory presented in Chapters 2 and 3 and is meant to test whether an SOSP model can simultaneously model parse distributions and processing times. After the model, we present the verb choice response time data from Smith et al. (2018) (along with the previously reported proportions of verb forms chosen). We conclude by discussing the fit of the model to the human data and by placing the results into the context of the literature.

4.2 An SOSP-TH model of verb choice

In the experiment below, participants were tasked with selecting a singular or plural verb form after reading the subject NP. We used a 4 (N1 type: Containers, Collections, Measure Phrases, or Quantifiers) x 2 (N2 presence, i.e., with or without of N2) design. Usually in agreement attraction experiments, one contrasts singular and plural N2s to test for attraction. With pseudopartitives, though, singular N2s are ungrammatical: *a box of orange. We therefore opted to measure the effect of the plural N2 by
including subject NPs without of N2, comparing, for example, a box of oranges to a box.

The present SOSP-TH simulations model the verb selection task by simulating the competition between the N1-verb link and the N2-verb link shown in Figure 4.1. We view this simplified structure formation process as an approximation for the construction of the full subject NP structure and its attachment to the to-be-produced verb. The system, therefore, consisted of two dimensions, one coding the strength of the N1-verb link and the other coding the N2-verb link. We include the Quantifier conditions here, in contrast to the simulations in Smith et al. (2018) which left out that condition for simplicity.

To specify the SOSP-TH model, we first specified the features used on the head and dependent attachment sites. We used the same settings as those used in Smith et al. (2018): ±Noun, ±Container, ±SpatialConfig, ±AbstractN2Prohib, and ±Present. When the Present feature is turned on, the word is actually present in the input, whereas when it is off, the word has been elided to model the -N2 conditions. The feature vectors on the attachment sites are given in Table 4.1. The system had two dimensions, one coding the N1-verb link and the other coding the N2-verb link. The local harmonies \( h_i \) for the two centers (one for each link successfully attaching to the verb) were calculated using Eq. 1 of Chapter 2; these are shown in Table 4.2. Because we used five features, each feature mismatch decrements the local harmony of a parse by 0.2. For example, in the +N2 Container condition, the N1-headed structure has a perfect feature match, so its \( h_i = 1.0 \), while the N2-headed structure has an \( h_i \) of 0.8 because it mismatches on the AbstractN2Prohib feature, which
Table 4.1: Feature vectors (one per row) for the SOSP-TH simulations. These are
head feature vectors for the N1 types and the N2 and subject dependent features for
the verb.

<table>
<thead>
<tr>
<th>Attachment site</th>
<th>NOUN</th>
<th>CONTAINER</th>
<th>SPATIAL CONFIG</th>
<th>ABSTRACT N2Prohib</th>
<th>PRESENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Collection</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Measure</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Quantifier</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>N2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0/1</td>
</tr>
<tr>
<td>Verb’s subject attachment site</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

does not apply to non-pseudopartitive nouns. The RBF width parameter $\gamma$ was set
to 0.2. This is smaller than the default of 0.4. This was necessary for separate N1-
and N2-headed attractors to exist. The noise magnitude $D$ was set to 0.0005. As
mentioned above, we used a 4 (N1 type: Containers, Collections, Measure Phrases,
or Quantifiers) x 2 (N2 presence: with of N2 or without of N2) design. For each
condition, we ran the system 2000 times starting from (0, 0), recording the settling
time (using the Chebyshev stopping criterion) and which attractor the system chose.
As in Smith et al. (2018), we assume that whichever noun attaches as the subject
selects the verb’s number, so if the N1 is selected, the model chooses the singular
verb, and if the N2 is selected, the model chooses the plural verb.  

The proportion of runs in which the model selected the N2-headed parse are
shown in Figure 4.2, and the mean settling times with 95% confidence intervals (CIs)

---

2As Schmid, Villata, Tabor, and Franck (2018) discuss, there are two ways that the presence
of the plural N2 can affect the verb’s number: either through the N2 attaching as the subject or
through the N2, in a short interaction with the verb, bumping the verb’s number to plural even if
the N1 is selected as the subject.
Table 4.2: Local harmonies \( (h_i) \) for the two possible parses in the eight simulated conditions.

<table>
<thead>
<tr>
<th>Parse</th>
<th>+N2</th>
<th></th>
<th></th>
<th></th>
<th>-N2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1-headed</td>
<td>1</td>
<td>0.8</td>
<td>0.4</td>
<td>0.2</td>
<td>1</td>
<td>0.8</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>N2-headed</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
</tr>
</tbody>
</table>

are plotted in Figure 4.3. Starting with the parse choice results, it is clear that the presence of the N2 (long condition) led to an increase in how often the N2-headed parse was chosen (except for the Quantifiers conditions). This is because the harmony of the N2-headed parse is higher when the N2 is actually present in the input than when it is not. Also, the probability of selecting the N2-headed parse increases from Containers to Collections to Measure Phrases and Quantifiers. This is because the harmony of the N2-headed parse increases relative to the harmony of the N1-headed parse: As the N1 becomes more quantifier-like in its features, it becomes a worse feature match for the verb’s subject attachment site, making it less likely to be selected.

The settling times show an interesting interactive pattern between N1 type and N2 presence. For Containers, the system settles at approximately the same average rate in the +N2 condition than in the -N2 condition. The harmony of the N2-headed parse is quite high in the +N2 condition, so it is selected quite often. But because the N2-headed parse still has lower harmony than the N1-headed parse, the system settles more slowly when it picks the N2-headed parse, bringing the average settling time up. For -N2 Containers, the N2-headed parse has lower harmony. So, even though selecting it causes a slow settling time, it was selected relatively rarely, so its slowness has little effect on the mean. Without the N2, there is little competition from the
Figure 4.2: Mean proportions of runs in which the model selected the N2-headed parse with 95% CIs.
N2-headed parse, so the system can easily build the other structure, which balances out the averages between the +N2 and -N2 conditions. In the Collections, neither parse is fully well-formed in either the +N2 or -N2 conditions, so settling times for Collections are slower than for Containers. In the +N2 condition, the two parses have the same harmony, so the noise chooses each one roughly equally often (Figure 4.2). In the -N2 condition, the N2-headed parse has lower harmony, so it is selected less often, but when it is selected, its slow settling pushes the average settling time up and counteracts the relatively fast settling to the N1-headed parse. For Measure Phrases and Quantifiers, the model produces a large difference between the +N2 and -N2 conditions, with much faster settling times in the +N2 condition. When the N2 is present, it can quickly and easily form a relatively high-harmony structure as the subject of the verb. When the N2 is not present, both the N1- and N2-headed structures are quite ill-formed, so which ever structure forms, the settling will move slowly. Finally, the Quantifiers were somewhat faster in the +N2 Condition than Measure Phrases. This is because the lower-harmony to higher-harmony ratio was greater for Quantifiers than for Measure Phrases, resulting in less competition-based slowdown for Quantifiers.

We note here that the grammar-controlled surprisal theory (Hale, 2001; Levy, 2008a) predicts a very similar pattern of results. Surprisal theory was designed for comprehension, but we can reasonably extend it to production: We assume that when it comes time to choose a verb in this task, the parser samples a verb from the probability distribution given what it has read so far. It then updates the probabilities over structures given the new word. The processing time to do this is proportional to

67
Figure 4.3: Mean settling times with 95% CIs for the SOSP-TH model of verb selection.
the surprisal between the probability distribution before choosing the verb and the probability distribution after choosing the verb. This predicts that when different structural options are roughly equally frequent, i.e., the probability distribution over parses (here, over verb number markings) has a high entropy, processing times should be long. This is because whichever option the parser chooses, it will have to update its probabilities quite a lot compared to when the distribution has lower entropy. In that case, the parse it chooses is most likely the higher-frequency one, so it will not have to update its probabilities very much. Haskell et al. (2010) report a corpus study where they found that subject NPs with a singular N1 and a plural N2, like our +N2 materials, had lower rates of correct agreement (and therefore higher entropy; 79.1%) compared to singular-singular and plural-plural subject NPs (lower entropy; 99.4% and 100%, respectively). If we assume that the -N2 conditions have low entropy like the latter structures, then surprisal predicts a difference between +N2 and -N2. It also predicts an increase in this difference across the N1 types: An N1 like box has fewer options for subsequent continuations, so it will have relatively low entropy and therefore fast processing compared to lot, for example, which can be followed by many different things due to its abstract, non-constraining meaning. The effect of N2 presence will be similar to what SOSP-TH predicts: The +N2 conditions are more constraining about what can follow, so it should be processed faster than the -N2 conditions due to the lower entropy for +N2. The place where surprisal and SOSP-TH differ is in the Quantifiers. For the both +N2 and -N2 quantifiers, surprisal predicts that processing times should be fast because the plural quantifiers or the plural noun strongly constrain upcoming structure to be plural. Thus, there should
be no effect of N2 presence and the verb choice speeds should be quite fast.

With the predictions from SOSP-TH and surprisal in mind, we turn now to the human experiment originally reported in Smith et al. (2018).

### 4.3 Experiment 1

Here, we briefly present the methods and verb choice results from Smith et al. (2018), but we focus on the response time data, which was not reported in that work.

#### 4.3.1 Method

**Participants**

Fifty-seven participants drawn from the University of Connecticut Psychological Sciences participant pool took part in the experiment for course credit.

**Materials**

For each N1 Type (Containers, Collections, Measure Phrases, and Quantifiers), there were eight different lexical variants of the N1, e.g., *box, tub* and *shoebox* were among the Container N1s. We used a 2 (N2 present or absent) x 4 (N1 type) design. In order to make the -N2 conditions felicitous, we included a context question before each item, e.g., *Do we have anything to juggle around here [, like balls]? A tube [of balls] is/are by the tennis racket.* There were 64 critical items along with 64 filler sentences. The materials were split into two lists, each with sixteen critical items and 32 fillers. Within each list, we counterbalanced for N2 presence, N1 type, verb tense
(past or present), and the number of fillers taking singular and plural agreement. The full set of materials are given in the Appendix.

**Procedure**

In many studies of agreement attraction, participants are presented with a subject NP and asked to repeat it aloud and then finish the sentence. This method, while it is somewhat similar to natural production, has the disadvantage of losing data due to participants not producing number-inflected verbs or making other mistakes. To avoid this problem while still eliciting verb productions, we used a verb selection task introduced by Staub (2009, 2010). In this paradigm, participants read the subject NP in rapid serial visual presentation (RSVP). When the region of the sentence containing the verb arrives, both the singular and plural forms appear on the screen, and participants press a button to choose one. After that, the remainder of the sentence is presented in RSVP. This way, no data is lost, as participants are required to choose a number-inflected form for each sentence.

The experiment was presented using E-Prime® (version 2.0 Schneider, Eschman, & Zuccolotto, 2012). After giving informed consent, participants read the instructions and completed four practice sentences before starting the actual experiment. For each item, the context question appeared in its entirety on the screen until the participant pushed a button to begin the test item. A fixation cross was presented for 1000ms, and the sentence was presented in one- or two-word chunks in the center of the screen. Each chunk appeared for 250ms followed by 150ms of blank screen, following Staub (2009, 2010). For the verb choice, the singular form (is or was) was presented on
the left side of the screen and could be selected using the “1” button on the number pad of the keyboard. The plural form (are or were) was presented on the right and was selected with the “3” button on the number pad. One thousand milliseconds separated each trial, with a break after half of the trials. The slashes in (2) and (3) show where the breaks between chunks were.

\begin{equation}
(2) \quad +N2 \text{ Condition} \\
\text{Do we have anything to juggle around here?} \\
+ / \text{ A tube / of balls / [VERB CHOICE] / by the tennis racket.}
\end{equation}

\begin{equation}
(3) \quad -N2 \text{ Condition} \\
\text{Do we have anything to juggle around here, like balls?} \\
+ / \text{ A tube / [VERB CHOICE] / by the tennis racket.}
\end{equation}

4.3.2 Results

Before analysis, we removed all trials with RTs less than 50ms or greater than 5000ms (less than 2% of the data). All data were analyzed using (generalized) linear mixed effects models (Bates, Maechler, Bolker, & Walker, 2015). N2 presence was coded using numerical sum coding (+N2 = 0.5, -N2 = -0.5). N1 type was coded using Helmert coding: The mean of Collections was compared to the mean of Containers, the mean of Measure Phrases was compared to the mean of Collections and Containers together, and the mean of Quantifiers was compared to the mean of all other levels of N1 type. This coding allowed us to test for changes in the pattern of data from one level to the next. The interaction of N1 Type with N2 presence tells us if the
effect of N2 presence changed between its effect at level level $n$ of N1 type and its effect at level $n + 1$.

Verb choice  The human verb choice data are shown in Figure 4.4. The verb choice responses were analyzed using generalized linear mixed effects models with a logit linking function. Here, we report odds ratios of plural to singular verb choices. These are calculated by taking the natural logarithm of the coefficients of the model. Estimates that have 95% CIs that include 1.0 are not significant at the .05 level. For this analysis, the random effects structure had to be simplified down to just random intercepts by participant and item in order for the model to converge. There was an effect of N1 Type such that the odds of choosing a plural verb increased from each level to the next: Containers to Collections, $e^b = 1.917$, 95% CI [1.400, 2.625]; Collections to Measure Phrases, $e^b = 2.316$, 95% CI [1.934, 2.775]; and Measure Phrases to Quantifiers, $e^b = 2.439$, 95% CI [2.093, 2.843]. There was also an effect of N2 presence such that participants were more likely to choose a plural verb in the +N2 conditions: $e^b = 8.054$, 95% CI [5.852, 11.085]. Finally, there were effects of two of the interaction coefficients: The difference between +N2 and -N2 conditions was smaller than the average of the preceding differences for Measure Phrases ($e^b = 0.773$, 95% CI [0.625, 0.957]) and Quantifiers ($e^b = 0.760$, 95% CI [0.613, 0.943]). The interaction at Collections had no effect ($e^b = 0.785$, 95% CI [0.530, 1.162]).

Response times  The log-transformed verb choice response times are plotted in Figure 4.5. They were analyzed using a linear mixed effects model with fixed effects for N2 presence and N1 type (same coding scheme as before; Bates, Maechler, et
al., 2015). The random effects structure was maximal given the design (Barr, Levy, Scheepers, & Tily, 2013): by-participant random intercepts and slopes for both fixed effects and their interaction and by-item random intercepts and slopes for N2 presence. Ninety-five percent confidence intervals that exclude zero would be considered significant at the .05 level.

There was a main effect of N2 presence such that the +N2 condition produced shorter RTs ($b = -0.178$, 95% CI = [-0.241, -0.115]). The coefficients for the main effect of N1 type were not significant, Collections vs. Containers, $b = 0.036$, 95% CI [-0.003, 0.076]; Measure Phrases vs. previous, $b = 0.015$, 95% CI [-0.012, 0.042]; Quantifiers vs. previous, $b = 0.001$, 95% CI [-0.016, 0.019]. However, there was an interaction: The effect of N2 presence for Collections differed from the effect of N2 presence for Containers.

---

3The model did not converge with the default settings, but increasing the number of optimization iteration steps allowed the model to converge without warnings or errors.
Figure 4.5: Mean verb choice response times with 95% CIs for the human experiment (left panel) and the SOSP-TH model (right panel).

presence for Containers ($b = -0.110$, 95% CI = [-0.191, -0.028]). The effect of N2 presence for Measure Phrases differed from the effect of N2 presence for Collections and Containers ($b = -0.075$, 95% CI = [-0.121, -0.029]). Finally, the effect of N2 presence for Quantifiers differed from the effect of N2 presence for Measure Phrases, Collections, and Containers ($b = -0.038$, 95% CI = [-0.074, -0.001]). These coefficients indicate that size of the difference was increasing level to level.

4.3.3 Discussion

In the human data, participants became more likely to choose a plural verb as the N1 became less concrete and more quantifier-like (i.e., progressing from Containers to Collections to Measure Phrases to Quantifiers) and when the N2 was present in the sentence, with the difference between +N2 and -N2 decreasing along the N1 type cline. As Smith et al. (2018) discuss, the effect of N2 presence parallels previous
agreement attraction effects that manipulate N2 number instead of presence. These results extend previous ones to effects of manipulating the semantics of the N1 along a gradient. Importantly, there is good agreement between the human data and the pattern of results from the SOSP-TH model. Figure 4.4 shows that, in both cases, rates of plural agreement are higher when the N2 is present, they increase along the N1 type cline, and the difference between +N2 and -N2 decreases in the same direction. The original model presented in Smith et al. (2018) also produced similar results, suggesting that, in general, competitive, feature-based SOSP parsers that include lower-harmony attractors offer a good approach to this type of structure. The results are also expected under Marking and Morphing; however, Marking and Morphing requires additional processing mechanisms in addition to structure building in order to capture the results.

In the human response times, the +N2 conditions tended to be faster, with the difference between +N2 and -N2 increasing along the N1 type cline. Marking and Morphing does not make clear predictions about response times for this type of subject NP (although see Brehm & Bock, 2013, for response time predictions with a different class of subject NPs), so we focus on SOSP-TH. Once again, the model’s fit to the human data is quite good. Figure 4.5 shows few points of difference: Processing times in the +N2 conditions generally increase from Containers toward Quantifiers in the model, whereas they show a decreasing numerical trend in the human data. All of the 95% CIs overlap in those conditions, though, so the decreasing trend could be completely spurious. Overall, though, the patterns predicted by the SOSP-TH model accord very well with the human data.
The data are also largely in accord with the predictions of surprisal theory, except for the Quantifiers. Surprisal predicts fast processing times for both N2 presence conditions, while the human data showed a large increase for the -N2 condition. Thus, while SOSP-TH and surprisal make similar predictions for most conditions, the data agree with SOSP-TH where the models diverge.

4.4 General Discussion

The verb choice and response time data from the human experiment of Smith et al. (2018) align well with the predictions of the simple SOSP-TH model presented here, with only small discrepancies. SOSP-TH predicts the results by allowing high- and low-harmony parses to compete. When two nouns in the subject NP are a good feature match to the verb's subject attachment site, we can expect competition-based slowdowns. In this case, the verb is expecting a subject with a set of empirically supported (Smith et al., 2018) semantic features. Containers and, to a slightly lesser extent, the N2 are assumed to match that set of features, so if the noise pushes the system into the basin of attraction where one of those attaches as the subject, the verb’s number can be decided relatively quickly. But the other N1 types match the verb’s expectations less and less, and so they slow processing (to the extent that the system selects the N1-headed parse at all in those conditions). We note that the use of the five features here leads to stark contrasts in the N1’s feature match with the verb between Containers and Collections versus Measure Phrases and Quantifiers. However, as long as the N1’s features decrease in how well they fit the verb’s subject
expectations, a similar pattern of results should hold. Thus, SOSP-TH provides a parsimonious way of explaining not only the verb choice patterns (which Marking and Morphing can also explain) but also the response times (which Marking and Morphing is not equipped to predict).

Surprisal theory predicts many of the differences observed in the verb choice response times with the important exception of the Quantifiers. We note that, if we assume that any verb (plural or not) is very infrequent after a plural Quantifier, then it might predict the slowdown in the -N2 condition as long as this effect can overpower the strong constraint of needing a plural continuation. But even allowing this, because surprisal is grammar-controlled, it lacks a way of producing plural verbs with Containers. According to any reasonable grammar, singular Containers should only ever have singular verbs. This applies to most cases of agreement attraction in production, which highlights a central weakness of surprisal.

We also note that SOSP-TH makes its predictions based on two separate aspects of its design, competition between structures and the inclusion of low-harmony attractors. Surprisal can only make predictions based on competition between strictly grammatical structures. SOSP-TH’s inclusion of sub-optimal parses is what allows it to fully capture the range of data in the human experiment. Where surprisal fails most clearly (with Containers) is where it would need to allow for ungrammatical structures.

---

4We note that the balance between these two is not fully transparent in the current model. One way of understanding this better is to explore the full range of local harmony values for the two competing peaks, allowing them to range between zero and one independently. Recording the average setting time for each combination of harmony peaks would allow us to plot an RT surface over the two-dimensional parameter space, providing a visual way of understanding how the local harmonies affect the degree of competition between attractors.
As Marking and Morphing does not make clear timing predictions and surprisal only captures parts of the data, we consider now the predictions of ACT-R (Lewis & Vasishth, 2005). Recall that, in ACT-R, all structure building involves retrieving chunks of syntactic structure from memory on the basis of the activation level of the to-be-retrieved chunks and cues determined by the word currently being processed. chunks receive activation according to how well their features match retrieval cues, and when multiple chunks are at least partial matches for the retrieval cues, they race to the activation threshold for retrieval (Logačev & Vasishth, 2015). While ACT-R is grammar controlled in that retrieval cues are determined strictly according to grammatical rules, the noisy retrieval mechanism can sometimes produce ungrammatical structures by retrieving incorrect memory chunks.

In order to produce a verb form in the present experiment, ACT-R needs to retrieve a noun in the subject NP so it can determine the verb’s number (Badecker & Kuminiak, 2007). We argued that the semantic features from the Smith et al. (2018) norming were relevant to subject-hood, so we assume that they can be used as retrieval cues under ACT-R. In that case, ACT-R can predict the observed verb choice effects on N1 type, and the timing effects, but only in the +N2 conditions. Just like in SOSP-TH, the features on the nouns gradually start to favor the N2 as the subject as we move from Containers towards Quantifiers, making the N2 more likely to be retrieved as the subject and thus producing more plural agreement. In the response times, ACT-R would likely struggle the most with the Containers. Container N1s are excellent subjects, and so ACT-R should retrieve the N1 quickly and almost every time, predicting very low rates of agreement attraction and very fast processing.
The human data, though, show high rates of agreement attraction and relatively slow processing, at least compared to the fastest condition, the +N2 Quantifiers. Thus, while ACT-R does a reasonable job with much of the observed data, it does not capture the full range of effects like SOSP-TH.

In sum, we have seen that Marking and Morphing can provide a good account of how often singular and plural verbs are produced, surprisal and ACT-R can predict (some) response times, but only SOSP-TH can do both in a single parsimonious model.
Chapter 5

Encoding interference

5.1 Introduction

Long-distance dependencies in language are common. 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*. During language processing, the dependency between *the key* and the verb has to be established in order to build an interpretable structure. One prominent theory of how such dependencies are created during language comprehension relies on cue-based retrieval: After reading *is*, the parser has to retrieve a subject on the basis of cues at the verb, such as +SINGULAR and +NOUN (Lewis & Vasishth, 2005; McElree et al., 2003).

This retrieval process can be subject to similarity-based interference when multiple NPs in memory share retrieval features. For example, the retrieval of the correct word can be delayed when it is approximately equally matched on retrieval cues with
another competing word, or the wrong word can be retrieved. An example of this comes in studies of agreement attraction in comprehension, where self-paced reading times or eye-tracking measures are used to compare the ungrammatical structures in (1):

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

Typically, participants read the verb in (1-b) more quickly than the verb in (1-a) (Dillon, Mishler, Sloggett, & Phillips, 2013; Jäger et al., 2017; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Pearlmutter et al., 1999; Wagers et al., 2009). The cue-based retrieval model of Lewis and Vasishth (2005) explains this effect: In ACT-R, words must reach an activation threshold in order to be retrieved, and when a word matches some retrieval cue, its activation is increased, bringing it closer to the threshold. Because key (with its +Subject feature) and keys (with its +Plural feature) are both partial matches for the verb’s +Subject and +Plural cues, some activation spreads to both nouns, leading to a race process as both nouns’ activations approach the threshold for retrieval (Logačev & Vasishth, 2015). So, in (1-b), activation spreads to both key and cabinets because each matches some retrieval cues (e.g., +Subject and +Plural, respectively). In (1-a), cabinet is neither the subject nor plural, so it gets no activation boost, leaving key to slowly accumulate activation based on its partial feature match, which results in slower average processing times than (1-b).

However, not all interference effects can be explained as retrieval interference. A number of studies have provided evidence that features not relevant for retrieval in
the service of structure building can interfere with language processing. For example, Gordon et al. (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. 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 for (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. This is an example of encoding interference, which is when features of a word that are present when the word is perceived (and then encoded in memory) affect sentence processing if they are similar to the encoding features of other words.

Hofmeister (2011) and Hofmeister and Vasishth (2014) found similar effects: They observed significant reading time slowdowns at the retrieval site in long-distance

\(^{1}\)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.
dependencies when the retrieval target was less syntactically and semantically distinct from competitors. That is, when there were multiple NPs that could be retrieved, retrieval was facilitated when they were less similar even though they did not differ in their match to retrieval cues. Villata et al. (2018) tested for encoding interference in subject-verb agreement in grammatical English and Italian sentences. With clearer effects in Italian than in English, they found evidence of interference in the comprehension questions, where participants were significantly more likely to correctly answer the question in the encoding feature mismatch conditions than in the encoding feature match condition. There was also a related effect in reading times at the verb, with faster reading times in the mismatch condition compared to the match condition. Finally, in production, Gennari et al. (2012) found that the semantic similarity between two animate nouns correlated with the rate at which one of them was elided in a relative clause question-answering paradigm. Gennari et al. interpret this as an effect of increased competition between the semantically similar items. Together, these studies provide evidence that when two items have similar encoding features, they tend to be harder to process than when their encoding features are distinct.

This chapter further explores an encoding interference effect in agreement attraction reported in Barker et al. (2001, their second experiment). Barker et al. found higher rates of agreement attraction (in production) for subject noun phrases like (3-b) than for (3-a).

(3)  a. The canoe by the cabins...
     b. The canoe by the sailboats...
Barker et al. argued that this effect was due to the similarity in meaning (semantic feature match) between canoe and sailboats, which allowed sailboats to influence the verb number more than cabins, a clear case of encoding interference in production. Cue-based retrieval cannot explain these results, as sailboats and cabins do not differ in their retrieval cues and so they should be erroneously retrieved with equal probability.

There is no current leading explanation for these effects in sentence processing, and the ACT-R model of Lewis and Vasishth (2005), while able to capture many (but not all, Engelmann et al., under review; Jäger et al., 2017) retrieval interference effects, does not predict encoding interference effects. However, a simple extension to cue-based retrieval does allow that theory to account for these results. The extension borrows a spreading activation mechanism from the RACE/A (Retrieval by ACcumulating Evidence in an Architecture; van Maanen & van Rijn, 2007; van Maanen et al., 2009, 2012) extension of ACT-R. Under this setup, activation spreads between memory chunks (e.g., NPs) that share encoding features. This means that similar items in memory will send activation to each other, increasing each other’s activation in a positive feedback loop. van Maanen and van Rijn (2007) showed that this mechanism can account for stimulus onset asynchrony effects in picture-word interference studies, and van Maanen et al. (2009) showed that it can provide good coverage of Stroop task data, so the mechanism has support beyond using it for sentence processing.

If we combine this activation spreading with the standard ACT-R assumptions that chunk activations are noisy and that a chunk is retrieved when its activation

---

2Villata et al. (2018) discuss a different extension of ACT-R that can also account for this effect. We discuss this in the general discussion of this chapter, as it makes similar predictions to SOSP-TH.
reaches a preset threshold\textsuperscript{3}, we can account for the Barker et al. (2001) data. In (3-b), \textit{canoe} and \textit{sailboats} share semantic features related to being boats, so they spread activation to each other. After producing the subject NP, the participant has to produce a verb, and so they search in memory for an appropriate noun control the verb’s number. Because of their semantic similarity, \textit{canoe} and \textit{sailboats} are egging each other on in the race to the activation threshold for retrieval, and their activations will tend to converge. This is because a word spreads activation to other words in proportion to its own activation (a utopian rich-get-more-generous system). Thus, \textit{canoe}’s larger initial activation (due to it being a better feature match for retrieval cues) will send enough activation to \textit{sailboats} to allow it to start to catch up. No such convergence will happen for \textit{cabins}, which does not exchange activation with \textit{canoe}. The features on \textit{canoe} make it most likely to reach the threshold first and be selected as the subject in both cases, but because its activation converges with that of \textit{sailboats}, sometimes the noise will bump \textit{sailboat}’s activation over the threshold first, causing an erroneous retrieval and the production of a plural verb. This will happen less often with \textit{cabins} because its activation will rarely catch up with that of \textit{canoe}, thus producing the Barker et al. effect.

In addition to accounting for the Barker et al. data, this mechanism also makes a novel prediction: The subject-retrieval process should finish faster in the \textit{sailboats} condition than in the \textit{cabins} condition, a pattern that should be reflected in self-paced reading times. However, as shown in the next section, SOSP-TH can also account for the Barker et al. findings, and it makes the opposite prediction in self-paced reading:

\textsuperscript{3}Note that RACE/A rescales the activations before evaluating whether the threshold has been reached, making relative activation the retrieval criterion.
Participants should read the verb more slowly when the two nouns in the subject NP are semantically similar. Experiments 2A and 2B were designed to test the diverging predictions of these two theories. Before turning to the experiments, I first discuss an implemented SOSP-TH model.

5.2 An SOSP-TH approach to encoding interference

SOSP-TH can account for the agreement attraction results in Barker et al. (2001) if we assume that features on the verb’s subject attachment site can already be affected by preceding words. That is, when the parser processes *canoe*, it starts anticipating a verb that has features associated with boats. Such anticipatory processing has strong support in the literature (e.g., Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003). With this expectation of a boat-like subject established after reading or producing *canoe*, the Barker et al. result is derived in the following way. In the *sailboats* conditions, the N2 has boat-like features, so it makes a relatively good feature match for the verb. This feature match allows its plural number feature to start pushing the verb’s number feature towards its plural attractor, in competition with the singular influence from *canoe*. Because the dynamics are noisy, *sailboats* will sometimes succeed in attaching as the verb’s subject, forcing the verb to be produced in its plural form.\(^4\) *Cabins* lacks the boat features of *sailboats*, making it a poorer

\(^4\)In the full incremental SOSP-TH model, it is possible that the verb’s plural feature can be activated even if *canoe* attaches as the subject (Schmid et al., 2018). However, this is a low-harmony attractor, so its influence on parsing will be weak compared to the attractor in which *sailboats* attaches as the subject as long as fragmentary parses are not penalized too much.
feature match for the verb and therefore weakening its influence on the verb’s number. Because of this, \textit{cabins} is less likely to attach as the subject, although it still happens occasionally due to the noise. The model described below demonstrates this effect in more detail.

Unfortunately, in the present instantiation of SOSP-TH as described in Chapters 2 and 3, this approach to anticipatory processing is problematic because of the slot-based representations: How should the parser know which slot to turn the boat feature on in, since there is no fixed position in the sentence where the verb must appear? One way of doing this would be to have attractors that have the verb’s boat features turned on at any slot where the verb might appear. In other words, there would be separate attractors with the verb’s boat features active in every slot following the ones that had already been filled by words in the input. However, this would be a major contributor to combinatorial explosion, which is already an issue in the incremental model. Vosse and Kempen (2009) solve this by allowing activation to spread from the current slot to a special slot for holding predictions about upcoming material in the next slot in the sentence. In SOSP-TH, as currently implemented, there is no way of ensuring that the head features on \textit{canoe} are associated with predicted boat features on a verb. We leave the implementation of a suitable prediction mechanism to future research and assume here, in a simpler SOSP-TH model of settling after reading the verb, that the prediction mechanism has turned on boat features on the verb’s subject attachment site.
5.2.1 Model design

To model self-paced reading times at the verb, we set up a one-word SOSP-TH model with two centers. One represented the N1-to-verb link and the other the N2-to-verb link. This models the establishment of an attachment link between the verb and one of the nouns after reading the verb. To calculate the local harmonies \( h_i \), we assumed there were three semantic features coding properties of being a boat and one number feature. This setup reflects our assumption that semantic representations of words are rich and distributed, while syntactic representations are discrete and low-dimensional. Moreover, these settings put the model in the range of parameters where we can expect competition-based slowdowns as opposed to competition-based speedups, as would be necessary for modeling ambiguity advantage effects (see Chapter 2 and Traxler, Pickering, & Clifton, 1998).

The \( h_i \) were calculated using Eq 1 from Chapter 2, 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 with an attachment site without a link, which was penalized by multiplying the \( h_i \) by the missing link parameter set to 0.8. The initial condition was both link strengths set to zero, i.e., equidistant from each attractor. The width parameter \( \gamma \) was set to 0.4, and the noise magnitude \( D \) was 0.001. 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).
5.2.2 Results and discussion

Table 5.1 shows the number of times the model settled into the two attractors in each condition. In the 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 frequently, as the N2-verb attractor had higher harmonies in these conditions. The N2 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. We note these data do not show agreement attraction at all in the semantically dissimilar N1[sg]-N2[pl] conditions, unlike what Barker et al. (2001) showed. However, the present model is meant to make predictions about self-paced reading, not production. To model production, a model similar to the one used for pseudopartitives in the previous chapter would be more appropriate: We would initialize the system with the N1 and N2 features set appropriately and allow the system to settle on a verb number. In this model, the fact that the verb number is provided to participants provides a strong constraint on the type of structures they can build, making N1-headed structures most common.

More important for the comparison with the predictions of the RACE/A extension to ACT-R are the predicted settling times at the verb, which are shown in Figure 5.1 with 95% confidence intervals. A few patterns are worth noting, starting first with the grammatical conditions (left panel). First, settling times are longer in the semantically similar conditions than in the dissimilar conditions. This is due to extra competition from the N2, which is a good feature match for the pre-activated boat features on the verb and the opposite of what the RACE/A extension to ACT-R predicts. In addition,
Table 5.1: Percentages of runs that settled to the two attractors; 2000 runs per condition. The N1 was always singular, so any configuration with a plural verb is ungrammatical.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Semantic similarity</th>
<th>N2 number</th>
<th>Verb number</th>
<th>N1-V</th>
<th>N2-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissimilar</td>
<td>Singular</td>
<td>Singular</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Dissimilar</td>
<td>Singular</td>
<td>Plural</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Dissimilar</td>
<td>Plural</td>
<td>Singular</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Dissimilar</td>
<td>Plural</td>
<td>Plural</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Similar</td>
<td>Singular</td>
<td>Singular</td>
<td>98.5%</td>
<td>1.5%</td>
<td></td>
</tr>
<tr>
<td>Similar</td>
<td>Singular</td>
<td>Plural</td>
<td>97.15%</td>
<td>2.85%</td>
<td></td>
</tr>
<tr>
<td>Similar</td>
<td>Plural</td>
<td>Singular</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Similar</td>
<td>Plural</td>
<td>Plural</td>
<td>35.05%</td>
<td>68.95%</td>
<td></td>
</tr>
</tbody>
</table>

the 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 the two nouns 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 panel) is somewhat different. First, processing times overall are slower (note the different y-axes in Figure 5.1). 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. It is worth noting that this is the opposite of what ACT-R predicts, and it is also the opposite of the results of Jäger et al. (2017)’s meta-analysis. Pilot simulations suggest that this ill-fit to previous data is sensitive to parameter settings, so a different way of setting the parameters might provide a better fit to the data. Finally, there is again a 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 results of the simple SOSP-TH model make the following predictions about reading times at the verb in the self-paced reading experiments below. First, there should be large main effects of semantic similarity and verb number such that the similar conditions and the plural-verb conditions are read more slowly. In the singular-verb conditions, we might also detect a smaller effect of N2 number, with the singular N2s leading to slower processing, especially in the similar condition. For the plural-verb condition, we might detect the opposite effect of N2 number, with plural N2 conditions being read longer. The interaction effects are relatively small and might be hard to detect if they are actually there in the human data.
5.3 Experiment 2A

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-TH and the RACE/A extension of ACT-R. Because the full design involves attempting to estimate a three-way interaction from human reading time data (which is 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, Colonna, & Rizzi, 2015), we decided to run an experiment with only grammatical (i.e., with singular verbs) first. This should reduce any noise due to shifting participant strategies and provide a first chance to evaluate whether the human data follow the predictions of SOSP-TH or RACE/A more closely.
5.3.1 Method

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., participants 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.

Materials

The 36 test sentences were constructed to be similar to those used in Barker et al. (2001). 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 (see (4)). Following Barker et al. (2001), we manipulated the semantic similarity of the N2 to the N1 and the number marking on 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 chose to use the past tense form of be, following Barker et al. (2001), who instructed participants to use the past tense. In addition, we reasoned that

---

5 We contacted the authors, but the original materials were no longer available, so they could not be used to more directly translate their experiment to comprehension.
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. 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 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 comprehension question. The comprehension questions for the test sentences all asked what the subject was, in effect; the correct answer was always the N1 (see (5)). The answers were presented as a two-alternative forced choice below the comprehension question. This type of comprehension question allows us to both test whether participants were able to build a correct parse and whether participants were more confused about the subject when the N1 and N2 were similar. 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). All sentences were distributed in a latin square design with four lists. All items are listed in the Appendix.

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 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 ((5)), with the order of the correct and incorrect answers randomized for each sentence.

**Analyses**

We analyzed the word-by-word reading times, comprehension question accuracy, and comprehension question response time using (generalized) linear mixed effects models (Bates, Maechler, et al., 2015). The fixed effects for all analyses included semantic similarity, N2 number, and their interaction. Factors were coded using deviation coding, with plural and semantically similar coded as +0.5 and singular and dissimilar coded as -0.5. We used the full random effects structure justified by the design whenever such models converged (Barr et al., 2013); otherwise we simplified the random effects structure following the backward stepping procedure of Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017) until a model converged without error. Any question responses or reading times with a response time less than 50ms or greater than 10s were excluded from analysis (affecting about 1% of the data), and all reading and response times were log-transformed before analysis to better approximate a normal distribution. Below, we report point estimates of the effects of interest (e.g., odds ratio of correct to incorrect responses or difference in reading time between conditions) along with 95% confidence intervals (CIs) to quantify the uncertainty around those point estimates.\(^6\) As Cumming (2014) argues,\(^6\)

---

\(^6\)For odds ratios, an effect is considered significant at the .05 level if the 95% CI excludes 1.0. Note that odds ratios are the exponential of the log-odds reported in the model output in R. A reading time effect would be considered significant if its 95% CI excludes zero.
this is both more informative than simply reporting the outcome of a significance test (e.g., reporting that effect of semantic similarity at the verb differed significantly from zero) 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. This emphasis also facilitates future meta-analyses and a more cumulative approach to psycholinguistics.

5.3.2 Results

Comprehension question accuracy

For the effect of semantic similarity, participants were about 1.375 times more likely to answer the question correctly in the dissimilar condition than in the similar condition (95% CI [1.014, 1.865]). There was no effect of N2 number (odds ratio: 0.995, 95% CI [0.764, 1.298]) or of the interaction (odds ratio: 0.726, 95% CI [0.429, 1.228]).

Comprehension question response times

For these analyses, only correct responses were analyzed. Also, a reduced random effects structure was used to allow the model to converge (random by-participants random intercepts and by-item random intercepts and slopes for similarity with the correlations between the random slopes and intercepts set to zero). For the effect of semantic similarity, participants answered questions more quickly in the dissimilar condition than the similar condition ($b = -0.039$, 95% CI [-0.069, -0.001]). For the number manipulation, participants were marginally faster in the the plural condition ($b = -0.021$, 95% CI [-0.043, 0.001]). There was no effect of the interaction ($b = 0.039$,
95% CI [-0.006, 0.083]).

**Reading times**

The analysis of reading times proceeded as follows. First, after removing trials in which participants answered the comprehension question incorrectly, very long (>10s) and very short (<50ms) reading times were removed. Then, a linear mixed effects model with by-participant random intercepts and fixed effects for log list position and word length in characters was fit to the log-transformed reading times to account for learning or fatigue and word-length effects (Bates, Kliegl, Vasishth, & Baayen, 2015; Ferreira & Clifton, 1986; Hofmeister & Vasishth, 2014). The main analyses were carried out using the residuals from this model. The residualized log reading times for the regions of interest are shown in Figure 5.2.

**N2** This model required a reduced random effects structure to converge (by-participants random intercepts and by-item random intercepts and slopes for semantic similarity with the correlation parameters set to zero). There were no effects at the N2: N2 number ($b = -0.003$ 95% CI [-0.027, 0.020]), semantic similarity ($b = -0.011$ 95% CI [-0.039, 0.016]), interaction ($b = -0.016$, CI [-0.062, 0.031]).

**Adverb** There were no effects at the adverb: N2 number ($b = 0.020$ 95% CI [-0.012, 0.051]), semantic similarity ($b = -0.020$ 95% CI [-0.066, 0.025]), interaction ($b = 0.38$, CI [-0.027, 0.103]).
Figure 5.2: Mean reading times with 95% CIs from Experiment 2A.
**Verb**  Participants read the verb faster in the dissimilar condition than in the similar condition ($b = -0.033$, 95% CI [-0.055, -0.010]), while there were no effects of N2 number ($b = 0.018$, 95% CI [-0.005, 0.041]) or the interaction ($b = -0.018$, 95% CI [-0.067, 0.032]).

**Spillover**  This model required a reduced random effects structure to converge (by-participants random intercepts and slopes for N2 number with the correlation parameters set to zero and by-item random intercepts). There was no effect of N2 number ($b = 0.009$, 95% CI [-0.015, 0.032]) or semantic similarity ($b = -0.012$, 95% CI [-0.034, 0.009]), but there was an interaction ($b = 0.051$, 95% CI [0.008, 0.094]): Reading times for plural N2s were longer than singular N2s in the dissimilar condition, but the opposite pattern held in the similar condition.

**5.3.3 Discussion**

This experiment was designed to test the diverging predictions of RACE/A and SOSP-TH in reading times at the verb. The RACE/A model predicts faster reading times in the semantically similar conditions compared to the dissimilar conditions, whereas the implemented SOSP-TH model produces slower reading times for the semantically similar conditions. In line with the predictions of SOSP-TH, we observed a significant 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 RACE/A. The results of the comprehension question accuracy and response
times are also consistent SOSP-TH. In line with previous findings (e.g., Villata et al., 2018), 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. Villata et al. (2018) suggest that these effects might be handled in an SOSP model if question answering involves reconstructing the links between the treelets used in the sentence. In sentences with much feature similarity between items, this predicts slower question answering times and lower accuracies (both of which we observed) because, as in online processing, competition between the similar nouns can often lead to slowdowns and the production of less-than-perfect structures.

In the reading times at the verb, the SOSP-TH 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. We found no evidence for slower reading times in the singular N2 conditions or for the interaction in the human data. However, despite the relatively large sample size, the confidence intervals for both effects are quite large, so we cannot dismiss SOSP-TH’s prediction until we have higher-precision estimates of the effects, e.g., from a meta-analysis or an experiment with an even larger sample size.

We note that this experiment failed to replicate Wagers et al. (2009)’s finding that having a plural N2 slows processing compared to a singular N2. These effects are quite small (Wagers et al. (2009) report slowdowns of approximately 15ms), so our failure to detect the effect might simply be a Type II error.

Going beyond what is directly predictable from the implemented model, we also observed an interaction in the spillover region following the verb. In the dissimilar
conditions, participants read the singular N2 condition faster than the plural N2 condition, opposite to what the model predicted at the verb. In the similar conditions, though, the pattern was reversed, with longer reading times in the singular N2 case. This pattern is consistent with what the model predicted at the verb and is similar to the one reported in Nicenboim et al. (2018). Nicenboim et al. 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 (all job titles), and the number on the two nouns intervening between the grammatical subject and the verb was manipulated. Currently, SOSP-TH does not predict spillover effects, but although spillover effects are commonly used in the literature to draw conclusions about processing in preceding sentence regions, there is no leading theory of why we should observe spillover effects at all. Given this theoretical uncertainty, we remain agnostic about the cause of the spillover interaction and focus on the results at the verb, which largely support SOSP-TH’s prediction of similarity-based slowdowns.

5.4 Experiment 2B

Overall, Experiment 2A provides support for SOSP-TH’s prediction of similarity-induced slowdowns at the verb. It included only grammatical test items, though.

---

7Cho, Szkudlarek, 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-TH seems limited. Unexpected or ill-fitting material can take extra time to integrate into the existing structure, but because SOSP-TH settles to some attractor after each word, it is not clear how effects could spill over to subsequent words. Each word is processed completely before moving on. Spillover effects are important and widespread, though, so the system should be developed further so that it can cover them.
Experiment 2B included both grammatical and ungrammatical sentences, allowing us to explore the full design commonly used in agreement attraction studies in comprehension.

5.4.1 Method

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). In total, 90 participants were entered into the analyses.

Materials

In addition to the grammatical items in (4) above, Experiment 2B included ungrammatical versions of those four conditions with plural verbs (6):

(6) a. *Dissimilar, N2 singular:*

The canoe by the cabin likely were damaged in the heavy storm.

b. *Dissimilar, N2 plural:*

The canoe by the cabins likely were damaged in the heavy storm.
c. *Similar, N₂ singular:*

The canoe by the kayak likely were damaged in the heavy storm.

d. *Similar, N₂ plural:*

The canoe by the kayaks likely were damaged in the heavy storm.

The number of test items was increased to 40 from 36, as we needed a number divisible by eight to accommodate the 2 x 2 x 2 design and to increase statistical power to observe any three-way interaction that might be present. The comprehension questions had the same format as Experiment 2A. We included 120 filler items, eighty of which had comprehension questions for which the correct answer was N₂. 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.

**Procedure**

The procedure was identical to that of Experiment 2A.

**Analyses**

The analyses followed the same procedures as Experiment 2A.
5.4.2 Results

Comprehension question accuracy

To test the effects of semantic similarity, N2 number, and verb number, three models were fit. The first model included the full design, with fixed effects for semantic similarity, N2 number, verb number, and all interactions. There were effects of N2 number ($e^b = 0.784$ (odds ratio of correct to incorrect), 95% CI [0.637, 0.966]), verb number ($e^b = 0.782$, 95% CI [0.6181527 0.900]), the interaction between semantic similarity and N2 number ($e^b = 0.613$, 95% CI [0.392, 0.959]), and the three-way interaction ($e^b = 2.996$, 95% CI [1.125, 7.979]). There was no effect of semantic similarity ($e^b = 1.289$, 95% CI [0.9612115 1.727]), the interaction between semantic similarity and verb number ($e^b = 0.736$, 95% CI [0.445, 1.216]), or the interaction between N2 number and verb number ($e^b = 1.031$, 95% CI [0.664, 1.599]).

To unpack the three-way interaction, two additional models were fit, one that included only grammatical (singular verb) sentences and one that included only ungrammatical (plural verb) sentences. The grammatical analysis revealed a main effect of semantic similarity, with participants more likely to answer the question correctly in the dissimilar condition ($e^b = 1.565$, 95% CI [1.047, 2.342]). There was no effect of N2 number ($e^b = 0.798$, 95% CI [0.580, 1.099]). There was a significant interaction, though, such that participants were more likely to answer a question correctly in the N2 singular condition than the N2 plural condition, but only for semantically dissimilar items ($e^b = 0.341$, 95% CI [0.171, 0.681]). In the ungrammatical analysis, there were no effects: semantic similarity ($e^b = 1.091$, 95% CI [0.778, 1.529]), N2 number ($e^b = 0.828$, 95% CI [0.634, 1.082]), interaction ($e^b = 1.114$, 1.082]).
95% CI [0.592, 2.096]).

**Comprehension question response times**

Only times from correct trials were used. The only effect in this analysis was the main effect of semantic similarity, which showed that response times were faster in the semantically dissimilar conditions ($b = -0.062$, 95% CI [-0.096, -0.0285]). There was no evidence for effects of the other fixed effects: N2 number ($b = -0.020$, 95% CI [-0.048, 0.008]), verb number ($b = 0.010$, 95% CI [-0.025, 0.046]), the semantic similarity-N2 number interaction ($b = 0.032$, 95% CI [-0.028, 0.093]), the semantic similarity-verb number interaction ($b = -0.022$, 95% CI [-0.076, 0.032]), the N2 number-verb number interaction ($b = -0.031$, 95% CI [-0.096, 0.033]), and the three-way interaction ($b = -0.038$, 95% CI [-0.156, 0.079]).

**Reading times**

As in Experiment 2A, only reading times between 50ms and 10s were included in the analysis. Reading times were log transformed and residualized as above before being entered into the analysis. The residualized log reading times are plotted in Figure 5.3.

**N2** This region required a simplified random effects structure: by-participant random intercepts and slopes for N2 number, the N2 number-verb number interaction, the semantic similarity-verb number interaction, and the three-way interaction, as well as by-item random intercepts and slopes for semantic similarity, the N2 number-semantic similarity interaction, and the N2 number-verb number interaction. The
Figure 5.3: Mean reading times with 95% CIs from Experiment 2B. The upper panel shows the ungrammatical (plural verb) conditions and the bottom panel the grammatical (singular verb) conditions. The whole item was *The canoe by the cabin(s)/kayak(s) was/were damaged in the heavy storm.*
random correlations were also set to zero. The only effect whose 95% CI did not include zero was verb number: participants read the N2 slower in the plural verb conditions, $b = 0.026$, 95% CI [0.004, 0.049]. As there is no reasonable theoretical reason for this and all conditions were identical up to this point, we interpret this result as a Type I error. There were no other effects: N2 number ($b = -0.021$, 95% CI [-0.044, 0.003]), semantic similarity ($b = 0.004$, 95% CI [-0.028, 0.035]), the N2 number by semantic similarity interaction ($b = -0.022$, 95% CI [-0.070, 0.027]), the N2 number by verb number interaction ($b = 0.013$, 95% CI [-0.040, 0.067]), the semantic similarity by verb number interaction ($b = -0.012$, 95% CI [-0.059, 0.034]), and the three-way interaction ($b = 0.071$, 95% CI [-0.020, 0.162]).

**Adverb** This model had the full random effects structure except that the by-participant random slope for the N2 number-semantic similarity interaction was removed, and the correlations were set to zero. No fixed effects had 95% CIs that did not include zero: N2 number ($b = 0.011$, 95% CI [-0.015, 0.036]), semantic similarity ($b = -0.003$, 95% CI [-0.035, 0.029]), verb number ($b = 0.993$, 95% CI [-0.027, 0.033]), the N2 number by semantic similarity interaction ($b = 0.010$, 95% CI [-0.051, 0.071]), the N2 number by verb number interaction ($b = -0.026$, 95% CI [-0.084, 0.032]), the semantic similarity by verb number interaction ($b = -0.002$, 95% CI [-0.051, 0.048]), and the three-way interaction ($b = -0.015$, 95% CI [-0.125, 0.095]).

**Verb** The analysis at the verb required a simplified random effects structure (by-participant random intercepts and random slopes by verb number and by-item random intercepts and slopes by semantic similarity, with the correlations set to zero for sets
of random effects). At the verb, the only effect was the interaction between semantic similarity and verb number such that the semantically dissimilar conditions were faster than the similar conditions in ungrammatical plural-verb sentences but not in singular verb sentences ($b = -0.050$, 95% CI [-0.095, -0.004]). The confidence intervals of all other fixed effects included zero: N2 number ($b = -0.005$, 95% CI [-0.028, 0.017]), semantic similarity ($b = -0.017$, 95% CI [-0.042, 0.008]), verb number ($b = 0.023$, 95% CI [-0.005, 0.051]), N2 number by semantic similarity interaction ($b = 0.021$, 95% CI [-0.024, 0.066]), N2 number by verb number interaction ($b = -0.026$, 95% CI [-0.071, 0.020]), and the three-way interaction ($b = 0.002$, 95% CI [-0.089, 0.092]).

**Spillover** The random effects structure for this model was reduced to by-participant random intercepts and slopes for N2 number, verb number, the N2 number-semantic similarity interaction, the N2 number-verb number interaction and the three-way interaction. The random effects also included by-item random intercepts and slopes for N2 number, semantic similarity, and all three two-way interactions. The random correlations were set to zero. The only effect was the main effect of verb number: the spillover word was read more slowly when the verb was plural compared to when it was singular ($b = 0.089$, 95% CI [0.058, 0.120]). There were no other effects: N2 number ($b = -0.026$, 95% CI [-0.054, 0.002]), semantic similarity ($b = -0.008$, 95% CI [-0.031, 0.015]), N2 number by semantic similarity interaction ($b = -0.042$, 95% CI [-0.094, 0.010]), N2 number by verb number interaction ($b = -0.044$, 95% CI [-0.089, 0.001]), semantic similarity by verb number interaction ($b = -0.002$, 95% CI [-0.048, 0.043]), and the three-way interaction ($b = 0.031$, 95% CI [-0.061, 0.124]).
5.4.3 Discussion

The results of Experiment 2B provide tentative support for some of the predictions of the implemented SOSP-TH model over those of the RACE/A extension of ACT-R. In the comprehension question accuracies for grammatical sentences, we replicated the result from Experiment 2A: Participants were about 50% more likely to answer questions correctly in the semantically dissimilar conditions than in the similar conditions. There was an interaction, however, between N2 number and semantic similarity such that accuracy was higher for singular N2s than plural N2s, but only for semantically dissimilar conditions, with no N2 number effect when the nouns were semantically similar. Neither cue-based retrieval (including RACE/A) nor SOSP can explain this effect: In both cases, we would expect worse performance because the singular number feature on both nouns should lead to competition. Given that the results of this experiment are generally noisy, this effect may also be spurious.

There are two effects to comment on from the reading time data. First, in the verb region, there was an effect of semantic similarity in the ungrammatical sentences such that the dissimilar conditions were read faster. This bears out SOSP-TH’s predictions, but it is puzzling why we did not observe this effect in the grammatical sentences like we did in Experiment 2A. It may be that reading an ungrammatical sentence caused participants to read more carefully, allowing the similarity effect to surface, but when reading a grammatical sentence, there was no cue to pay more attention, thus washing out any effects. Second, there was a main effect of verb number such that ungrammatical, plural-verb sentences showed slower reading times in the spillover region. This makes sense, as participants should be perturbed by
ungrammatical sentences, or it might simply be an effect of the markedness of plurals in general (Wagers et al., 2009). However, we did not replicate previous findings of a reduction in this ungrammaticality cost in the plural N2 conditions (Jäger et al., 2017; Lago et al., 2015; Pearlmutter et al., 1999; Wagers et al., 2009). While there is a numerical trend that suggests such a reduction, the confidence intervals for the relevant parameters included zero, and so our results cannot be viewed as strongly replicating that well-attested effect.

Overall, the results of this experiment are somewhat messy. It may be that the inclusion of ungrammatical items distracted participants or caused them to adopt a different strategy in how they approached this experiment from how they approached Experiment 2A (Franck et al., 2015). It is possible that the null and unexpected results we observed reflect reality; however, we know of no reasonable theory that would lead to the exact pattern of results observed, so we can only conclude that the results of Experiment 2B, such as they are, tentatively support the predictions of SOSP.

5.5 General Discussion

The two experiments presented here were designed to test the predictions the RACE/A extension to ACT-R and the implemented SOSP-TH model using sentences similar to those used in Barker et al. (2001)’s production study of agreement attraction. In self-paced reading, RACE/A 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-TH model predicts the opposite effect: Semantic similarity between the nouns should lead to competition-induced slowdowns in reading times at the verb. Experiment 2A clearly supports the prediction of SOSP-TH in this regard. Participants read the verb more slowly when the preverbal nouns were semantically related. The ungrammatical conditions of Experiment 2B also support SOSP-TH, but Experiment 2B did not replicate the similarity-based slowdown in the grammatical items. The comprehension question accuracies in Experiment 2A supported SOSP-TH: Participants were more accurate when the nouns were dissimilar. In both experiments, the comprehension question response times also supported SOSP-TH, with participants responding faster to questions about sentences with dissimilar nouns than with similar nouns, as predicted by the link reassembly account described in Villata et al. (2018). Overall, the human data clearly support the predictions of SOSP-TH better than RACE/A, although caution is warranted given the complex pattern of results, especially in Experiment 2B.

SOSP-TH is not alone in being able to account for the general pattern of results, though. Hofmeister and Vaisishth (2014), building on a feature overwriting approach by Nairne (1990), argue that features of words can be overwritten by features of subsequent words. Thus, when the parser tries to retrieve the first noun, it is less distinct from other words and therefore more difficult to retrieve. This approach also predicts the similarity-based slowdowns we observed in Experiments 2A and 2B. However, it requires an additional mechanism beyond what is needed for assembling syntactic structure in order to capture the results.
Villata et al. (2018) propose a different type of extension to cue-based retrieval for handling encoding effects: activation leveling. Activation leveling extends the fan effect in ACT-R—where activation is split among all chunks sharing a retrieval cue—to encoding features. If two chunks have similar encoding 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 encoding features are shared between multiple words, just like SOSP-TH and against the prediction of RACE/A. However, as Villata et al. (2018) argue, SOSP-TH’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 ACT-R that requires separate mechanisms for encoding and retrieval effects.

This is indeed SOSP-TH’s great strength: It provides a unifying account for different types of interference—retrieval and encoding—using a single mechanism. Both types of interference are predicted in SOSP-TH because all features that treelets carry are considered relevant to structure building; no distinction is made between, e.g., semantic features and syntactic features, because individual treelets are considered chunks of conceptual structure that can be pieced together in different ways. Thus, when they combine, the overall coherence of the resulting thought is what affects processing, not how separate processing modules interact.
Chapter 6

General discussion

The main goals of this dissertation were twofold: first, to develop, implement, and test a general-purpose mathematical theory of sentence processing, and second, to show that the model can provide an explanation for important timing effects from sentence processing experiments. This chapter reviews the results from the preceding chapters, assesses how well the goals of the dissertation were met in comparison with other theories, and outlines directions for future work.

6.1 Summary and assessment of results

6.1.1 Chapter 1: Two approaches to sentence processing

Chapter 1 contrasted two approaches to theories of sentence processing. The first approach, grammar-controlled theories, is based on the idea that the linguistic structures that the mind entertains during sentence processing are strongly constrained
by the rules of a symbolic grammar. In grammar-controlled theories, sentences are put together so that only perfectly grammatical structures result, both in incremental processing and as a final result at the end of a sentence. These theories (e.g., Crocker & Brants, 2000; Frazier & Fodor, 1978; Futrell & Levy, 2017; Hale, 2001, 2011; Jurafsky, 1996; Levy, 2008a, 2008b) have been largely successful in explaining important timing effects in sentence processing. However, they have struggled to provide parsimonious explanations for three central sentence processing phenomena. The first are local coherence effects—where people seem to entertain locally coherent structures that are ungrammatical in the context of the rest of the sentence (Bicknell et al., 2009; Cai et al., 2012; Konieczny, 2005; Konieczny et al., 2009; Levy et al., 2009; Paape & Vasishth, 2015). The second are agreement attraction effects, where a verb agrees in number (or gender) with a noun other than the grammatical subject (e.g., Bock & Miller, 1991; Dillon et al., 2013; Eberhard, 1997; Jäger et al., 2017; Lago et al., 2015; Pearlmutter et al., 1999; Wagers et al., 2009). Again, grammar-controlled theories are only somewhat successful here. The third are encoding interference effects, when properties of words that are not relevant for memory retrieval (in a particular context) interfere with how they are processed (e.g., Gordon et al., 2001; Hofmeister & Vasishth, 2014; Villata et al., 2018). These are also difficult to account for under leading grammar-controlled theories.

These phenomena have motivated the development of another class of psycholinguistic theories, self-organizing theories (beim Graben et al., 2008; Cho et al., 2018, 2017; Cho & Smolensky, 2016; Gerth & beim Graben, 2009; Kempen & Vosse, 1989; Smith et al., 2018; Smith & Tabor, 2018; Stevenson, 1994a; Stevenson & Merlo,
1997; Tabor & Hutchins, 2004; van der Velde & de Kamps, 2006; Vosse & Kempen, 2000, 2009). Self-organizing theories, while guided by symbolic grammars, are not constrained to only construct perfectly grammatical structures. Instead, larger-scale structures emerge through local interactions between lexically anchored syntactic treelets. These interactions usually lead to grammatical structures, but sometimes the system settles on a suboptimal structure instead. While self-organizing theories provide good coverage of many psycholinguistic phenomena, they have struggled to capture many timing effects, which are the main source data informing psycholinguistic theories. As self-organizing theories are typically specified in terms of ordinary differential equations, where linguistic variables are described in terms of how they change in time, this is a particularly glaring failure on the part of these self-organizing theories. Thus, the remainder of the dissertation sought to develop a self-organizing theory that can capture important timing effects, especially those related to local coherence, agreement attraction, and encoding interference effects.

6.1.2 Chapter 2: Introducing SOSP-TH

Chapter 2 (Smith & Tabor, 2018) served a number of purposes. First, it presented the equations of SOSP-TH. SOSP-TH is a stochastic gradient dynamical system with attractors that correspond to both well-formed (high-harmony) and ill-formed (low-harmony) linguistic structures. Next, we showed that the model predicts a simple relationship between processing times and the harmony of the structure being built. Specifically, processing times are inversely proportional to the harmony of the
structure. Because the processing dynamics in SOSP-TH are noisy, the average processing time over many trials is the weighted average of the settling times to each attractor that the system chooses in the different trials.

We additionally uncovered an interesting nonlinear relationship between average processing times and the ratio of the harmonies of two competing structures: When one structure has very low harmony compared to the other, processing times are fast because there is only weak competition between the structures, and the system can easily build the high-harmony structure. As the harmony of the lower-harmony structure increases, average processing times rise as the system starts building the low-harmony, and therefore slow-to-build, parse more often. But if the ratio of the lower harmony to the higher harmony passes about 0.8, average processing times begin to drop again. This is because both attractors are relatively well-formed and therefore fast to build, so even if the noise nudges the system to choose the lower-harmony structure, it does not pull the average time up very much.

Finally, Chapter 2 presented the first published self-organizing model of local coherence effects (Konieczny, 2005; Konieczny et al., 2009; Levy et al., 2009; Paape & Vasishth, 2015; Tabor et al., 2004). The model reproduced Tabor et al. (2004)’s report of increased reading times at the participle in ...smiled at the player tossed/thrown... for the tossed condition over the thrown condition, demonstrating that this effect, which has been argued to require self-organizing models, is indeed possible in an implemented SOSP-TH model. A major outcome of this chapter was showing that it

---

1We demonstrated this using a one-dimensional, one-attractor model, but simulations with larger models, e.g., the incremental local coherence model in Chapter 3, suggest that the pattern holds more generally.

2As discussed in Chapter 1, the noisy channel extension of surprisal (Futrell & Levy, 2017; Levy,
is possible to derive timing predictions for both local coherence effects and ambiguity advantage effects from the same model.

### 6.1.3 Chapter 3: From equations to full-fledged models

Chapter 3 provided additional details of the implementation of SOSP-TH with the goal of providing enough detail to use the equations introduced in Chapter 2 to create meaningful models of sentence processing, which was demonstrated in two models. This chapter highlighted the fact that SOSP-TH is a flexible framework, as it can implement a large variety of grammar formalisms, although the range of formalisms is strongly limited by the need to account for human processing data. The feature-based dependency grammar that we have used here provides a theoretically well-grounded set of representations for creating attractors of the system dynamics. Chapter 3 also discussed choices for default settings of the free parameters. Importantly, it also showed that the noiseless system is stable, i.e., for any initial condition, the system will eventually settle into one of its attractors and not oscillate or exhibit other dynamical behaviors. This is important because it is often argued that self-organizing models might work on a small scale, but when used to make larger, broader-coverage processing models, they might exhibit unrealistic behaviors. The proof at least shows that larger-scale models must behave coherently, although their dynamical behavior is limited to fixed points; oscillations and chaos, for example, are ruled out.

Chapter 3 also presented a model of classic agreement attraction (e.g., Bock & Miller, 1991) that captured the typical agreement production data from previous 2008b; Levy et al., 2009) also purports to account for local coherence effects. However, we argue that they cannot account for the full range of local coherence data.
studies while highlighting the importance of carefully encoding assumptions about linguistic structure for correctly predicting human sentence processing data. We needed to adjust this model to reflect markedness or frequency differences between N1[sg]-N2[pl]-V[pl] and N1[pl]-N2[sg]-V[sg] cases (Farkas & de Swart, 2010; Haskell et al., 2010; Haspelmath, 2006), but doing so allowed the model to reproduce the well-attested asymmetry in attraction rates. The second model in Chapter 3 was an incremental, word-by-word model of local coherence effects. The main contribution of this model was a proof of concept showing that SOSP-TH is capable of scaling up to multi-word sequences. The model drew attention to the fact that we need a better understanding of how linguistic structures shape the harmony surface and how less-than-perfect structures are handled because, without substantial modification, the incremental model failed to build the desired locally coherent structures. This model thus provides important cues to where to look to make the model work for incremental processing.

We note that Bicknell and Levy (2009)’s grammar-controlled noisy channel model, while not fully specified, provides an interesting comparison case to the incremental SOSP-TH model. This model allows locally coherent structures to form on the basis of short sequences of words, e.g., \textit{player tossed the frisbee} might be analyzed as a main sentence in \textit{the coach smiled at the player tossed the frisbee}. Processing difficulty arises when integrating these probabilistic beliefs about the local structure (a Bayesian prior) to posterior beliefs about the whole structure. This provides a tantalizing suggestion of a Bayesian bridge between self-organizing models and certain grammar-constrained models.
6.1.4 Chapter 4: Pseudopartitives

Chapter 4 presented the response time data from the verb selection experiment in Smith et al. (2018). Participants were asked to choose a singular or plural verb after reading a subject noun phrase (NP) belonging to one of four classes of English pseudopartitives (first noun (N1) type) and quantifiers: Containers (e.g., *a box of chocolates*), Collections (*a stack of sandwiches*), Measure Phrases (*a lot of newspapers*) or Quantifiers (*several pamphlets*). To test for effects of agreement attraction from the plural second noun (N2), we also included conditions that elided the N2, leaving only the singular N1 in the input. In the human response time data (Experiment 1), we found an interactive pattern between N2 presence and N1 type: With the exception of Containers, the -N2 conditions were consistently slower than the +N2 conditions, with the difference increasing from Collections to Measure Phrases to Quantifiers. An implemented SOPS-TH model of the verb choice provided a good fit to the human data. The model produces this behavior due to the changing balance in feature match for different conditions. These results (which are consistent with the original modeling results in Smith et al. (2018)) provide a better account for the human data than Marking and Morphing (Bock et al., 2001; Eberhard et al., 2005), surprisal theory (Levy, 2008a), and ACT-R (Lewis & Vasishth, 2005).

The model’s behavior is also largely consistent with the original timing results of Staub (2009), who initially developed this experimental paradigm for studying agreement attraction. Staub (2009) found that participants were slower to respond and more likely to pick the incorrect verb form when there was a number-mismatched distractor noun. This is consistent with competition between the fully grammatical
parse and a lower-harmony agreement attraction parse, as predicted by SOSP-TH. In addition, when participants in Staub (2009) made an agreement error when both nouns had the same number marking (e.g., N1[sg]-N2[sg]-V[pl]), response times were especially elevated, consistent with the parser settling on a very low-harmony structure. Thus, for agreement attraction, SOSP-TH covers both the new data of Smith et al. (2018), and it is also consistent with other timing data from the same task.

6.1.5 Chapter 5: Encoding interference

Chapter 5 investigated encoding interference effects in subject-verb agreement using self-paced reading (Experiments 2A and 2B). Specifically, we considered the effect of semantic similarity between the nouns in the subject NP. An implemented SOSP-TH model of reading times at the verb predicted a slowdown when the N1 and the N2 in the subject NP were semantically similar (canoe and kayak(s)) compared to when they were dissimilar (canoe and cabin(s)). An extension to ACT-R based on RACE/A (van Maanen & van Rijn, 2007; van Maanen et al., 2009, 2012) predicted the opposite results, i.e., faster reading times at the verb for similar nouns compared to dissimilar nouns.

To test these differing predictions, we ran two self-paced reading experiments examining the effects of semantic similarity, N2 number marking, and verb number marking. Experiment 2A included only grammatical sentences like The canoe by the cabin(s)/kayak(s) unsurprisingly was damaged in the heavy storm. Experiment 2B included grammatical and ungrammatical sentences like The canoe by the
cabin(s)/kayak(s) unsurprisingly were... Experiment 2A showed a clear effect of semantic similarity at the verb such that participants read the verb more slowly in the similar conditions than in the dissimilar conditions. The results of Experiment 2B were not as clear, as we only observed a slowdown for semantic similarity in the ungrammatical sentences. This might be due to the inclusion of ungrammatical items in the stimuli, which possibly induced a shift in participants’ strategy toward the task. On the whole, however, we see the results of the Experiments 2A and 2B as tentative support for SOSP-TH over the RACE/A extension to ACT-R.

For ungrammatical sentences, the SOSP-TH model incorrectly predicts that the semantically dissimilar, plural N2 condition should be processed slower than the semantically dissimilar, singular N2 condition. This is the opposite of the results of the meta-analysis of Jäger et al. (2017). ACT-R, by contrast, correctly predicts the human data. The SOSP-TH model showed some parameter sensitivity in this effect, so a different set of treelet features or free parameters might bring its predictions more in line with established results.

Still, SOSP-TH is well-equipped for handling encoding interference more generally because of the way structure building is distributed across the treelets that make up the structure. When a new word is perceived, it can change aspects of words that have come before, the structures already established, and even words and structures that come later. In this way, the treelets conspire together to interactively co-determine the final structure. This interactivity allows treelets to define each other, at least in part. In the case of the canoe/kayaks example, both nouns push the verb to have certain features which it would not carry in their absence, and,
by activating these features, the nouns interfere with the process of working out which of them becomes subject, leading to the observed slowdown effects. This contrasts with ACT-R, which allows interaction between chunks in memory, but disallows the type of co-definition that happens in SOSP-TH. Thus, ACT-R can only handle the reported encoding interference effects if additional mechanisms are added: Either the semantic features must be relevant for retrieval, features must be capable of being overwritten (Hofmeister & Vasishth, 2014; Nairne, 1990), retrieval cues must be associable with multiple items (Engelmann et al., under review), or the activations of semantically similar items must be made to level off Villata et al. (2018). However, these additional mechanisms go beyond what is strictly necessary for constructing a syntactic representation for a sequence of words. While further testing and a better understanding of SOSP-TH’s parameter- and structure-dependence are necessary, SOSP-TH is equipped to handle these effects using only its standard, highly-interactive structure-building mechanism.

6.2 Future directions

6.2.1 Accounting for human timing data

Garden paths Garden path effects are one of the most-studied sentence processing phenomena and have shaped psycholinguistic theory for decades (e.g., Altmann & Steedman, 1988; Bever, 1970; Crain & Steedman, 1985; Ferreira & Henderson, 1991; Frazier, 1978; Frazier & Fodor, 1978; Kimball, 1973; Paape & Vasishth, 2015; Warner & Glass, 1987), so it is critical to show that SOSP-TH can plausibly account for this
important class of timing effects. In a garden path, a temporarily ambiguous phrase (like the dog in (1)) is interpreted in a way that is incompatible with disambiguating material that comes later (e.g., attaching the dog as the direct object of scratched instead of as the subject of yawned). One important case of garden paths was reported in Ferreira and Henderson (1991). Ferreira and Henderson compared sentences like

(1)  
   a. While the boy scratched the dog yawned loudly.  
   b. While the boy scratched the big and hairy dog yawned loudly.  
   c. While the boy scratched the dog that Sally hates yawned loudly.

Compared to unambiguous control sentences (e.g., While the boy scratched the dog the girl yawned loudly), participants were less likely to rate the sentences in (1) as grammatical. This effect was even larger for (1-c) (compared to (1-a) and (1-b)), where the head of the ambiguous phrase is far from yawned, where it needs to attach for the sentence to be grammatical. (Similar effects in reading times are provided in, e.g., Tabor & Hutchins, 2004). Ferreira and Henderson (1991) argue that this because the incorrect parse has more time to entrench in (1-c) than the other conditions, making reanalysis more difficult (Arai & Nakamura, 2016; Bailey & Ferreira, 2003; Paape & Vasishth, 2015; Tabor & Hutchins, 2004; Warner & Glass, 1987, report similar findings). Tabor and Hutchins (2004)’s self-organizing model captures this effect because it will approach an attractor where the links are less susceptible to competition. It is likely that SOSP-TH will behave similarly: The incremental local coherence model in Chapter 3 (before modification) failed in part because previously established links were basically impervious to interference from subsequent competitor
links. This aspect of the model needs further work, though, because a complete model of sentence processing should be able to account for both local coherence and these garden path effects in the same model.

Another important property of garden pathing is the effect of lexical bias (e.g., Trueswell, Tanenhaus, & Kello, 1993). Using self-paced reading, Trueswell et al. (1993) compared reading times after verbs that preferred either an NP complement or a sentential complement, e.g., (2):

\[(2) \quad \text{a. NP complement: The student forgot the answer was in the back of the book.} \]
\[ \text{b. Sentential complement: The student hoped the answer was in the back of the book.} \]

After forgot, which is biased to take an NP complement, Trueswell et al. found that reading times were elevated after was, which disambiguates answer as being its subject. For the hoped condition, which is biased to take a sentential complement, they found that reading times were elevated immediately after the verb. These results suggest that the subcategorization preferences of verbs can affect whether and where a garden path effect is detected.

Currently in SOSP-TH, the way ambiguity is handled is to average the representations of the different forms and use that as the initial condition when a word is introduced. However, this predicts no difference between forgot and hoped in (2). Thus, this example will require a way of incorporating these (frequency-sensitive) preferences. One way of doing this in SOSP-TH is use a weighted (instead of un-
weighted) averaging of treelet features when an ambiguous word is introduced. This can be accomplished in a non-ad-hoc fashion using parameter fitting for the weights or by using learned feature vectors instead of hand-coded features (see the discussion below). As long as the trained weights or learned features encode the difference in preferences in (2), SOSP-TH should easily handle these effects, making further competition for the strongly frequency-sensitive, grammar-controlled surprisal theory (Hale, 2001; Levy, 2008a; Smith & Levy, 2013).

**Ambiguity advantage** An intuitive prediction of competition-based parsing theories like SOSP-TH is that stiffer competition between structures should lead to longer processing times. If two structures are competing strongly, it would seem that the system should get temporarily hamstrung between them, slowing processing times (see, e.g., Nicenboim et al., 2018). Indeed, Chapter 2 showed that exactly this happens for a range of local harmony values on the competing parses. However, Traxler et al. (1998) found the opposite effect in structures like (3) (see also Logačev & Vasishth, 2015, 2016; Swets, Desmet, Clifton, & Ferreira, 2008; van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2000, 2001).

(3)  

a. **High attachment:** The driver of the car that had the mustache was pretty cool.

b. **Low attachment:** The car of the driver that had the mustache was pretty cool.

c. **Ambiguous attachment:** The son of the driver that had the mustache was pretty cool.
They found a speedup in eye-tracking reading times when an adjunct could plausibly attach to more than one noun phrase (3-c) compared to unambiguous attachments (3-a) (high attachment) and (3-b) (low attachment). This effect is complicated by task effects (Logačev & Vasishth, 2015, 2016; Swets et al., 2008), but overall, these findings have been interpreted as strong evidence against competition-based parsing, and so they provide a critical test case for SOSP-TH as a general theory of sentence processing.

SOSP-TH seems to offer an explanation for this effect that makes crucial use of competition. Recall the finding from Chapter 2 where average settling times in a two-attractor model change as a function of the ratio of the lower harmony value to the higher harmony value. Settling times rise as the ratio increases from very low toward about 0.8. After the ratio passes 0.8, the average setting times begin to speed up compared to lower harmony ratios. This nonlinearity in SOSP-TH’s sensitivity to the local harmony parameters suggests a way of explaining ambiguity advantage effects: One attractor corresponds to attaching *that had the mustache* to the first noun in the sentence, and the other to attaching to the second noun. For (3-a) and (3-b), one attractor will have high harmony and the other will have lower harmony due to its poor semantic feature match. As long as the ratio of the lower to the higher harmony is close to 0.8, reading times should be slow for both of those conditions. For (3-c), both attachments are good semantic and syntactic feature matches, so as long the ratio of their harmonies is greater than 0.8 (and less than one), (3-c) should be processed faster than (3-a) and (3-b). Future work needs to actually implement an SOSP-TH model of ambiguity advantage effects to see if the simple one-word model
from Chapter 2 is an accurate approximation of a more full-featured system.

**Parasitic gaps** One area of linguistic import where self-organization might offer new insights is parasitic gaps, e.g., *What did the attempt to repair ___ ultimately damage ___?* (Phillips, 2006). Extracting a wh-word from the first gap (a so-called syntactic island) usually results in ungrammaticality (Ross, 1967). But when a gap in an island is accompanied by another gap outside the island, the sentence is rated as more acceptable. In sentence comprehension, people posit gaps whenever one is licensed (e.g., Stowe, 1986), which is consistent with Phillips’ finding that people posited a gap only when it could be rescued later on.

In the self-organizing approach, we can account for this by assuming that the parser always expects a licit gap incorporating *slash* features whenever it encounters a word that might be displaced. Slash features are a way of making the features of a moved element available at other locations in a sentence by passing the features through the structure until the relevant gap is reached (Gazdar, 1981). In the example above, inputting *what* would cause the system to approach attractors where the features of *what* are repeated in special dimensions in each slot until the gap is found. At this point, the features can interact with words near the gap. This type of feature passing would be constrained by the grammar to only pass the features along paths that exclude ungrammatical gaps. But if the system encounters an illicit gap, it must create a new, lower-harmony structure in which the slash features are projected back from the illicit gap to the extracted element. The system can do this because of the pattern-completion nature of its processing: There is a pattern corresponding to the extraction from the island; it just does not posit it until it needs to because of its low
harmony. In the case of parasitic gaps as above, if the system finds second gap after the ungrammatical first one, the overall structure is imperfect but relatively good because the slash chain can connect what with both the illicit and the allowed gap. If there is no additional gap, though, it must settle for a hard-to-construct, ill-formed structure. The slash mechanism has not yet been implemented, but in addition to providing an account for how parasitic gaps can form, it makes testable predictions for self-paced reading (specifically, a slowdown at the right boundary of the island due to the construction of the lower-harmony slash chain), warranting new human experiments in addition to the modeling.

6.2.2 Extending the model

Incremental model The incremental version of the model is perhaps the highest-impact future direction for the model. Getting the incremental model working would allow for large-scale comparisons with competing theories like ACT-R (Lewis & Vasishth, 2005) and surprisal (Hale, 2001; Levy, 2008a). Engelmann et al. (under review) recently demonstrated that an updated version of ACT-R has a good fit to the human data from seventy-seven published studies. Surprisal has been shown to predict many timing effects in sentence processing effects (Levy, 2008a), including word-by-word reading times over several orders of magnitude of frequency (Smith & Levy, 2013). However, comparisons with these models will require developing the theory of how new incoming words affect the structure that has already been established in SOSP-TH. This is a fundamental question in sentence processing and is a strong motivation for developing the theory explicitly enough to test against large
data sets.

**Ambiguity and underspecification**  Currently, lexical ambiguity is handled by initializing the state of the system to a point that is the unweighted average of the treelet specifications for the different forms. Pilot simulations suggest that this approach can lead to unrealistic competition-based slowdowns for ambiguous words. Moreover, it assumes that all ambiguous forms are equally likely to occur, which is obviously false (see also the discussion of lexical effects in garden paths above). A more flexible approach is underspecification (Egg, 2010): certain lexical features are left unspecified (e.g., set to zero) unless they are turned on via interactions with other words. This would further strengthen the self-organizing principles of the approach by allowing treelets to affect each other instead of being inert building blocks. It might also allow the model to account for the results of Patson and Husband (2015), who found that participants are likely to incorrectly recall the number marking on the subject when in the presence of a number-mismatched distractor noun. The added feature flexibility that the underspecification mechanism affords could, once a participant has decided on a verb number, allow the verb number to percolate back to the subject and change its number feature.

**Parameter fitting**  The values of the free parameters for the models reported here were chosen mainly by manually exploring the parameter space to find values that produce reasonable model behavior. For example, the SOSP-TH models reported here were constrained to use the same, fixed RBF width parameter $\gamma$. This was set according to the heuristic value recommended in Muezzinoglu and Zurada (2006) that
ensures that maximal-harmony peaks will remain distinct and not merge into a single intermediate harmony peak. While this approach is sensible and produces reasonable results, other parameter settings might produce a better fit to the empirical data or reveal new, testable predictions by making the system sensitive to particular, manipulable aspects of its input.

The manual, trial-and-error method of parameter setting, e.g., as used with the noise and proximity tolerance parameters, while common, makes replicating the results difficult because the decisions that led to the shared code are not usually made public, even when the code is shared (Dotlačil, 2018). A better approach is to use a general-purpose, replicable method of fitting the parameters. One way of doing this is approximate Bayesian computation (Toni, Welch, Strelkowa, Ipsen, & Stumpf, 2009), which can be used not only to get point estimates of optimal parameter values given the model and a data set but also for estimating the uncertainty around those point estimates. This approach would allow us to explicitly compare different parameter settings in a systematic way by training parameters in different ways on the same data set and then testing their performance on a test data set, e.g., using 80% of the data from an experiment to fit the model and then testing it on the remaining 20%. For example, we might test whether using a single $\gamma$ value for all RBFs and manipulating the $h_i$s produces better model behavior than allowing different $\gamma$s for each center and keeping the $h_i$s fixed in the agreement attraction model of Chapter 3. Thus, one application could be to test different ways of restricting the free variables.

**Learning** Currenty, SOSP-TH is not a learning model; the linguistic representations are designed by hand to reflect useful generalizations from linguistic theory.
However, as we saw in the simulations for local coherence effects, encoding interference, and classical agreement attraction effects, the feature specifications can have a large impact on the model’s behavior. One way of making the model less sensitive to researcher choices and have broader coverage would be to learn features using modern machine learning techniques. Zhao et al. (2014), Levy and Goldberg (2014), and Bansal (2015) use neural network techniques to develop dependency-relation-specific lexical feature vectors. Unlike other popular approaches to learning these word-embeddings (e.g., Mikolov et al., 2013), these approaches learn a feature vector associated with a word in particular, pre-specified dependency relationships with other words. Lexical features learned in this way could easily be plugged in to SOSP-TH as head and dependent feature vectors to provide more realistic, empirically derived measures feature similarity for the local harmony values. Since these learned features are sensitive to the frequency distribution of words in context, this would also be a natural way to implement frequency effects in SOSP-TH and empirically estimate the biases of ambiguous words and structures.

**Mathematical analysis**

Finally, the mathematical framing of SOSP-TH was chosen in part for its transparency and its simplicity. Stochastic gradient dynamical systems like SOSP-TH are well understood in physics and chemistry (Gardiner, 1985; Haken, 1983; Risken, 1989; Wales, 2003), and analytical techniques for understanding their behavior and generating new, testable predictions are available. For example, more exact processing time predictions might be derivable using first-passage time analyses, where we ask how
long, on average, a stochastic dynamical system takes to exit a region surrounding
an initial condition (Gardiner, 1985; Jacobs, 2010). More precise predictions about
the relative probabilities of building different parses might be derived from transient
solutions to the Fokker-Planck equation that describes the probability over states as
a function of time (Gardiner, 1985; Risken, 1989). Finally, disconnectivity graphs
(Wales, 2003; Wales et al., 1998) that visualize the structure of the high-dimension
harmony landscape could provide more insight into the types of linguistic representa-
tions that are necessary for the model to successfully capture human language
processing, a result that would also have implications for theoretical linguistics.

These explorations might reveal that the simple dynamical processing that SOSP-
TH is limited to (only approaching attractors) might be too restrictive. One alternative
dynamical regime would be to design the system so that it follows pre-defined paths
that travel close to attractors but never approach them directly. Words in the input
would bump the system at branch points, choosing between ambiguous structures
and allowing the system to approach attractors corresponding to partial or complete
parses that incorporate the new word. Such dynamical objects are called stable
heteroclinic channels, and they have been used to model sequential behavior, decision
making, and changes in emotion (Afraimovich, Tristan, Huerta, & Rabinovich, 2008;
Afraimovich, Zhigulin, & Rabinovich, 2004; Horchler, Daltorio, Chiel, & Quinn, 2015;
Muezzinoglu, Tristan, Huerta, Afraimovich, & Rabinovich, 2010; Rabinovich, Huerta,
& Laurent, 2008; Rabinovich et al., 2001).
6.3 Conclusion

As Smaldino (2016) notes, formal modeling exercises like the ones in this dissertation are sometimes criticized on the grounds that models are too dumb, that they oversimplify complex phenomena to the point of being useless for telling us anything other than how the model works. Smaldino agrees that, indeed, models are dumb and oversimplified. But that is precisely why we need them. They force us to identify, evaluate, and then argue for the assumptions we make about systems. This does not usually happen for verbal models, where assumptions can easily remain implicit in vague language, hidden in the dark corners of assumed common ground, or simply forgotten. But once the assumptions have been made clear and a model is running, we can test it to see how well it does at accounting for the phenomena it was designed for.

SOSP-TH does a good job on most of the phenomena we have set out to explain here—local coherence, agreement attraction, and encoding interference. Using a quite simple mathematical framework, the theory provides a parsimonious, unifying, and yet extensible account of a number of sentence processing effects that had previously required non-parsing mechanisms by doing just one thing: maximizing the well-formedness of structures via local interactions. But it sometimes fails under the simplest assumptions (e.g., the incremental local coherence model). That the model fails in some regards puts it in good company: Lewis and Vasishth (2005)’s ACT-R model and surprisal theory (Levy, 2008a), while they do many things well, make incorrect predictions for a number of important sentence processing effects, both ones that are central arguments for self-organization and ones that they ought to capture
according to verbal ACT-R and surprisal models (Engelmann et al., under review; Futrell & Levy, 2017; Jäger et al., 2017; Levy, 2008b; Vasishth, Mertzen, Jäger, & Gelman, 2018). For SOSP-TH, where the model fails, it highlights clear paths for new experimental, theoretical, and modeling research. SOSP-TH’s promising successes and useful failures therefore make it a strong basis on which to build a high-impact research program.
Appendix: Experimental materials

A.1 Experiment 1: Pseudopartitives

Items in the “other experiment” rows were included as a part of a different experiment reported elsewhere.

<table>
<thead>
<tr>
<th>N1 Type</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Containers</td>
<td>a sack of potatoes</td>
</tr>
<tr>
<td>Containers</td>
<td>a sack of rocks</td>
</tr>
<tr>
<td>Containers</td>
<td>a bin of baseballs</td>
</tr>
<tr>
<td>Containers</td>
<td>a bin of screwdrivers</td>
</tr>
<tr>
<td>Containers</td>
<td>a crate of tangerines</td>
</tr>
<tr>
<td>Containers</td>
<td>a crate of newspapers</td>
</tr>
<tr>
<td>Containers</td>
<td>a vase of cattails</td>
</tr>
<tr>
<td>Containers</td>
<td>a vase of daisies</td>
</tr>
<tr>
<td>Containers</td>
<td>a tube of balls</td>
</tr>
<tr>
<td>Containers</td>
<td>a tube of mints</td>
</tr>
<tr>
<td>Containers</td>
<td>a bottle of pills</td>
</tr>
<tr>
<td>Containers</td>
<td>a bottle of olives</td>
</tr>
<tr>
<td>Containers</td>
<td>a matchbox of needles</td>
</tr>
<tr>
<td>Containers</td>
<td>a matchbox of keys</td>
</tr>
<tr>
<td>Containers</td>
<td>a pack of matches</td>
</tr>
<tr>
<td>Containers</td>
<td>a pack of razors</td>
</tr>
<tr>
<td>Collections</td>
<td>a pile of pucks</td>
</tr>
<tr>
<td>Collections</td>
<td>a pile of shirts</td>
</tr>
<tr>
<td>Collections</td>
<td>a heap of blankets</td>
</tr>
<tr>
<td>Collections</td>
<td>a heap of tires</td>
</tr>
<tr>
<td>Collections</td>
<td>a stack of sandwiches</td>
</tr>
<tr>
<td>Collections</td>
<td>a stack of bricks</td>
</tr>
</tbody>
</table>

137
<p>| Collections | a group of llamas |
| Collections | a group of boxes |
| Collections | a collection of coins |
| Collections | a collection of figurines |
| Collections | a set of dvds |
| Collections | a set of cufflinks |
| Collections | a crowd of onlookers |
| Collections | a crowd of fans |
| Collections | a troop of preschoolers |
| Collections | a troop of scouts |
| Measure Phrases | a lot of postcards |
| Measure Phrases | a lot of newspapers |
| Measure Phrases | a number of bags |
| Measure Phrases | a number of applications |
| Measure Phrases | a couple of notepads |
| Measure Phrases | a couple of sponges |
| Measure Phrases | a handful of cookbooks |
| Measure Phrases | a handful of pens |
| Measure Phrases | a bunch of folders |
| Measure Phrases | a bunch of novels |
| Measure Phrases | a smattering of magazines |
| Measure Phrases | a smattering of raisins |
| Measure Phrases | a variety of mugs |
| Measure Phrases | a variety of wrenches |
| Measure Phrases | a ton of ties |
| Measure Phrases | a ton of whisks |
| Quantifiers | all pies |
| Quantifiers | all trucks |
| Quantifiers | many clocks |
| Quantifiers | many violets |
| Quantifiers | more chips |
| Quantifiers | more tables |
| Quantifiers | some paintings |
| Quantifiers | some thumbtacks |
| Quantifiers | several pamphlets |
| Quantifiers | several forks |
| Quantifiers | fewer people |</p>
<table>
<thead>
<tr>
<th>Quantifiers</th>
<th>fewer mosquitos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantifiers</td>
<td>both pans</td>
</tr>
<tr>
<td>Quantifiers</td>
<td>both dictionaries</td>
</tr>
<tr>
<td>Quantifiers</td>
<td>five baseballs</td>
</tr>
<tr>
<td>Quantifiers</td>
<td>five CDs</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a box with transistors</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a box with chocolates</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a jar with pickles</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a jar with capers</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a bucket with tees</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a bucket with mussels</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a can with beans</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a can with cashews</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a pan with biscuits</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a pan with hamburgers</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a pail with mushrooms</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a pail with pebbles</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a container with thumbtacks</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a container with crayons</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a chest with sweaters</td>
</tr>
<tr>
<td>Other experiment</td>
<td>a chest with collectibles</td>
</tr>
</tbody>
</table>
### A.2 Experiments 2A and 2B: Encoding interference

Test items:

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>The canoe by the cabin likely was damaged in the heavy storm.</td>
<td>What was damaged?</td>
</tr>
<tr>
<td>The canoe by the cabin likely were damaged in the heavy storm.</td>
<td>What was damaged?</td>
</tr>
<tr>
<td>The canoe by the cabins likely was damaged in the heavy storm.</td>
<td>What was damaged?</td>
</tr>
<tr>
<td>The canoe by the cabins likely were damaged in the heavy storm.</td>
<td>What was damaged?</td>
</tr>
<tr>
<td>The canoe by the kayak likely was damaged in the heavy storm.</td>
<td>What was damaged?</td>
</tr>
<tr>
<td>The canoe by the kayak likely were damaged in the heavy storm.</td>
<td>What was damaged?</td>
</tr>
<tr>
<td>The canoe by the kayaks likely was damaged in the heavy storm.</td>
<td>What was damaged?</td>
</tr>
<tr>
<td>The canoe by the kayaks likely were damaged in the heavy storm.</td>
<td>What was damaged?</td>
</tr>
<tr>
<td>The laptop near the stapler probably was dirty from years of sweaty fingers.</td>
<td>What was dirty?</td>
</tr>
<tr>
<td>The laptop near the stapler probably were dirty from years of sweaty fingers.</td>
<td>What was dirty?</td>
</tr>
<tr>
<td>The laptop near the staplers probably was dirty from years of sweaty fingers.</td>
<td>What was dirty?</td>
</tr>
<tr>
<td>The laptop near the staplers probably were dirty from years of sweaty fingers.</td>
<td>What was dirty?</td>
</tr>
<tr>
<td>The laptop near the computer probably was dirty from years of sweaty fingers.</td>
<td>What was dirty?</td>
</tr>
<tr>
<td>The laptop near the computer probably were dirty from years of sweaty fingers.</td>
<td>What was dirty?</td>
</tr>
<tr>
<td>The laptop near the computers probably was dirty from years of sweaty fingers.</td>
<td>What was dirty?</td>
</tr>
</tbody>
</table>
The laptop near the computers probably were dirty from years of sweaty fingers.

The van behind the sign evidently was rusty from last winter's road salt.

The van behind the sign evidently were rusty from last winter's road salt.

The van behind the signs evidently was rusty from last winter's road salt.

The van behind the signs evidently were rusty from last winter's road salt.

The van behind the truck evidently was rusty from last winter's road salt.

The van behind the truck evidently were rusty from last winter's road salt.

The van behind the trucks evidently was rusty from last winter's road salt.

The van behind the trucks evidently were rusty from last winter's road salt.

The pencil beside the backpack clearly was worn after a long semester.

The pencil beside the backpack clearly were worn after a long semester.

The pencil beside the backpacks clearly was worn after a long semester.

The pencil beside the backpacks clearly were worn after a long semester.

The pencil beside the pen clearly was worn after a long semester.

The pencil beside the pen clearly were worn after a long semester.

The pencil beside the pens clearly was worn after a long semester.

The pencil beside the pens clearly were worn after a long semester.

The cabinet by the window often was open on days when the children had visited.

What was dirty?

What was rusty?

What was rusty?

What was rusty?

What was rusty?

What was rusty?

What was rusty?

What was rusty?

What was rusty?

What was rusty?

What was rusty?

What was worn?

What was worn?

What was worn?

What was worn?

What was worn?

What was worn?

What was worn?

What was worn?

What was worn?

What was open?
The cabinet by the window often were open on days when the children had visited. What was open?
The cabinet by the windows often was open on days when the children had visited. What was open?
The cabinet by the windows often were open on days when the children had visited. What was open?
The cabinet by the cupboard often was open on days when the children had visited. What was open?
The cabinet by the cupboard often were open on days when the children had visited. What was open?
The cabinet by the cupboards often was open on days when the children had visited. What was open?
The cabinet by the cupboards often were open on days when the children had visited. What was open?

The painting above the vase unfortunately was ruined from a leak in the roof. What was ruined?
The painting above the vase unfortunately were ruined from a leak in the roof. What was ruined?
The painting above the vases unfortunately was ruined from a leak in the roof. What was ruined?
The painting above the vases unfortunately were ruined from a leak in the roof. What was ruined?
The painting above the portrait unfortunately was ruined from a leak in the roof. What was ruined?
The painting above the portrait unfortunately were ruined from a leak in the roof. What was ruined?
The painting above the portraits unfortunately was ruined from a leak in the roof. What was ruined?
The painting above the portraits unfortunately were ruined from a leak in the roof. What was ruined?

The skateboard by the halfpipe unsurprisingly was scratched after the heavy collision. What was scratched?
The skateboard by the halfpipe unsurprisingly were scratched after the heavy collision. What was scratched?
The skateboard by the halfpipes unsurprisingly was scratched after the heavy collision. What was scratched?
<table>
<thead>
<tr>
<th>Statement</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>The skateboard by the halfpipes unsurprisingly were scratched after the heavy collision.</td>
<td>What was scratched?</td>
</tr>
<tr>
<td>The skateboard by the scooter unsurprisingly was scratched after the heavy collision.</td>
<td>What was scratched?</td>
</tr>
<tr>
<td>The skateboard by the scooters unsurprisingly were scratched after the heavy collision.</td>
<td>What was scratched?</td>
</tr>
<tr>
<td>The skateboard by the scooter unsurprisingly was scratched after the heavy collision.</td>
<td>What was scratched?</td>
</tr>
<tr>
<td>The skateboard by the scooters unsurprisingly were scratched after the heavy collision.</td>
<td>What was scratched?</td>
</tr>
<tr>
<td>The magazine under the paperweight doubtlessly was mildewy from the flash flood.</td>
<td>What was mildewy?</td>
</tr>
<tr>
<td>The magazine under the paperweight doubtlessly were mildewy from the flash flood.</td>
<td>What was mildewy?</td>
</tr>
<tr>
<td>The magazine under the paperweight doubtlessly was mildewy from the flash flood.</td>
<td>What was mildewy?</td>
</tr>
<tr>
<td>The magazine under the paperweights doubtlessly was mildewy from the flash flood.</td>
<td>What was mildewy?</td>
</tr>
<tr>
<td>The magazine under the paperweights doubtlessly were mildewy from the flash flood.</td>
<td>What was mildewy?</td>
</tr>
<tr>
<td>The magazine under the catalog doubtlessly was mildewy from the flash flood.</td>
<td>What was mildewy?</td>
</tr>
<tr>
<td>The magazine under the catalog doubtlessly were mildewy from the flash flood.</td>
<td>What was mildewy?</td>
</tr>
<tr>
<td>The magazine under the catalog doubtlessly were mildewy from the flash flood.</td>
<td>What was mildewy?</td>
</tr>
<tr>
<td>The magazine under the catalogs doubtlessly were mildewy from the flash flood.</td>
<td>What was mildewy?</td>
</tr>
<tr>
<td>The magazine under the catalogs doubtlessly were mildewy from the flash flood.</td>
<td>What was mildewy?</td>
</tr>
<tr>
<td>The letter near the briefcase accidentally was forgotten after the long meeting.</td>
<td>What was forgotten?</td>
</tr>
<tr>
<td>The letter near the briefcase accidentally were forgotten after the long meeting.</td>
<td>What was forgotten?</td>
</tr>
<tr>
<td>The letter near the briefcases accidentally was forgotten after the long meeting.</td>
<td>What was forgotten?</td>
</tr>
<tr>
<td>The letter near the briefcases accidentally were forgotten after the long meeting.</td>
<td>What was forgotten?</td>
</tr>
<tr>
<td>The letter near the memo accidentally was forgotten after the long meeting.</td>
<td>What was forgotten?</td>
</tr>
<tr>
<td>The letter near the memo accidentally were forgotten after the long meeting.</td>
<td>What was forgotten?</td>
</tr>
<tr>
<td>The letter near the memos accidentally was forgotten after the long meeting.</td>
<td>What was forgotten?</td>
</tr>
<tr>
<td>The letter near the memos accidentally were forgotten after the long meeting.</td>
<td>What was forgotten?</td>
</tr>
<tr>
<td>The novel on the table regrettably was dusty since the library’s closure.</td>
<td>What was dusty?</td>
</tr>
<tr>
<td>The novel on the tables regrettably were dusty since the library’s closure.</td>
<td>What was dusty?</td>
</tr>
<tr>
<td>The novel on the table regrettably were dusty since the library’s closure.</td>
<td>What was dusty?</td>
</tr>
<tr>
<td>The novel on the tables regrettably were dusty since the library’s closure.</td>
<td>What was dusty?</td>
</tr>
<tr>
<td>The novel on the memoir regrettably was dusty since the library’s closure.</td>
<td>What was dusty?</td>
</tr>
<tr>
<td>The novel on the memoir regrettably were dusty since the library’s closure.</td>
<td>What was dusty?</td>
</tr>
<tr>
<td>The novel on the memoir regrettably were dusty since the library’s closure.</td>
<td>What was dusty?</td>
</tr>
<tr>
<td>The novel on the memoirs regrettably were dusty since the library’s closure.</td>
<td>What was dusty?</td>
</tr>
<tr>
<td>The fork by the wineglass probably was part of the fancy dinnerware set.</td>
<td>What was part of the dinnerware set?</td>
</tr>
<tr>
<td>The fork by the wineglass probably were part of the fancy dinnerware set.</td>
<td>What was part of the dinnerware set?</td>
</tr>
<tr>
<td>The fork by the wineglass probably were part of the fancy dinnerware set.</td>
<td>What was part of the dinnerware set?</td>
</tr>
<tr>
<td>The fork by the wineglass probably were part of the fancy dinnerware set.</td>
<td>What was part of the dinnerware set?</td>
</tr>
<tr>
<td>The fork by the spoons probably was part of the fancy dinnerware set.</td>
<td>What was part of the dinnerware set?</td>
</tr>
<tr>
<td>The fork by the spoons probably were part of the fancy dinnerware set.</td>
<td>What was part of the dinnerware set?</td>
</tr>
<tr>
<td>The fork by the spoons probably were part of the fancy dinnerware set.</td>
<td>What was part of the dinnerware set?</td>
</tr>
<tr>
<td>Original Statement</td>
<td>Modified Question</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>The fork by the spoons probably were part of the fancy dinnerware set.</td>
<td>What was part of the dinnerware set?</td>
</tr>
<tr>
<td>The folder on the desk never was emptied after the employee left.</td>
<td>What was emptied?</td>
</tr>
<tr>
<td>The folder on the desk never were emptied after the employee left.</td>
<td>What was emptied?</td>
</tr>
<tr>
<td>The folder on the desks never was emptied after the employee left.</td>
<td>What was emptied?</td>
</tr>
<tr>
<td>The folder on the desks never were emptied after the employee left.</td>
<td>What was emptied?</td>
</tr>
<tr>
<td>The folder on the notebook never was emptied after the employee left.</td>
<td>What was emptied?</td>
</tr>
<tr>
<td>The folder on the notebook never were emptied after the employee left.</td>
<td>What was emptied?</td>
</tr>
<tr>
<td>The folder on the notebooks never was emptied after the employee left.</td>
<td>What was emptied?</td>
</tr>
<tr>
<td>The folder on the notebooks never were emptied after the employee left.</td>
<td>What was emptied?</td>
</tr>
<tr>
<td>The snowboard beside the boot definitely was ready for the new winter season.</td>
<td>What was ready for winter?</td>
</tr>
<tr>
<td>The snowboard beside the boot definitely were ready for the new winter season.</td>
<td>What was ready for winter?</td>
</tr>
<tr>
<td>The snowboard beside the boots definitely was ready for the new winter season.</td>
<td>What was ready for winter?</td>
</tr>
<tr>
<td>The snowboard beside the boots definitely were ready for the new winter season.</td>
<td>What was ready for winter?</td>
</tr>
<tr>
<td>The snowboard beside the ski definitely was ready for the new winter season.</td>
<td>What was ready for winter?</td>
</tr>
<tr>
<td>The snowboard beside the ski definitely were ready for the new winter season.</td>
<td>What was ready for winter?</td>
</tr>
<tr>
<td>The snowboard beside the skis definitely was ready for the new winter season.</td>
<td>What was ready for winter?</td>
</tr>
<tr>
<td>The snowboard beside the skis definitely were ready for the new winter season.</td>
<td>What was ready for winter?</td>
</tr>
<tr>
<td>The shovel near the pail often was balanced in the corner of the tool shed.</td>
<td>What was balanced in the corner?</td>
</tr>
</tbody>
</table>
The shovel near the pail often were balanced in the corner of the tool shed.
The shovel near the pails often was balanced in the corner of the tool shed.
The shovel near the pails often were balanced in the corner of the tool shed.
The shovel near the rake often was balanced in the corner of the tool shed.
The shovel near the rake often were balanced in the corner of the tool shed.
The shovel near the rakes often was balanced in the corner of the tool shed.
The shovel near the rakes often were balanced in the corner of the tool shed.

The goblet behind the crown allegedly was adorned with sparkling jewels.
The goblet behind the crown allegedly were adorned with sparkling jewels.
The goblet behind the crowns allegedly was adorned with sparkling jewels.
The goblet behind the crowns allegedly were adorned with sparkling jewels.
The goblet behind the chalice allegedly was adorned with sparkling jewels.
The goblet behind the chalice allegedly were adorned with sparkling jewels.
The goblet behind the chalices allegedly was adorned with sparkling jewels.
The goblet behind the chalices allegedly were adorned with sparkling jewels.

The sneaker by the box certainly was soaked after the sudden thunderstorm.
The sneaker by the box certainly were soaked after the sudden thunderstorm.
The sneaker by the boxes certainly was soaked after the sudden thunderstorm.

What was balanced in the corner?
What was balanced in the corner?
What was balanced in the corner?
What was balanced in the corner?
What was balanced in the corner?
What was balanced in the corner?
What was balanced in the corner?
What was balanced in the corner?

What was adorned with jewels?
What was adorned with jewels?
What was adorned with jewels?
What was adorned with jewels?
What was adorned with jewels?
What was adorned with jewels?
What was adorned with jewels?
What was adorned with jewels?

What was soaked?
What was soaked?
<table>
<thead>
<tr>
<th>Sentence</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>The sneaker by the boxes certainly were soaked after the sudden thunderstorm.</td>
<td>What was soaked?</td>
</tr>
<tr>
<td>The sneaker by the sandal certainly was soaked after the sudden thunderstorm.</td>
<td>What was soaked?</td>
</tr>
<tr>
<td>The sneaker by the sandals certainly were soaked after the sudden thunderstorm.</td>
<td>What was soaked?</td>
</tr>
<tr>
<td>The sneaker by the sandals certainly were soaked after the sudden thunderstorm.</td>
<td>What was soaked?</td>
</tr>
<tr>
<td>The mug above the shelf rarely was used in the last few years.</td>
<td>What was rarely used?</td>
</tr>
<tr>
<td>The mug above the shelf rarely were used in the last few years.</td>
<td>What was rarely used?</td>
</tr>
<tr>
<td>The mug above the shelves rarely was used in the last few years.</td>
<td>What was rarely used?</td>
</tr>
<tr>
<td>The mug above the shelves rarely were used in the last few years.</td>
<td>What was rarely used?</td>
</tr>
<tr>
<td>The mug above the cup rarely was used in the last few years.</td>
<td>What was rarely used?</td>
</tr>
<tr>
<td>The mug above the cup rarely were used in the last few years.</td>
<td>What was rarely used?</td>
</tr>
<tr>
<td>The mug above the cups rarely was used in the last few years.</td>
<td>What was rarely used?</td>
</tr>
<tr>
<td>The mug above the cups rarely were used in the last few years.</td>
<td>What was rarely used?</td>
</tr>
<tr>
<td>The chair near the television quickly was sold at the garage sale.</td>
<td>What was sold?</td>
</tr>
<tr>
<td>The chair near the television quickly were sold at the garage sale.</td>
<td>What was sold?</td>
</tr>
<tr>
<td>The chair near the televisions quickly was sold at the garage sale.</td>
<td>What was sold?</td>
</tr>
<tr>
<td>The chair near the televisions quickly were sold at the garage sale.</td>
<td>What was sold?</td>
</tr>
<tr>
<td>The chair near the stool quickly was sold at the garage sale.</td>
<td>What was sold?</td>
</tr>
<tr>
<td>The chair near the stool quickly were sold at the garage sale.</td>
<td>What was sold?</td>
</tr>
<tr>
<td>Sentence</td>
<td>Question</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>The chair near the stools quickly was sold at the garage sale.</td>
<td>What was sold?</td>
</tr>
<tr>
<td>The chair near the stools quickly were sold at the garage sale.</td>
<td>What was sold?</td>
</tr>
<tr>
<td>The hammer under the board surprisingly was stolen from the construction site.</td>
<td>What was stolen?</td>
</tr>
<tr>
<td>The hammer under the board surprisingly were stolen from the construction site.</td>
<td>What was stolen?</td>
</tr>
<tr>
<td>The hammer under the boards surprisingly was stolen from the construction site.</td>
<td>What was stolen?</td>
</tr>
<tr>
<td>The hammer under the boards surprisingly were stolen from the construction site.</td>
<td>What was stolen?</td>
</tr>
<tr>
<td>The hammer under the saw surprisingly was stolen from the construction site.</td>
<td>What was stolen?</td>
</tr>
<tr>
<td>The hammer under the saw surprisingly were stolen from the construction site.</td>
<td>What was stolen?</td>
</tr>
<tr>
<td>The hammer under the saws surprisingly was stolen from the construction site.</td>
<td>What was stolen?</td>
</tr>
<tr>
<td>The hammer under the saws surprisingly were stolen from the construction site.</td>
<td>What was stolen?</td>
</tr>
<tr>
<td>The shed behind the hedge evidently was mangled after the autumn hurricane.</td>
<td>What was mangled?</td>
</tr>
<tr>
<td>The shed behind the hedge evidently were mangled after the autumn hurricane.</td>
<td>What was mangled?</td>
</tr>
<tr>
<td>The shed behind the hedges evidently was mangled after the autumn hurricane.</td>
<td>What was mangled?</td>
</tr>
<tr>
<td>The shed behind the hedges evidently were mangled after the autumn hurricane.</td>
<td>What was mangled?</td>
</tr>
<tr>
<td>The shed behind the shelter evidently was mangled after the autumn hurricane.</td>
<td>What was mangled?</td>
</tr>
<tr>
<td>The shed behind the shelter evidently were mangled after the autumn hurricane.</td>
<td>What was mangled?</td>
</tr>
<tr>
<td>The shed behind the shelters evidently was mangled after the autumn hurricane.</td>
<td>What was mangled?</td>
</tr>
<tr>
<td>The shed behind the shelters evidently were mangled after the autumn hurricane.</td>
<td>What was mangled?</td>
</tr>
</tbody>
</table>
The flag beside the bugle probably was meant for the military ceremony.
What was meant for the ceremony?

The flag beside the bugles probably was meant for the military ceremony.
What was meant for the ceremony?

The flag beside the pennant probably was meant for the military ceremony.
What was meant for the ceremony?

The flag beside the pennants probably was meant for the military ceremony.
What was meant for the ceremony?

The engine beside the signal probably was repaired in the early morning.
What was repaired?

The engine beside the signals probably was repaired in the early morning.
What was repaired?

The engine beside the caboose probably was repaired in the early morning.
What was repaired?

The engine beside the cabooses probably was repaired in the early morning.
What was repaired?

The screen by the keyboard typically was dusted on Friday afternoons.
What was dusted?
The screen by the keyboards typically was dusted on Friday afternoons.

What was dusted?

The screen by the keyboards typically were dusted on Friday afternoons.

What was dusted?

The screen by the monitor typically was dusted on Friday afternoons.

What was dusted?

The screen by the monitor typically were dusted on Friday afternoons.

What was dusted?

The screen by the monitors typically was dusted on Friday afternoons.

What was dusted?

The screen by the monitors typically were dusted on Friday afternoons.

What was dusted?

The banjo near the microphone luckily was undamaged after getting knocked over.

What was undamaged?

The banjo near the microphone luckily were undamaged after getting knocked over.

What was undamaged?

The banjo near the microphones luckily was undamaged after getting knocked over.

What was undamaged?

The banjo near the microphones luckily were undamaged after getting knocked over.

What was undamaged?

The banjo near the mandolin luckily was undamaged after getting knocked over.

What was undamaged?

The banjo near the mandolin luckily were undamaged after getting knocked over.

What was undamaged?

The banjo near the mandolins luckily was undamaged after getting knocked over.

What was undamaged?

The banjo near the mandolins luckily were undamaged after getting knocked over.

What was undamaged?

The flyer beneath the thumbtack crookedly was hung on the corkboard.

What was hung on the corkboard?

The flyer beneath the thumbtack crookedly were hung on the corkboard.

What was hung on the corkboard?

The flyer beneath the thumbtacks crookedly was hung on the corkboard.

What was hung on the corkboard?

The flyer beneath the thumbtacks crookedly were hung on the corkboard.

What was hung on the corkboard?
The flyer beneath the poster crookedly was hung on the corkboard.
The flyer beneath the poster crookedly were hung on the corkboard.
The flyer beneath the posters crookedly was hung on the corkboard.
The flyer beneath the posters crookedly were hung on the corkboard.

What was hung on the corkboard?

The pot by the knife unfortunately was placed at the edge of the counter.
The pot by the knife unfortunately were placed at the edge of the counter.
The pot by the knives unfortunately was placed at the edge of the counter.
The pot by the knives unfortunately were placed at the edge of the counter.
The pot by the kettle unfortunately was placed at the edge of the counter.
The pot by the kettle unfortunately were placed at the edge of the counter.
The pot by the kettles unfortunately was placed at the edge of the counter.
The pot by the kettles unfortunately were placed at the edge of the counter.

What was placed at the edge of the counter?

The lamp by the bunkbed clearly was visible in the nanny camera footage.
The lamp by the bunkbed clearly were visible in the nanny camera footage.

What was visible?
<table>
<thead>
<tr>
<th>What was visible?</th>
<th>The lamp by the bunkbeds clearly was visible in the nanny camera footage.</th>
</tr>
</thead>
<tbody>
<tr>
<td>What was visible?</td>
<td>The lamp by the bunkbeds clearly were visible in the nanny camera footage.</td>
</tr>
<tr>
<td>What was visible?</td>
<td>The lamp by the nightlight clearly was visible in the nanny camera footage.</td>
</tr>
<tr>
<td>What was visible?</td>
<td>The lamp by the nightlights clearly was visible in the nanny camera footage.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What was rarely seen?</th>
<th>The knickknack near the candle rarely was seen on the back of the shelf.</th>
</tr>
</thead>
<tbody>
<tr>
<td>What was rarely seen?</td>
<td>The knickknack near the candle rarely were seen on the back of the shelf.</td>
</tr>
<tr>
<td>What was rarely seen?</td>
<td>The knickknack near the candles rarely was seen on the back of the shelf.</td>
</tr>
<tr>
<td>What was rarely seen?</td>
<td>The knickknack near the candles rarely were seen on the back of the shelf.</td>
</tr>
<tr>
<td>What was rarely seen?</td>
<td>The knickknack near the trinket rarely was seen on the back of the shelf.</td>
</tr>
<tr>
<td>What was rarely seen?</td>
<td>The knickknack near the trinket rarely were seen on the back of the shelf.</td>
</tr>
<tr>
<td>What was rarely seen?</td>
<td>The knickknack near the trinkets rarely was seen on the back of the shelf.</td>
</tr>
<tr>
<td>What was rarely seen?</td>
<td>The knickknack near the trinket rarely were seen on the back of the shelf.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What was illuminated?</th>
<th>The motorcycle by the building barely was illuminated under the flickering street light.</th>
</tr>
</thead>
<tbody>
<tr>
<td>What was illuminated?</td>
<td>The motorcycle by the building barely were illuminated under the flickering street light.</td>
</tr>
<tr>
<td>What was illuminated?</td>
<td>The motorcycle by the buildings barely was illuminated under the flickering street light.</td>
</tr>
<tr>
<td>What was illuminated?</td>
<td>The motorcycle by the buildings barely were illuminated under the flickering street light.</td>
</tr>
</tbody>
</table>
The motorcycle by the moped barely was illuminated under the flickering street light.
What was illuminated?

The motorcycle by the moped barely were illuminated under the flickering street light.
What was illuminated?

The motorcycle by the mopeds barely was illuminated under the flickering street light.
What was illuminated?

The motorcycle by the mopeds barely were illuminated under the flickering street light.
What was illuminated?

The rope under the bucket evidently was splattered with green when the paint spilled.
What was splattered with green?

The rope under the bucket evidently were splattered with green when the paint spilled.
What was splattered with green?

The rope under the buckets evidently was splattered with green when the paint spilled.
What was splattered with green?

The rope under the buckets evidently were splattered with green when the paint spilled.
What was splattered with green?

The rope under the cord evidently was splattered with green when the paint spilled.
What was splattered with green?

The rope under the cord evidently were splattered with green when the paint spilled.
What was splattered with green?

The rope under the cords evidently was splattered with green when the paint spilled.
What was splattered with green?

The rope under the cords evidently were splattered with green when the paint spilled.
What was splattered with green?

The jet above the mountain suddenly was bathed in light as the sun rose.
What was bathed in light?

The jet above the mountain suddenly were bathed in light as the sun rose.
What was bathed in light?

The jet above the mountains suddenly was bathed in light as the sun rose.
What was bathed in light?

The jet above the mountains suddenly were bathed in light as the sun rose.
What was bathed in light?

The jet above the airplane suddenly was bathed in light as the sun rose.
What was bathed in light?

The jet above the airplane suddenly were bathed in light as the sun rose.
What was bathed in light?
The jet above the airplanes suddenly was bathed in light as
the sun rose.
The jet above the airplanes suddenly were bathed in light
as the sun rose.

What was bathed in light?

The basketball near the court clearly was abandoned during
the snowy winter.
The basketball near the court clearly were abandoned during
the snowy winter.
The basketball near the courts clearly was abandoned during
the snowy winter.
The basketball near the courts clearly were abandoned dur-
ing the snowy winter.
The basketball near the volleyball clearly was abandoned
during the snowy winter.
The basketball near the volleyball clearly were abandoned
during the snowy winter.
The basketball near the volleyballs clearly was abandoned
during the snowy winter.
The basketball near the volleyballs clearly were abandoned
during the snowy winter.

What was abandoned?

The donut by the plate typically was placed on the breakfast
table every day.
The donut by the plate typically were placed on the breakfast
table every day.
The donut by the plates typically was placed on the breakfast
table every day.
The donut by the plates typically were placed on the break-
fast table every day.
The donut by the bagel typically was placed on the breakfast
table every day.
The donut by the bagel typically were placed on the break-
fast table every day.
The donut by the bagels typically was placed on the break-
fast table every day.
The donut by the bagels typically were placed on the break-
fast table every day.

What was placed on the table?
The brush beside the mirror probably was dampened from the leaky faucet’s spray.

The brush beside the mirror probably were dampened from the leaky faucet’s spray.

The brush beside the mirrors probably was dampened from the leaky faucet’s spray.

The brush beside the mirrors probably were dampened from the leaky faucet’s spray.

The brush beside the comb probably was dampened from the leaky faucet’s spray.

The brush beside the comb probably were dampened from the leaky faucet’s spray.

The brush beside the combs probably was dampened from the leaky faucet’s spray.

The brush beside the combs probably were dampened from the leaky faucet’s spray.

The clock by the wall always was covered with sticky cobwebs.

The clock by the wall always were covered with sticky cobwebs.

The clock by the walls always was covered with sticky cobwebs.

The clock by the walls always were covered with sticky cobwebs.

The clock by the watch always was covered with sticky cobwebs.

The clock by the watch always were covered with sticky cobwebs.

The clock by the watches always was covered with sticky cobwebs.

The clock by the watches always were covered with sticky cobwebs.

The couch by the rug clearly was muddy from the dirty puppy’s paws.

The couch by the rug clearly were muddy from the dirty puppy’s paws.

What was dampened?

What was dampened?

What was dampened?

What was dampened?

What was dampened?

What was dampened?

What was dampened?

What was dampened?

What was covered with cobwebs?

What was covered with cobwebs?

What was covered with cobwebs?

What was covered with cobwebs?

What was covered with cobwebs?

What was covered with cobwebs?

What was covered with cobwebs?

What was covered with cobwebs?

What was muddy?

What was muddy?
The couch by the rugs clearly was muddy from the dirty puppy’s paws.
What was muddy?
The couch by the rugs clearly were muddy from the dirty puppy’s paws.
What was muddy?
The couch by the sofa clearly was muddy from the dirty puppy’s paws.
What was muddy?
The couch by the sofa clearly were muddy from the dirty puppy’s paws.
What was muddy?
The couch by the sofas clearly was muddy from the dirty puppy’s paws.
What was muddy?
The couch by the sofas clearly were muddy from the dirty puppy’s paws.
What was muddy?

Fillers:

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eliza received a gaudy gift from Erasmus wrapped in simple packaging.</td>
<td>Who gave the gift?</td>
</tr>
<tr>
<td>George, a clerk, discovered a cavern while camping with the explorer Jennifer.</td>
<td>Who was an explorer?</td>
</tr>
<tr>
<td>Ginny performed a classical piano sonata after she played the minuet.</td>
<td>When was the sonata performed?</td>
</tr>
<tr>
<td>Oliver went to a rock concert that was almost sold out by the time Bryne bought tickets.</td>
<td>Who bought tickets?</td>
</tr>
<tr>
<td>Victoria opened a fine red wine which she drank with Kyle last night.</td>
<td>Who opened the wine?</td>
</tr>
<tr>
<td>Katherine recommended a complicated board game that bored Ashley.</td>
<td>Who was bored?</td>
</tr>
<tr>
<td>Kyle made a paper boat in his origami class which he gave to Sasha.</td>
<td>Who got a present?</td>
</tr>
<tr>
<td>Denise excitedly planted a rosebush in Miriam’s honor last spring that she waters daily.</td>
<td>Who was honored?</td>
</tr>
<tr>
<td>Beth read a philosophical tome which Dmitri found in a dimly lit cafe.</td>
<td>Who found the book?</td>
</tr>
<tr>
<td>Olaf bought a cuddly brown teddy bear that he gave to his niece, Sarah.</td>
<td>Who was gifted a bear?</td>
</tr>
</tbody>
</table>
Mona received a new dollhouse from Suzanne that she likes to play with with her friends.

Benny furnished the renovated apartment with things from Bernie's old flat.

Paula recently bought a jogging stroller for Kelly that is very light.

Thomas shot a documentary film that Nicolas criticized as being too long.

Richie inexplicably created a new version of tag that has confused Russell.

Joseph moved his giant new television into his cramped living room.

After dinner, Ray brought the dirty dishes from the table to the soapy sink.

John quickly loaded the pickup truck with hay after sweeping out the remaining straw.

Manya replaced her favorite cassette tape with a CD from a garage sale.

Nora pushed an extra chair up to the table for the unexpected guest.

Karina arrayed the crackers and cheese neatly on a handful of plates for her guests.

Emily moved the cooked lasagna noodles from the boiling water into the dish.

After the players left, Brian gathered the hockey sticks, pucks, and masks.

Pietro carefully removed the broken bricks from the barrel in the courtyard.

Iris threw another magazine onto the growing stack on the end table.

Nicole stuffed her notebooks into her backpack by the pile of school newspapers.

On most Fridays, Charles mysteriously found a cookie on the corner of his desk.

The cows got their sodium by licking a block of salt the farmer left for them.

The more senior ex-members of the Strangled Canaries eventually appreciated silence.
Diamandus had lost his hair, and Burt wondered how it felt to go out in the rain.
Peony searched the ads in the paper every night in the hopes of finding escarole on sale.
The famished girl wanted chicken noodle soup but ordered grilled cheese instead.
Sheila’s doctor lauded her daily jogging, noting that he himself prefers rowing.
Tammy was so exhausted that she slept through her date with Karen.
The school bus was stranded at the top of the hill until the snowplow made its rounds.
With many mice marching to the sound of his drum, the pied piper bongoed toward the river.
The lawyer entered the courtroom in his best suit, avoiding eye contact with the staring jurors.
The silly birds perched upside down in an attempt to reach the last few hidden treats.
The cat the dog bit had been asking for it all day.
The computer simulation that the error interrupted had run all night.
The exams that the quiz last week replaced were much easier than expected.
A mouse that the raccoon hissed at in the shed seemed sick with its tiny coughs.
A dog that Marge saw in the park later waited outside a store for its owner.
The authors whose influential article Ethan cited have been discredited.
Some DVDs that Kayleigh borrowed were scratched when she opened the cover.
The chocolate chip cookies that Oliver baked beat the cake at the bake off.
All of the trees that Katie felled were ponderosa pines, perfect for completing her log cabin.
A pot that Rachel was carrying slid off the tray and onto the floor.

Who wondered?
What was sought?
What was ordered?
Who praised exercise?
Who got stood up?
What was making rounds?
What made noise?
Who stared?
What was hidden?
What was asking for it?
What was disruptive?
What replaced something?
What hissed?
What was in the store?
What was discredited?
What was scratched?
What was beaten?
What was under construction?
What fell to the ground?
Some rubber cement that Parker had used failed to hold the model together.  
The rug that the milkshake drenched was ruined after the impromptu ice cream party.  
Lucille looked on in horror as yet another party guest left without a gift bag.  
Kacie inquired if her new convertible was ready to be picked up from the dealership.  
Henry found that lewd jokes offended the prudish senator less than his political statements.  
His mouth full of rolls, Gerry asked if there was salt and pepper for the bland potatoes.  
Troy, having mastered air conditioning repair, gave up on plumbing.  
After the squirrel sniffed the acorn, it chomped it vigorously.  

Before Leslie ran for office, she served eight years in the parks department.  
Before starting the final, the nervous student whispered a pleading prayer.  
After changing the flat, Ron ran over another nail.  

Before preparing the romantic dinner, Wilfred put on some soothing music.  
After the storm blew it over, the tree seemed to wave goodbye to the blustery world.  
After he blew his nose, Jose wiped his puffy, mournful eyes.  

After Frodo left on his long, dark quest, he thought often of his beloved homeland.  
After the sun set, a lonesome cloud passed before the rising moon.  
Before chasing a brightly colored butterfly, Quentin fell into the mire.  
The trumpet cases were open, so the instruments fell out and got scratched.  
The book’s title was not decided until the manuscript had been reviewed by the editor.
The desktop had already snapped in half before the amateur wrestlers landed on it with a thud. When did it break?

Marissa’s headache got worse when the neighbors traded drumming for whistling. When did the headache worsen?

When the crate clashed against the rickety wall, it collapsed in an instant. When did it collapse?

Before listening to the skittering mice for an hour, Danea dreamt of tiny whiskers. When did she hear mice?

When the dishwasher stopped, a heavy silence filled the kitchen. When did the dishwasher end?

We could always recognize failure, but not always success. What could we always recognize?

There were several slices of pecan pie left, but the pumpkin pie was all gone. What had no leftovers remaining?

The two vectors were orthogonal because the sum of the products of their elements was zero. What was orthogonal?

The kindergartners liked the little girl who was brought a toy by her parents. Who received a toy?

We saw a movie about an artist who was painted a picture by his father on his deathbed. Who painted a picture?

The health officials pounced on a restaurant which was sent a shipment of salmon. Who got the salmon?

At the dinner party, I met a man who was allowed the pleasure of eating sweets by his doctor. Who allowed candy?

One expects a person who is forgiven their sins by their own god to have tolerance for weaknesses in others. Who should be tolerant?

An elderly gentleman addressed the woman who was given a beer by the hostess. Who was addressed?

Balthazar praised the professor who was given Swahili lessons by a graduate student. Who did the praising?

The manager watched a waiter who was served pea soup by a trainee. Who served the soup?

James entertained the children who were dyed Easter eggs by their teachers. Who colored the eggs?

The foreman yelled at a carpenter who was sawn a board by his buddy on the job site. Who sawed the board?

The preschool teacher congratulated the little boy who was sewn a hat by his grandmother. Who was congratulated?
The janitor chatted with the young man who was shown an apartment by his uncle.
The nurse admonished a student who nabbed a muffin for her friends from the dining hall.
The play centered around an innkeeper who was sung a verse by a travelling monk.
The hotel owner questioned a guest who was brought a drink by the bellboy.
The coach smiled at the player who was tossed a frisbee by the opposing team.
The anthropologist interviewed a woman who wove a shawl for her mother.
The FBI questioned a congressman who was mailed a letter by the activist.
The deliveryman teased the accountant who was given a coupon by her boss.
The prophet spoke of a man who planted a tree for his newborn daughter.
The brothers of the actress who shot herself on the balcony were under investigation.
The sister of the firemen who criticized themselves far too often painted the bedroom.
The secretaries of the salesman who amused himself quite a bit wrote a letter to the editor.
The father of the bride who embarrassed herself at the reception complained to the priest.
The aunt of the bishops who injured herself last summer was concerned about the infection.
The hostess of the mayors who complimented themselves constantly bothered the reporter.
The nephews of the waiter who hurt himself on the bicycle became angry about the incident.
The brothers of the seamstress who entertained herself most evenings read books.
The parents of the schoolgirl who burned herself the other day failed to be very careful.
The uncles of the nun who lost herself in thought disturbed the ceremony.
The children began to build sandcastles, but no moats, as soon as they arrived on the beach.

As the soldiers marched, toward the tank lurched an injured enemy combatant.

While Mary was mending, the sock she was working on fell off her lap.

Today, we had a meeting for the cognition class, but not for speech perception.

The carpet was pulled up in one corner, but the underlying wood floor was intact.

Nearly half of U.S. adults will have high blood pressure and low cholesterol under the new guidelines.

Statler found the performance impressive, but Waldorf could only make bad puns about the performer’s name.

All of the printers but none of the fax machines in the surplus room had been used since 1990s.

Michael found the milk frother the most superfluous of Jean’s coffee accoutrements.

Tabak made the computational model, but Fror was responsible for the syntax trees.

When the Martians invaded, the town near the decimated city descended into chaos.

One of the earpieces on the headphones could only play treble, but not bass tones.

He could see the plate of fresh lettuce by the wilted broccoli.

She could feel the shadows playing and the sun dancing on the outside of her closed eyelids.

When the Republicans introduced their tax bill, the Democrats knew their plan was doomed.
References


Barker, J., Nicol, J., & Garrett, M. F. (2001). Semantic factors in the production of
number agreement. *Journal of psycholinguistic research, 30*(1), 91–114.


structural uncertainty in gradient symbols. In *Proceedings of the 8th workshop on cognitive modeling and computational linguistics*.


Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject-verb agreement in English. *Journal of Memory and Language, 41*, 165


and Control, 8, 304–337.


In *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics on language technologies* (pp. 1–8). Association for Computational Linguistics. doi: 10.3115/1073336.1073357


Smaldino, P. E. (2016). Models are stupid, and we need more of them. In R. R. Vallacher, A. Nowak, & S. J. Read (Eds.), *Computational models in social psychology*. Psychology Press.


