COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2018) 1–32

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12591

A Self-Organizing Approach to Subject–Verb Number Agreement

Garrett Smith,[a] Julie Franck,[b] Whitney Tabor[c]

[a]*Department of Psychological Sciences, University of Connecticut*
[b]*Psycholinguistics Laboratory, University of Geneva*
[c]*Department of Psychological Sciences, University of Connecticut and Haskins Laboratories*

Received 21 August 2017; received in revised form 6 December 2017; accepted 8 December 2017

Abstract

We present a self-organizing approach to sentence processing that sheds new light on notional plurality effects in agreement attraction, using pseudopartitive subject noun phrases (e.g., *a bottle of pills*). We first show that notional plurality ratings (numerosity judgments for subject noun phrases) predict verb agreement choices in pseudopartitives, in line with the "Marking" component of the Marking and Morphing theory of agreement processing. However, no account to date has derived notional plurality values from independently needed principles of language processing. We argue on the basis of new experimental evidence and a dynamical systems model that the theoretical black box of notional plurality can be unpacked into objectively measurable semantic features. With these semantic features driving structure formation (and hence agreement choice), our model reproduces the human verb production patterns as a byproduct of normal processing. Finally, we discuss how the self-organizing approach might be extended to other agreement attraction phenomena.

Keywords: Sentence processing; Agreement attraction; Marking and Morphing; Notional plurality; Dynamical systems modeling

1. Introduction

A time-tested strategy in studying human sentence processing is to adopt a well-motivated theory of grammar and then find loci where, under carefully controlled experimental conditions, human behavior diverges from what would be expected under the grammar. One prominent example is the study of *agreement attraction*, which occurs when a word agrees in number or gender with a word other than its canonical controller,

Correspondence should be sent to Garrett Smith, Department of Psychological Sciences, University of Connecticut, Storrs, CT 06269. E-mail: garrett.smith@uconn.edu

as specifed by the well-motivated grammar. A common case is that of a verb agreeing in number with a noun other than the putative subject. Bock and Miller (1991) and many subsequent studies show that when participants are asked to complete a sentence starting with a singular first noun (N1) followed by a plural modifying noun (N2), they produce more plural verbs than with a singular N1 and a singular N2. For example, when provided with the subject NP *the key to the cabinets*, a participant might complete the sentence as *the key to the cabinets* ***are*** *on the table*. Why should participants make this verb agreement choice if they are faithfully following the rules of grammar?

One proposed explanation relies on the notional plurality of the subject NP, which is assumed to impact the agreement process on top of the normal grammatical rules. Notionally plural subject NPs denote multiple tokens of the object to which the NP refers. For example, Vigliocco, Hartsuiker, Jarema, and Kolk (1996) compared notionally singular Dutch subject NPs like *de kooi met de gorillas* (*the cage with the gorillas*), which refers to a single cage containing multiple gorillas, to notionally plural ones like *de handtekening op de cheques* (*the signature on the checks*), which implies that the same signature was repeated multiple times on different checks. They found higher rates of agreement attraction for notionally plural subjects, suggesting that reference to multiple objects might affect verb marking in spite of the morphologically singular N1 subject. Many studies have replicated this result, suggesting that notional plurality might play a key role in this class of agreement attraction effects (e.g., Eberhard, 1999; Foote & Bock, 2012; Humphreys & Bock, 2005; Vigliocco, Butterworth, & Garrett, 1996).

The Marking and Morphing model (Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001; Eberhard, Cutting, & Bock, 2005), the most prominent theory of agreement errors in processing and the only existing theory that addresses notional effects, treats notional plurality as a primitive: The notional plurality of a noun phrase is measured via a questionnaire that asks people to rate the notional number of the noun phrase. In this model, sentence production begins with the mental encoding of a message to be conveyed. Notional or semantic properties of this representation contribute a continuous-valued number (the measured notional plurality ratings)[1] to the syntactic representation of the subject NP as a whole in an initial "Marking" stage. After Marking, a second stage, "Morphing" occurs, during which semantic properties have no more influence. Instead, morphosyntactic properties of the constituent elements each exert an influence, modulated by their hierarchical position in the subject phrase. The Marking and Morphing biases are combined additively to produce a probability of choosing plural agreement. Notional plurality effects arise during the Marking stage, when a notionally plural N1 contributes a bias toward the plural end of the number value continuum, making plural agreement more likely.

We believe that Marking and Morphing theory is on the right track in recognizing the important role of conceptual and morphosyntactic factors in agreement processing, but there are a number of issues with the approach. First, notional plurality is not a theoretically defined notion; rather, it is defined empirically, based on participants' subjective judgments on whether a subject NP presented in isolation refers to one thing or more than one thing. Second, the way notional plurality affects sentence processing is not motivated by independent evidence or theoretical considerations. Eberhard et al. (2005) assume that a weighted sum of notional plurality values and morphological marking information jointly determine the likelihood of producing a plural. The weights for each are adjusted to fit the agreement attraction data and not otherwise constrained by theoretical considerations, preventing it from making predictions outside of its original, intended scope. Finally, the semantic/notional Marking process is an additional process on top of the mechanism for actually assembling a syntactic structure.

In this paper, we propose to unpack the black box of notional plurality and offer a one-mechanism account of agreement processing by adopting a different processing architecture. In *self-organized sentence processing* (SOSP), each perceived or produced word activates a treelet (a small piece of syntactic structure in memory; Fodor, 1998, 2017), and the treelets combine to form meaning-bearing tree structures (Kempen & Vosse, 1989; Stevenson, 1994a,b; Tabor & Hutchins, 2004; Van der Velde & de Kamps, 2006; Vosse & Kempen, 2000, 2009). Rather than treating semantic effects (Marking) and morphosyntactic effects (Morphing) as independent and sequential contributions to number marking, SOSP assumes that lexically anchored syntactic treelets bearing semantic and syntactic features interact continuously to build tree structures by forming attachment links. We present an implemented dynamical systems model that shows how the treelet interactions produce agreement patterns similar to the human data as a side-effect of the structure-building process. We illustrate this using the pseudopartitive constructions discussed next and speculate on extensions to other cases of agreement attraction in the General Discussion.

1.1. Pseudopartitives

To clearly illustrate how SOSP can explain notional plurality effects, we focus on a class of subject NP structures, pseudopartitives, for which there is evidence that two different grammatical structures with different agreement requirements compete with each other. Pseudopartives in English take the form *a N1 of N2*, where the N1 denotes a quantity or amount of the N2, for example, *a cup of sugar* or *a group of people*. Linguistic analyses (e.g., Deevy, 1999; Selkirk, 1977; Stickney, 2009) propose that phrases like *a lot of houses* and *a bunch of people*, arguably grammaticalized from phrases of measurement like *cup of* (Brems, 2003; Koptjevskaya-Tamm, 2001; Rutkowski, 2007), can either be headed by N1 (1) or by N2 (2) (here, we follow the structural proposal of Selkirk, 1977, adapting some labels)[2]:

(1) [NP(sg) [Det (sg) a] [N' [N_1(sg) bunch] [PP [P of] [NP(pl) [N_2(pl) people]]]]]
(2) [NP(pl) [DP [NP [Det a] [N_1(sg) bunch]] [Det (pl) of]] [N' (pl) [N_2(pl) people]]]

We sketch one of the motivations for the different structures here (additional syntactic evidence is provided in Selkirk, 1977; Deevy, 1999; and Stickney, 2009). Selkirk (1977) shows that words like *group* and *bunch*, which tolerate both singular and plural verb morphology, also show an alternation in number agreement with reflexive pronouns:

(3) a. That group of crazies really got itself in hot water, didn't it?
b. That group of crazies really got themselves in hot water, didn't they?

The reflexive pronoun must be bound by a noun that c-commands it (Büring, 2005). Since *itself* in (3a) agrees with *group* and not *crazies*, *group* must be in a position to c-command it. Similarly, *crazies* must c-command *themselves* in (3b). The only position available here that c-commands the reflexive is the head of the subject NP, thus providing evidence that both N1-headed and N2-headed structures are available.

We focused on three subtypes of pseudopartitives (N1 Types): "Containers" (*a bottle of pills*), "Collections" (*a stack of sandwiches*), and "Measure Phrases" (*a lot of postcards*), with plural Quantifiers as a control case (*several pamphlets*). We chose these classes because they a priori seemed to form a cline of increasing notional plurality from Containers through Collections to Measure Phrases and Quantifiers. Thus, under Marking and Morphing, one would expect the rate of plural usage to increase as the N1 type changes from Container through Collection, Measure Phrase, and Quantifier. Below, we report experiments that test these claims empirically. Before turning to the experimental investigation of pseudopartitives, we first describe the SOSP alternative to Marking and Morphing.

1.2. The self-organized sentence processing framework

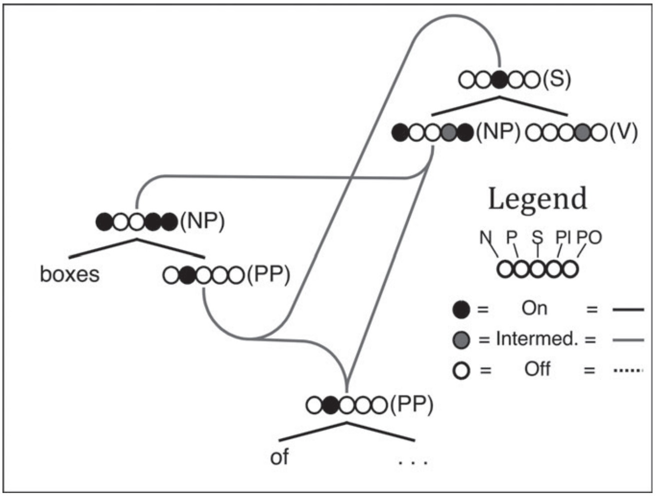
In this section, we outline the framework of SOSP, and later in the paper, we describe an implemented model of a part of this framework. SOSP is an instance of a self-organizing system, where bidirectional, local interactions between micro-elements give rise to coherent structure at the scale of the ensemble without any "leader" or external control (e.g., Haken, 1983). Examples include the alignment of magnetic spins in ferromagnetic materials (e.g., Solé, 2004), biological morphogenesis (Turing, 1952), and flocking behavior in birds (Reynolds, 1987). Models of language processing that emphasize the bottom-up construction of linguistic units such as the interactive activation model of letter and word perception (McClelland & Rumelhart, 1981) and TRACE (McClelland & Elman, 1986) also fall into this category.

Self-organized sentence processing builds on previous dynamical, self-organizing parsing models (Cho, Goldrick, & Smolensky, 2017; Kempen & Vosse, 1989; Kukona, Cho, Magnuson, & Tabor, 2014; Kukona & Tabor, 2011; Stevenson, 1994a,b; Tabor & Hutchins, 2004; Van der Velde & de Kamps, 2006; Vosse & Kempen, 2000, 2009). The basic units of the framework are lexically anchored syntactic treelets (see also Fodor, 1998, 2017) and links between attachment sites on the treelets. We adopt a dependency grammar formalism (e.g., Hudson, 2007; McDonald et al., 2013) for the treelets, which have feature vectors for a word itself as well as feature vectors for expected dependents, and the links, which embody the dependencies between words. In this formalism, each treelet can attach as a dependent of another treelet and take other treelets as dependents. As words are perceived or produced (the same process operates in both modes, as explained below), the associated treelets begin to form links with attachment sites on other treelets (see Fig. 1). The links have graded strengths and compete with each other on the basis of feature match between attachment sites (i.e., the strength of a link in which the feature bundles on its two attachment sites match well will grow more quickly than one for which they do not match well). In addition, there is noise in the link strengths, which makes the system sometimes settle into one configuration and sometimes into another if the feature matches for the competing links are similar, producing a probability distribution over final tree structures for ambiguous phrases and sentences.

Like the links, the feature values on the attachment sites are themselves able to change within limits specified by the lexical types. For example, the subject dependent of a generic verb treelet has a number value on a continuous scale, where the ends of the continuum code singular and plural. As the link between a number-marked subject noun phrase (e.g. *boxes*) and the subject attachment site on the verb increases in strength, the feature values at the two ends of the link tend to converge (see Fig. 1). Since the noun is marked plural, it stays fixed at the plural end of the continuum, and the subject attachment site moves continuously until it reaches the plural value. Due to linkages within the verb treelet, the expected marking on the verb itself simultaneously gravitates to the plural value. In this way, if the system first encounters a number marked subject (e.g., in an SVO language), it will expect a plural verb in comprehension and produce a plural verb in production. In general, the system implements feature passing similar to the kind employed in unification-based grammatical theories (e.g., Bresnan, 1982; Gazdar, Klein, Pullum, & Sag, 1985; Pollard & Sag, 1994) via the principle that features on opposite ends of a link converge in proportion to the strength of the link.

Models employing various subparts of this general framework have been implemented and shown to account for recency (or late closure) effects in attachment and binding preferences (Stevenson, 1994b), garden pathing, center embedding, subject- versus object-relative clauses, and predictive parsing (Kempen & Vosse, 1989; Vosse & Kempen, 2000, 2009),, and length effects or “digging-in” in garden paths (Tabor & Hutchins, 2004). These models also provide an explanation for processing effects that seem to flout traditional rules of grammar, such as local coherence effects (Konieczny, 2005; Paape & Vasishth, 2016; Tabor, Galantucci, & Richardson, 2004; see also Levy, Bicknell, Slattery, & Rayner, 2009). For example, Tabor et al. (2004) studied sentences like *The coach smiled at the player tossed the frisbee*. The string *the player tossed the frisbee* could stand alone as a main clause with the verb *tossed*, but in the context of the rest of the sentence, this structure should be ruled out: The preposition *at* cannot take a sentence as its complement. Despite this, participants showed slower reading times at *tossed* compared to *thrown* in (...) *the player thrown the frisbee*, since *thrown* is not compatible with the locally coherent string, suggesting competition between the incorrect locally coherent structure and the correct globally coherent structure. Most theories of parsing are guided in a top-down fashion by a well-motivated grammar, so they never build the structures that SOSP claims produce the interference in local coherence effects (e.g., Eberhard et al., 2005; Gibson, 1998; Hale, 2001, 2011; Levy, 2008; Lewis & Vasishth, 2005).[3] Given that SOSP naturally accounts for these cases of grammar-flouting interference, it is natural to extend the approach to agreement attraction, another case where the processing system seems to violate the rules of a plausible grammar.

(A)


(S)
(NP)
(V)
(NP)
boxes
(PP)
(PP)
of
. . .
Legend
N P S Pl PO
= On =
= Intermed. =
= Off =

(B)

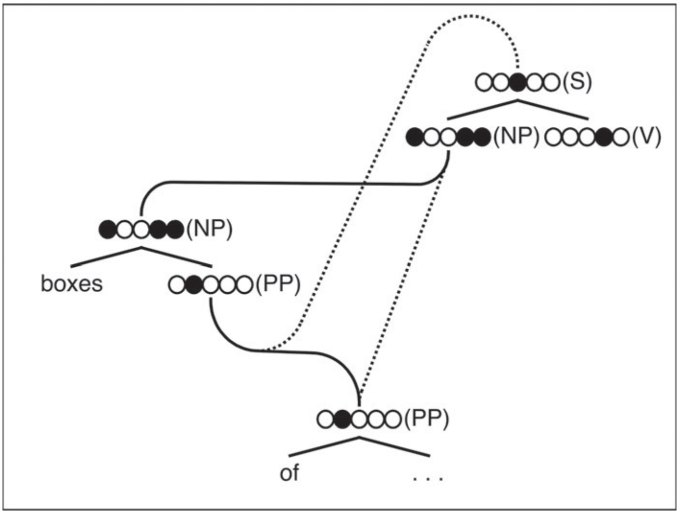

(S)
(NP)
(V)
(NP)
boxes
(PP)
(PP)
of
. . .

Fig. 1. Example of treelet interactions. The curved links can change their strengths; the straight lines show fixed relations within treelets that do not change. The circles represent features on attachment sites. Feature abbreviations: N = Noun, P = Preposition, S = Sentence, Pl = Plural, PO = Physical Object. For ease of viewing, only a proper subset of the features is shown. The features include syntactic features (e.g., N) and semantic features (e.g., PO). Initially (A), all possible links are activated at a low level and the features on opposite ends of links do not perfectly match. After settling (B), the links that have better matches have achieved near-maximal activation, while the links that have worse matches have approached zero. For the links that have grown strong, the feature values on opposite ends have converged to the same values.

1.3. Roadmap

Our pseudopartitive classes lie, intuitively, on a notional plurality cline. To test this, we first normed our pseudopartitive noun phrases to determine their notional plurality values in the standard sense (Experiment 1). Most of the previous notional agreement effects were found in production, so in Experiment 2, we tested the agreement preferences of these noun phrases using the forced-choice production paradigm of Staub (2009) and related these to the notional judgments of Experiment 1. Our goal with Experiments 1 and 2 was to establish that pseudopartitives fall in the realm of phenomena that Marking and Morphing theory is concerned with. Our goal is not to reject Marking and Morphing —indeed, SOSP agrees with it in the cases at hand—but rather to show how the self-organizing framework offers advantages in terms of simplicity of mechanism and insight into assumptions made by the classical theory. So in Experiment 3, we establish the basis for our proposed alternative model via a package of judgment tests that probe a set of fine-grained semantic features relevant to characterizing our N1 Types at the theoretical level. We then describe a self-organizing model based on the feature values derived from Experiment 3 that parses the pseudopartitive stimuli, generating a distribution of parses aligned with the results of Experiment 2. We conclude by addressing the question of how our approach may extend to other cases of agreement attraction and notional effects.

2. Experiment 1: Notional plurality norming

The purpose of Experiment 1 was to test for systematic variability in the notional plurality of pseudopartitive subject NPs. We expected, based on our intuitions as native speakers, that the notional plurality ratings would increase across the four N1 Types (Containers, Collections, Measures, Quantifiers), providing a basis for prediction of an effect of N1 Type on the rate of plural production in Experiment 2.

2.1. Method

2.1.1. Participants

We recruited 20 participants via Amazon Mechanical Turk and paid each one \$1.50 for participating (www.mturk.com; Gibson, Piantadosi, & Fedorenko, 2011; Sprouse, 2011). We included only participants who reported speaking English as a native language and whose IP addresses were located in the United States.

2.1.2. Design and materials

For each N1 Type, we selected four N1s or Quantifiers and paired each one with two different N2s for a total of 64 critical items. In addition to the four N1 Types we focused on here, we also tested eight items (each with two lexical variants) from an experiment reported elsewhere that included subject NPs of the form *a N1 with N2*, where the N1 was a different Container than the ones used for the present experiment. The *with*-Containers were not included in the analyses reported here. The whole set of materials was divided into two 40-item randomized lists with one lexical variant of the N2 for each N1 Type in a different list. Each participant was assigned at random to one of the lists. The full set of materials is given in Appendix S1.

After reading the information sheet and instructions (see Appendix S1), participants were presented with one subject NP at a time in the center of the screen with buttons for "one thing" or "more than one thing" below the subject NP. After completing the survey, participants received a completion code to enter on Mechanical Turk to receive payment.

2.2. Results

One observation was removed because the participant did not enter a response; the remaining data were included in the analyses. In contrast to Bock, Carreiras, and Meseguer (2012),[4] we used mixed effects logistic regression[5] to analyze the binary ratings of the notional plurality norming (Bates, Maechler, Bolker, & Walker, 2014; Jaeger, 2008). Participants' ratings (coded 0 = "one thing," 1 = "more than one thing") were entered as a function of the N1 Type (Container, Collection, Measure, Quantifier). N1 Type was coded using backward difference coding: The mean of each level was compared to the mean of the previous level. A likelihood ratio test showed that the effect of N1 Type was significant ($\chi^2(3) = 19.889$, $p < .001$). Collections were significantly more likely than Containers to be rated "more than one thing" ($b = 3.357$, 95% CI [1.786, 4.928], $p < .001$), and Measure Phrases were rated marginally more likely than Collections ($b = 1.848$, 95% CI [−0.092, 3.788], $p = .062$). There was no significant difference between Quantifiers and Measures ($b = 1.201$, 95% CI [−1.584,9 3.986], $p > .1$). Converting the fitted log-odds (logits) to probabilities, participants chose "more than one thing" with the following probabilities: Containers 0.331, Collections 0.756, Measures 0.906, and Quantifiers 0.913, showing a stepwise increase in the probability of choosing "more than one thing" across the first three levels of N1 Type. The data analysis scripts for this and the other experiments are available at https://github.com/garrett-m-smith/.

2.3. Discussion

In Experiment 1, we observed a stepwise increase in the notional plurality ratings for the first three levels of N1 Type (marginally from Collections to Measure Phrases).

Marking and Morphing predicts that this pattern of notional plurality ratings should produce increasing rates of plural verb agreement across the three types. As the notional plurality of the subject NP increases, it can exert a stronger force in the Marking phase toward making the subject NP plural overall, and this should increase the rate of plural verb agreement. In line with our first goal of showing that our pseudopartitive structures are an instance of what Marking and Morphing is meant to explain, Experiment 2 tested this prediction.

3. Experiment 2: Verb choice experiment

Most studies of agreement attraction compare a number mismatch condition (e.g., *the label*[sg] *on the bottles*[pl]) to a control condition with two singular NPs (e.g., *the label* [sg] *on the bottle*[sg]). Because the singular-singular condition is not available in the pseudopartitive (**a box of orange*) and because we wanted to manipulate both a Marking factor and a Morphing factor, we adopted a different control condition in which the modifying PP containing the N2 was elided. To make this elision felicitous, we included a context sentence before all items, for example, *Do we have anything to juggle around here? A tube of balls is by the tennis racket* and *Do we have anything to juggle around here, like balls? A tube is by the tennis racket*. The condition without the PP (−N2 condition) provided a baseline for the agreement preferences of the N1 itself. The +N2 condition, by contrast, makes any influence of the N2 on the verb choice apparent. Including both the factors N1 Type and +/−N2 allowed us to test both the Marking and Morphing components of the Marking and Morphing theory.

3.1. Method

Instead of using the typical sentence completion task, we used a task first used by Staub (2009). In Staub's task, the words prior to the critical verb are presented on a computer screen using rapid serial visual presentation (RSVP). Then, both singular and plural versions of the critical verb are presented, and the participant must choose between them by pushing a button. Although this task is not as close to natural production as the commonly used method of having participants read a preamble, repeat it, and then invent a completion of the sentence, it has been shown to replicate standard sentence-completion results of increased latencies for a plural N2 compared to a singular N2 and the structural hierarchy effect (Franck, Vigliocco, & Nicol, 2002; Haskell & MacDonald, 2003; Staub, 2009, 2010). The verb selection task also has the advantage of forcing a choice between the two verb forms; thus, no data are lost from participants using uninflected or otherwise unusable verbs. To our knowledge, this method has not been used to test for effects of notional plurality.

3.1.1. Participants

Fifty-seven University of Connecticut undergraduates took part for course credit.

3.1.2. Design and materials

A subset of the critical subject NPs from Experiment 1, embedded in complete sentences and preceded by context questions, were used in the present experiment along with additional items from each N1 Type. For each N1 Type, we used eight different lexical variants of the N1. We used a 2 × 4 design, crossing +/−N2 with N1 Type (Containers, Collections, Measures, or Quantifiers), resulting in 64 total critical items. Sixty-four filler sentences were also included. The materials were divided into two lists with sixteen critical items and 32 fillers and counterbalanced for +/−N2, N1 Type, verb tense (past or present), and the number of fillers taking singular and plural agreement. All materials are listed in Appendix S1.

3.1.3. Procedure

The experiment was carried out using E-Prime® software (version 2.0, Schneider, Eschman, & Zuccoloto, 2012). After giving informed consent, participants sat at a computer and read the instructions on the screen. Four practice items were presented before the actual experiment. Thereafter, the context question was presented in its entirety in the center of the screen (see examples (4a) and (4b)). When ready, participants pressed the “1” button on the number pad of a keyboard to go to the test sentence. After a fixation cross (1,000 ms), the test sentence was presented in one- or two-word chunks in the center of the screen. Each chunk was presented for 250 ms followed by 150 ms of blank screen. When the chunk containing the verb came, both the singular and plural verb forms (*is* and *are* or *was* and *were*) were presented side by side. Singular was always on the left and plural was always on the right. Participants chose the verb form they thought fit the sentence best using the “1” or “3” buttons on the number pad. Participants were instructed to enter their responses as quickly as possible without sacrificing accuracy. After the verb choice, the rest of the sentence continued in RSVP. The next trial began after 1,000 ms of blank screen. There was a break halfway through the experiment. In the following two examples, the slashes show where test sentences were broken into chunks.

(4a) +N2 Condition: Do we have anything to juggle around here?
+ / A tube / of / balls / [VERB CHOICE] / by the / tennis racket.

(4b) −N2 Condition: Do we have anything to juggle around here, like balls?
+ / A tube / [VERB CHOICE] / by the / tennis racket.

3.2. Results

Responses that differed more than three standard deviations from each participant’s average log reaction time[6] were excluded, resulting in the loss of 11 data points (about 1% of the total). The verb choice results were analyzed using mixed effects logistic regression with the dependent measure being the log odds of choosing a plural verb. Factors were coded using numerical sum contrasts (−1 or 1). The full model included fixed effects for +/−N2 and N1 Type and their interaction in addition to the maximal random effects

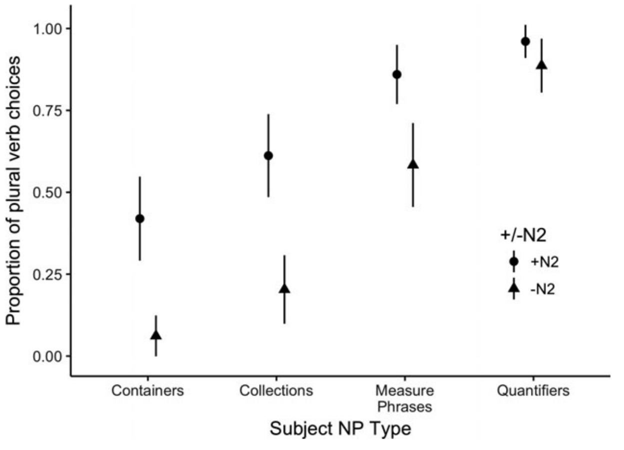


Fig. 2. Mean probability of choosing a plural verb in Experiment 2. Error bars represent subject 95% confidence intervals.

structure[7] (by-participant random intercepts and slopes for +/−N2, N1 Type, and their interaction and by-item random intercepts and slopes for +/−N2; Barr, Levy, Scheepers, & Tily, 2013). Significance tests were done using likelihood ratio tests comparing the model with an effect of interest to one that differed only in its exclusion of that effect.

The proportion of trials (with 95% confidence intervals) when a plural verb was chosen are plotted in Fig. 2. The main effect of N1 Type was significant ($\chi^2(3) = 74.875$, $p < .001$). As shown in Fig. 2, the effect of N1 Type was such that the probability of choosing a plural verb increased from Containers to Quantifiers. The main effect of +/−N2 was also significant ($\chi^2(1) = 52.931$, $p < .001$), with +N2 being more likely to receive a plural verb than –N2. The interaction between +/−N2 and N1 Type was also significant ($\chi^2(3) = 8.624$, $p < .04$). Post hoc pairwise analyses of +N2 versus −N2 within each N1 Type showed that each difference between +N2 and −N2 (after Bonferroni corrections for four comparisons) was significant except for Quantifiers, with the +N2 conditions all significantly more likely to receive a plural verb than the –N2 conditions: Containers: $z = 4.196$, $p < .001$; Collections: $z = 6.110$, $p < .001$; Measure Phrases: $z = 4.187$, $p < .001$; Quantifiers: $z = 2.191$, $p = .114$.

3.3. Discussion

In Experiment 2, we observed evidence of agreement attraction with pseudopartitive subject NPs: The significant main effect of +/−N2 showed that participants were more likely to choose a plural verb when the N2 was present for all N1 Types, except Quantifiers which were near ceiling even in the –N2 condition. This replicates the typical agreement attraction finding when a plural N2 intervenes between N1 and the verb. The main effect of N1 Type provided evidence for an increasing cline in plural agreement from Containers towards Measure Phrases and Quantifiers. As predicted by Marking and Morphing, this result is consistent with the results of Experiment 1, where we found increasing notional plurality across N1 Types. Indeed, there is a significant correlation between the mean notional plurality norms by item and the mean probability of choosing a plural verb ($r = .749$, $p < .001$). Thus, the agreement patterns we observed in Experiment 2 are consistent with previous studies that showed increased rates of plural verb agreement with increased notional plurality (e.g., Vigliocco, Butterworth, et al., 1996; Vigliocco, Hartsuiker, et al. 1996), adding support to our hypothesis that this pseudopartitive domain is a case that falls under the set of phenomena that Notional Plurality is suited to explain.

Marking and Morphing explains the effect of N1 Type and the effect of +/−N2 by using the Marking and Morphing components, respectively. The effect of N1 Type is due to the notional Marking pushing in the direction of plural marking from Containers towards Measure Phrases and Quantifiers. The effect of +/−N2 is due to Morphing: When the N2 is present, there is some chance that its plural feature erroneously percolates up to the root of the subject NP, making the whole phrase plural and increasing the probability of plural verb agreement.

In order to explain all of the results of Experiment 2, Marking and Morphing must rely on two separate mechanisms that are independent of the process of actually building structure. In the next section, we explore SOSP's prediction that these notional effects should stem from contrasts in independently motivated semantic features that guide the choice among structures to be built (N1-headed vs. N2-headed).

4. Experiment 3: Semantic feature hierarchy experiments

In Experiments 1 and 2, we showed that pseudopartitive subject NPs increased monotonically in notional plurality from Containers towards Quantifiers and that plural verb agreement increased monotonically in the same way, consistent with the general finding that higher notional plurality ratings lead to higher rates of plural verb agreement. The SOSP framework expects that both the N1Type and +/−N2 effects found in Experiment 2 can be derived from feature differences in the self-organizing treelets. Keenan (1976) provides a list of over 30 typical subject features. One of these is especially relevant to choosing a subject in pseudopartitives. Keenan argues that subjects typically have autonomous reference; that is, they do not depend on other NPs for their own reference. Containers clearly refer without relying on other NPs: *the box is on the table* refers just as successfully as *the box of chocolates is on the table* even if the context does not specify the contents of the box. Because they have this autonomous reference property, we would expect Container N1s to be relatively good subjects. Collections (*a stack of sandwiches*) refer to an abstract grouping of the items denoted by N2, making their meaning dependent on another NP. Thus, Collections should be poorer subjects because their reference is dependent on that of another NP. Measure Phrases are even worse subjects: The first three words (e.g., *a lot of* in *a lot of newspapers*) do not even specify a grouping, but simply indicate quantity. These NPs are typically treated as operators on the meaning of N2 (Champollion, 2009; Deevy, 1999; Stickney, 2009), similar to determiners like Quantifiers. With this strong dependence on other NPs, Measure Phrases make quite poor subjects. Finally, moving beyond Keenan's properties, this abstract operator status also leads us to expect that Measure Phrases (like Quantifiers) can be acceptable with concrete and abstract N2s (*a lot of postcards* and *a lot of ideas*). This suggests that Measure Phrases are quite similar in meaning to Quantifiers (with Collections and Containers being increasingly different in meaning from Quantifiers). We note that Measure Phrases do differ in some respects from Quantifers, though: having multiple morphemes versus being monomorphemic; the possibility for some modification in Measure Phrases as in *a great/small/ appreciable variety of mugs*. Thus, there seems to be systematic differences in the features on our N1 Types that make them vary in how suitable they are as subjects/quantifiers.

To operationalize these differences in subject-suitability, we selected three features that we hypothesized to vary between our N1 Types: +/−container, +/−spatial configuration, and +/−prohibited abstract N2. The first two (+/−container and +/− spatial configuration) correspond to Keenan's independent reference property, and the final one to the apparent abstract operator status of Measure Phrases and Quantifiers. The values of these features form a cline of similarity in meaning to Quantifiers, from Containers (which denote an independent container, constrain the physical arrangement of the N2, and do not allow an abstract N2) to Collections (which do not denote a container but do imply spatial configuration of the collected elements) to Measure Phrases and Quantifiers (which have neither a container nor an implied spatial configuration) and, therefore, also work with abstract N2s. This contrasts with Collections, which should be less acceptable with an abstract N2.

Experiment 3 was designed to carefully assess these fine-grained semantic features in our materials, with a goal to establishing the feature bundle properties needed to build an SOSP model of the data.

4.1. Method

4.1.1. Participants

Seventy-four undergraduate participants took part in the experiment for course credit using an online participant pool platform. We had to remove the data from 13 non-native speakers of English, one participant who reported a speech or language problem, and two participants for choosing almost the same rating for each sentence. Participants received credit regardless of whether their data were used or not. The removal of these data points did not change the results. Fifty-eight participants were entered into the analyses reported here.

4.1.2. Design and materials

We used four different semantic tests to try to identify a set of semantic features that would capture the gradation between the typical NP-like Containers and the abstract, Quantifier-like Measure Phrases. The first test focused on the acceptability of pairing a Container or Collection subject NP with the verb *overflowed* (Container-hood Test 1; see (5)). Only subject NPs with a Container N1 should be acceptable with this verb, since there is no physical container in Collections that could overflow.

(5) Container: We added so many strawberries that the dish of strawberries overflowed.
Collection: We added so many shirts that the pile of shirts overflowed.

The second test (Container-hood Test 2; (6)) compared the acceptability of Containers and Collections with the verb *broke*. Since Collections lack a physical container that could break, participants were expected to rate these sentences lower than the Containers.[8]

(6) Container: A dish of strawberries broke.
Collection: A pile of shirts broke.

The third test, a Spatial Configuration test, (7) used shape adjectives like *tall* to modify Collection and Measure Phrase N1s. Collections were expected to receive higher ratings since they seem to imply a spatial configuration that could have a particular shape, whereas Measure Phrases should be rated lower because they do not constrain the spatial configuration of the N2.

(7) Collection: She moved a tall pile of shirts into the garage.
Measure Phrase: She moved a tall bunch of shirts into the garage.

The fourth test (Abstract N2; (8)) tested whether it was possible to pair Collections and Measure Phrases with an abstract N2. Measure Phrases were expected to be rated higher since their meanings seem to be abstract, conveying only number information, whereas Collections seem to require a more concrete N2.

(8) Collection: She defined a pile of concepts.
Measure Phrase: She defined a bunch of concepts.

In running these tests comparing just two N1 Types, we made the assumption that predictable patterns would hold in the remainder of N1 Types; that is, we assumed an implicational hierarchy in the N1 Types: For example, if Containers have a certain semantic feature and Collections do not, we assumed that Measure Phrases and Quantifiers also lack that feature. Furthermore, we assumed that Measure Phrases and Quantifiers will have the same features, so we did not include Quantifiers in the tests. Thus, we only made pairwise comparisons between adjacent levels of N1 Type. This approach receives support from Mahowald, Graff, Hartman, and Gibson (2016), who argue that acceptability judgments that the experimenters expect to be unanimous in a forced-choice paradigm can be put on surer statistical footing by polling a small number of participants without having to run a large-sample experiment. The three authors of this paper agreed that Containers were more acceptable than Measures Phrases and Quantifiers for the two Container-hood tests, that Collections were more acceptable than Measure Phrases and Quantifiers for the Spatial Configuration test, and that Measures were more acceptable than Collections for the Abstract N2 tests. Mahowald et al. (2016) calculate the expected probability of unanimity in such a test with three participants and three items to be 0.89

(95% CI [0.7, 1.0]). Thus, we feel justified in having done only the reported pairwise comparisons to investigate the hypothesized feature hierarchy.

We included 32 critical items, 56 items from an experiment to be reported elsewhere,[9] and 35 fillers, a total of 123 items in the whole experiment. There were eight critical items for each semantic test. All participants rated each item.

4.1.3. Procedure

The sentences were presented using the IbexFarm online software (http://spellout.net/ibexfarm/). After reading an information sheet and agreeing to participate, participants were instructed to enter acceptability judgments on a 1- to 7-point Likert scale, where a rating of 1 indicated that a native speaker of English would not naturally produce that sentence, and 7 indicated that the sentence was perfectly natural for a native speaker of English. Participants were given example sentences with ratings to illustrate the use of the scale (see Appendix S1).

Following the instructions, participants completed a short demographics form and proceeded to practice sentences (given in Appendix S1). Practice and test items were presented individually in the center of the screen with clickable buttons labeled 1–7 directly below them. When participants clicked on a rating, the next item appeared. After rating all items, participants were granted course credit.

4.2. Results

All analyses were done using linear mixed effects models with by-participant random intercepts and random slopes for N1 Type and by-item random intercepts. Factors were coded using numerical sum coding (-1 or 1). The p-values were obtained via likelihood ratio tests.

4.2.1. Container-hood 1: Overflowing test

A linear mixed model showed a significant main effect of N1 Type ($\chi^2(1) = 5.24$, $p < .05$). Containers were rated significantly higher than Collections ($M_{\text{Container}} = 4.99$, $SD_{\text{Container}} = 1.61$; $M_{\text{Collections}} = 4.10$, $SD_{\text{Collections}} = 1.88$), indicating that participants found overflowing Containers significantly more acceptable than overflowing Collections.

4.2.2. Container-hood 2: Breakable test

A linear mixed model showed a significant effect of N1 Type ($\chi^2(1) = 15.60$, $p < .001$), with Containers rated higher ($M = 4.388$, $SD = 2.086$) than Collections ($M = 2.453$, $SD = 1.648$). This suggests that participants found sentences more acceptable when a Container broke than when a Collection did.

4.2.3. Spatial configuration test

A linear mixed model again showed a significant effect of N1 Type ($\chi^2(1) = 6.79$, $p < .01$). Here, Collections ($M = 4.668$, $SD = 1.911$) were rated higher than Measure Phrases ($M = 3.474$, $SD = 1.932$); thus, shape modifiers were rated more acceptable with a Collection N1 than with a Measure Phrase.

4.2.4. Abstract N2 test

A linear mixed model showed a significant effect of N1 Type ($\chi^2(1) = 16.84$, $p < .001$). Measure Phrases were rated higher ($M = 5.384$, $SD = 1.650$) than Collections ($M = 3.384$, $SD = 1.803$), indicating that Measure Phrases are more acceptable than Collections with an abstract N2.

4.3. Discussion

In line with our intuitions, our semantic tests suggested a hierarchy of semantic features. While the results of the Likert scale judgments were on a 7-point ordinal scale, we focus here on the relative acceptability between conditions, effectively boiling the acceptability results down to a binary comparison. The results of the Container-hood Tests 1 (overflowing) and 2 (breakability) showed that whereas Containers imply a physical container, Collections do not. The results of the Spatial Configuration test showed that Collections put constraints on the physical layout of the objects denoted by N2, but Measure Phrases and Quantifiers do not. Finally, Measure Phrases place no concreteness restriction on the denotation of N2, whereas Collections do. Thus, the meanings of the phrases range from concrete, standard NPs that make good subjects according to Keenan (1976)'s features to more abstract, quantificational phrases. Together with the assumptions discussed above, these results suggest the semantic feature hierarchy in Table 1, ranging from least quantifier-like in meaning (conveying much non-number information) to most quantifier-like (conveying only number information), keeping in mind the assumptions that the relative acceptability is most relevant and the authors' judgments on the items not tested in the experiments:

Under SOSP, the differences in semantic features between the N1 Types should bias the formation of different structures by making links with good feature matches stronger competitors. Specifically, if the N1 has features that make it a good subject and a poor Quantifier (as do Containers), the parser will be more likely to build the N1-headed parse because the head of the N1 treelet is a good match to the verb's subject attachment site. But if the N1 makes a poor subject and a good quantifier (as do Measure Phrases), the N2-headed parse is more likely to form (note that we used N2s that have a good feature match for the verb's subject attachment site in all items). For Collections, which are

Table 1
Feature values for the different N1 Types based on the results of Experiment 3

	Semantic Feature		
N1 Type	Container	Spatial Configuration	Abstract N2 Prohibited
Containers	+	+	+
Collections	–	+	+
Measure Phrases	–	–	–
Quantifers	–	–	–

Notes. The table uses the feature "prohibited abstract N2" rather than "abstract N2" allowed in order to align the polarity of the features across the test: Subject NPs with more + values are more compatible with N1-headed structures; those with more − values are more compatible with N2-headed structures.

intermediate in their feature match to the verb, we expect a more balanced distribution of N1- and N2-headed parses. Since these distributions align with the N1 Type effects observed in Experiment 2, the SOSP approach of deriving notional plurality effects from independently motivated semantic features in a single parsing mechanism receives some support. However, we have so far only sketched the principles of this account. In the next section, we make it plausible that the principles can work together in an explicit mechanism by showing how an SOSP processor generates the pattern of Experiment 2, including the +/−N2 effect, based on the feature bundles derived from Experiment 3.

We note that taking the path of feature-guided parsing raises the question of why we should have four N1 Types. Could we not have used any other number of categories? Indeed, dividing the N1 Types into four categories is somewhat arbitrary, and other divisions and features will likely be relevant when extending this approach to other constructions. In principle, we could posit a different N1 Type for each combination of features present in individual items, distinguishing, for example, between *a bottle of pills* and *a basket of oranges*. However, grouping the N1 Types into four categories captures our intuitions about the general meanings of each type: There are enough shared features within each category for them to make the grouping natural. Moreover, these intuitions are backed up by the significant differences between categories observed in Experiment 3. While using four N1 Types is still an approximation to even more fine-grained semantic representations, it does start to unveil some of the semantic underpinnings of notional plurality.

Finally, rather than nullifying the results of Experiment 1, the semantic features uncovered in Experiment 3 offer a new way of understanding notional plurality. The notional plurality norms for our materials can be thought of as a way of roughly placing subject NPs along the semantic feature cline related to subjecthood versus quantifierhood. Because our N1s were always singular, N1-headed structures should be more likely to be rated notionally singular. N2-headed structures, on the other hand, were more likely to be rated notionally plural because our N2s were always plural. Because sentence processing effects in SOSP are due to differences in feature specifications affecting parse formation, we were led to look for semantic features that could drive the formation of different parses acros the N1 Types. Doing so with these materials allowed us to unpack the pretheoretical concept of notional plurality into independently motivated semantic features. In the Simulation section below, we show how the assumptions that SOSP makes can work together in a model of online linguistic structure building that generates the patterns observed in our experiments.

5. Simulation

In the Introduction, we sketched a theory of sentence parsing based on competition between attachment links guided by syntactic and semantic features. In this section, we describe an implemented version of relevant parts of this framework, showing how it produces distributions of parses of pseudopartitive subject constructions consistent with the results of Experiments 1 and 2 when we incorporate the semantic features reported in

Experiment 3. The full framework involves lexically anchored treelets that are themselves dynamical (with continuously changing features) that competitively form attachments with one another. In the present model, we implemented just the link competitions because this part suffices to model the data from the experiments above.

5.1. Architecture

In addition to the link attachment competitions, each attachment site in the full SOSP framework has a vector of syntactic and semantic features, the values of which change continuously under the influence of the values of the features on treelets linked to that attachment site. In addition, the only restrictions placed on which links can form are that there can be no within-treelet links and that links must join a head attachment site to a dependent attachment site (no head-to-head or dependent-to-dependent attachments), similar to some previous dynamical parsing models (e.g., Tabor & Hutchins, 2004). Links can, in principle, grow between head and dependent sites that are extremely poorly matched (e.g., a verb head attaching to a nominal dependent site on another word). However, the poor feature match for such links will generally prevent them from winning the competition. In most cases, links with a good feature match thrive, and those with a bad feature match lose out, leading to the construction of well-formed trees.

In the present simulation, we only simulated the subset of all possible links that is necessary to build the N1- and N2-headed parses, with links competing with other links attaching to the same attachment site. To approximate the effects of letting the feature vectors themselves change dynamically, we also allowed links to compete that were part of incompatible parses (see Appendix S2 for details). The inclusion of only N1- and N2-parse links plausibly has the same dynamics as the full framework because the ignored links have very poor feature match and thus do not play a substantial role in the competition. Considering only the restricted link set allows us to analyze the implemented model more easily. Using tools from dynamical systems theory (Guckenheimer & Holmes 1983; Strogatz, 1994), we determined (see Appendix S2) that the simplified model has only two attractors, that is, points in the system to which the system returns after small perturbations. The two attractors correspond to either all of the links supporting the N1-headed parse winning or all of the links supporting the N2-headed parse winning. Since there are no other attractors in the system, the parser will form one of the relevant parses and nothing else.

Informally, processing in the implemented model works as follows. When a word is perceived in the input, the links connecting that treelet with other treelets are boosted in strength, effectively "activating" them and initiating competitions. As additional words are perceived, their links get a boost and continue competing with other incompatible links. How quickly a link can grow is determined by how compatible the features are at the attachment sites it connects. Links with good feature matches are, therefore, able to compete more strongly with their competitors. The competitions continue until a clear winning structure emerges. Finally, small-magnitude Gaussian noise is added to the links at each time step. This makes it so that the system produces a distribution over N1- and N2-headed parses biased by the feature matches on the links.

We turn now to a more detailed description of the model. As motiviated in the Introduction, in both the N1- and N2-headed parses, *of* is the head of a functional phrase. In the N1-headed structure, *of* attaches to the PP-dependent node on the N1 and takes N2 as a nominal dependent. In the N2-headed structure, *of* takes the N1 as a nominal dependent and functions as a determiner dependent of N2. In the full SOSP framework, in which treelet feature values are dynamic, the head node of *of* adjusts its feature values continuously between values encoding determiner properties and values encoding preposition properties, depending on whether the parse is being pushed more toward the N2- or the N1-headed parse, and the *of*-treelet has attractors for these two states. In the present simulations, we approximated the full framework by giving *of*'s head node features of both prepositions and determiners. Additionally, we included competition between the N1 to verb link and the *of* to N2 link and between the N2 to verb link and the *of* to N1 link. These extra competitions, which would not be present in the full framework, simplify the dynamical-treelet-with-attractors assumption, while still ensuring that the system has only the attractors corresponding to *of* as determiner and *of* as preposition, and no stable blends of these states (cf. Cho et al., 2017).

The link competition dynamics are governed by a set of differential equations originally developed for modeling predator–prey interactions and between-species competition in ecology, the Lotka-Volterra equations (Frank, 2014; Fukai & Tanaka, 1997; Hirsch, Smale, & Devaney, 2004. Lotka, 1920). These equations are based on models of logistic population growth. In the absence of competition, a single species with a positive initial population will approach its maximum carrying capacity, here normalized to be 1. The equation for each species i of n species has the following form:

$$\frac{dx_i}{dt} = x_i\left(1 - \sum_{j=1}^{n} W_{ij}x_j\right),$$

where x_i is the population density of species i relative to its maximum carrying capacity, and W_{ij} is an interaction matrix determining which species compete and how strongly.

We incorporated the semantic-feature-driven link competition by having the feature match scale the rate at which a link can grow. Thus, a link i will be able to grow faster if the feature match m_i between the attachment sites it connects is high and grow more slowly if m_i is low. This implements the idea that it should be easier to build structures when the constituents fit together well. We simulated $n = 6$ links (Fig. 3 below) based on the considerations discussed above about which links are most relevant to the current parsing problem. The form we used for the simulations is given below, where x_i now stands for the link strength, and η is a Gaussian noise process:

$$\frac{dx_i}{dt} = m_i x_i\left(1 - \sum_{j=1}^{n=6} W_{ij}x_j\right) + \eta$$

We used the following procedure to calculate the feature match values m_i. First, we included a +noun feature on N1 and N2. Based on the results of Experiment 3, we also assumed that there were three semantic features on the N1 and N2 relevant to the formation of structure in our materials: +/− container, +/−spatial configuration, and +/−prohibited abstract N2. These were encoded using binary features, with the presence of a feature coded as 1 and absence coded as 0 (see Table B1 in Appendix S2 for feature values for noun-related attachment sites). We assumed a further feature in order to capture the +/−N2 effect: This feature had a value of 1 if the word was present in the input and 0 if it was elided. In the −N2 conditions, therefore, this feature on the *of* and N2 treelets was set to zero. Under the assumption that the features identified in Experiment 3 contribute to making a noun a good subject and that a word being present in the input is easier to integrate into a structure than one that must be inferred from the context, we assumed that the verb's subject dependent attachment site was a relatively good match for both Containers and Collections and a fairly poor match for the N1 of a measure phrase. The +present feature also made it so that the N2 in the −N2 conditions was a poor subject compared to the +N2 conditions because the elided N2 would not match the last feature of the verb's subject attachment site. Since N2 was consistently a plausible subject noun throughout the experiment, we set N2's features to be as good a feature match with the verb as the best N1 Type (Containers). For *of*, we assumed its head features were +preposition, +determiner, and +/−present ([1, 1, 1/0]). Having both the +preposition and +determiner features encoded the assumption that *of* can function equally well in this context in either role. The NP-dependent attachment site on *of* had the features +noun and +present (we assumed that *of* can take any kind of noun as a dependent equally well, so we did not include the other features when determining the feature match). Finally, the PP attachment site on N1 had the features +preposition, −determiner, and +present ([1, 0, 1]), while the determiner attachment site on N2 had the features −preposition, +determiner, and +present ([0, 1, 1]).

Since the feature vectors are binary, we used the Hamming distance (the number of features values that differed between the two sets of shared feature dimensions) between the feature vectors connected by each link i as a distance measure, although any distance metric that preserves the ordering of the N1 Types should produce similar results. We converted the Hamming distances to similarities by taking the exponential of the negative of the distances (Shepard, 1987):

$$m_i = e^{-distance}$$

Since the distances were always non-negative, the m_i therefore range between 0 and 1.

Quantifiers have a different structure than the other pseudopartitive N1 Types, so they were handled differently. The only relevant grammatical structure can form between a quantifier and a noun is for the quantifier to attach as the determiner dependent of the noun, so we only included the link attaching the quantifier as the determiner dependent of N2 and the link attaching the N2 as the subject of the verb. We used the treelet corresponding to *of* in the other conditions as the quantifier. Since we only used plural quantifiers paired with plural N2s in our materials, we set the feature match between the head of the quantifier to be a perfect feature match for the N2's determiner attachment site (i.e., the relevant $m_i = 1$). The N2's head attachment site was the same as in the other simulations, so its feature match to the verb was unchanged (see Table B1 in Appendix S2). The other link strengths (and their feature matches) were clamped to 0. This setup for the Quantifier condition can only produce parses headed by N2—there is no link from the quantifier to the verb. Thus, all Quantifier simulations resulted in N2-headed parses.

The interaction matrix W_{ij} constrains which links compete with each other, and thus which sets of winning links are possible. In the full theory of competitive parse formation, the W_{ij} should encode competition between all links that can exist between all treelets. However, as noted above, we only included links that participate in one of the two viable structures. In addition, we only simulated the parsing of the N1, *of*, and the N2, leaving out N1's determiner, as its inclusion does not differentially affect the conditions. The interaction matrix W_{ij} is given in Table B1 (Appendix S2). Fig. 3 depicts the simulated links.

Attachment links to the verb are included, but we did not simulate inputting an actual verb in order to model the Staub (2009) paradigm used in Experiment 2. The goal was to model the formation of a structure that supports choosing either singular or plural verb agreement. In our materials, N1 was always singular, and N2 was always plural, so if the system settles on the N1-headed structure, this amounts to choosing a singular verb. Similarly, settling on an N2-headed parse corresponds to choosing a plural verb. For the

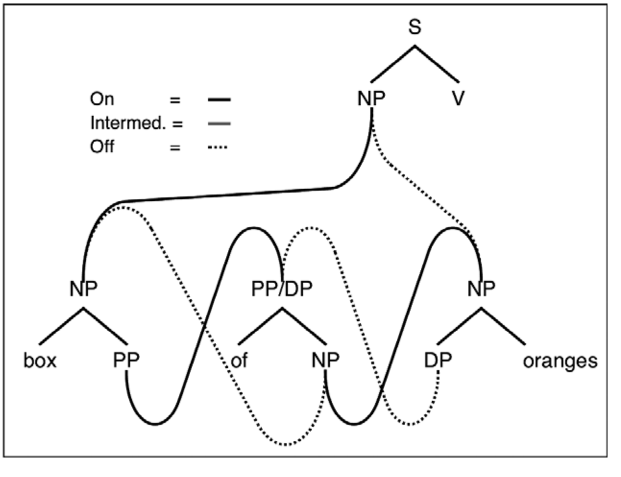


Fig. 3. Architecture of the present simulation. The dark (nearly maximally activated) curved links participate in the N1-headed structure, while the light (nearly minimally activated) links participate in the N2-headed structure. Thus, the figure shows the system at a point when it has nearly converged on the N1-headed structure. As noted, this model does not have dynamic feature values, so the link dynamics constitute the entire system. (The feature values on the nodes, which determine the match values and thus influence link growth rates, are not shown.)

model to reproduce the results of Experiment 2, it should produce proportions of N1- and N2-headed parses comparable to the proportions of singular and plural agreement we observed. Given the linguistic evidence for both parses in pseudopartitives, it is reasonable to assume that participants are making the same choice in Experiment 2.

Finally, we modeled the perception of a new word by boosting the activation of the links attaching its treelet to others. When a new word is perceived, the strengths of its links are boosted by adding 0.1 to their current strength. Words not present in the input (*of* and N2 in the −N2 conditions) received no boost. To simulate the equations, we discretized them and input a new word every 100 time steps. Pilot simulations showed that the qualitative pattern of results was insensitive to this parameter, and this value provided a reasonable fit to the human data.

To test the rates at which the system stabilizes on each parse with noisy link strengths, we ran Monte-Carlo simulations in each of the N1 Types in both +/−N2 conditions. For each condition, we ran 1,000 simulations using the Python NumPy library (Van der Walt, Colbert, & Varoquaux, 2011). We used simple Euler-Maruyama integration with a time step of 0.01 to numerically integrate the equations (Higham, 2001). The code and data are available at https://github.com/garrett-m-smith/; see Appendix S2 for further details.

5.2. Results

To assess which parse the model settled on, we integrated the system until either the link from N1 to the verb or the link from N2 to the verb had a strength greater than 0.5. This approximates the system beginning to settle on one parse or the other. As shown in Fig. 4 below, the model produced distributions over N1- and N2-headed parses that are qualitatively similar to the rates of singular and plural agreement from Experiment 2. Specifically, the model always produces more N2-headed parses in the +N2 conditions than in the −N2 conditions, with the exception of the Quantifiers, which only produced N2-headed parses. In addition, the model also replicates the effect of N1 Type: The probability of building an N2-headed parse increases monotonically from Containers to Collections to Measure Phrases to Quantifiers.

5.3. Discussion

The simulations presented in this section show that a self-organizing parser that incorporates fine-grained semantic features can produce distributions over N1- and N2-headed parses that qualitatively match the effects observed in the verb agreement data from Experiment 2. This model illustrates that the effects observed in the human data can be plausibly explained by a feature-driven, self-organizing parsing process. Unlike Marking and Morphing, which relies on notional plurality in an extra processing step, the simple SOSP model presented here relies only on the local feature information on linked treelets, deriving notional agreement attraction effects in pseudopartitives from the normal structure building process.

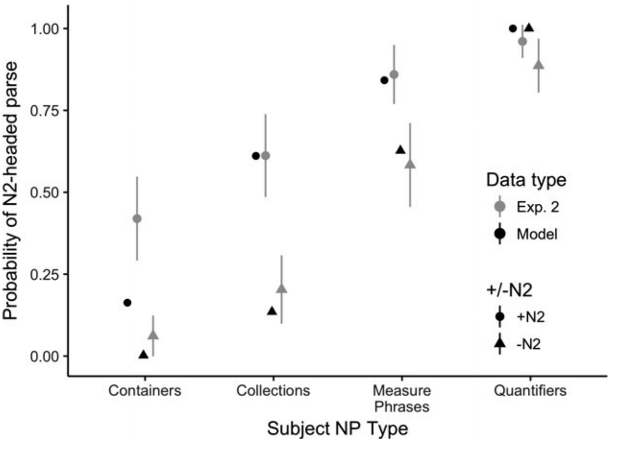


Fig. 4. Comparison of proportion of N2-headed parses between the model (black points) and the human verb production data from Experiment 2 (gray points). The human mean proportions are shown with 95% CIs as in Fig. 2.

The model was able to reproduce the qualitative shape of the human data via integration of the feature match parameters, the m_i, into the dynamics of parse formation. If a link connects attachment sites that have a good feature match, that link's strength grows quickly, making it a stronger competitor against other links. For example, a Container N1 is a good feature match for the subject attachment site of the verb and a relatively poor dependent for *of* in its determiner reading. The good feature match between the Container N1 and the verb will cause that link to grow quickly while strongly inhibiting the growth of the link between N1 and *of* as *of*'s dependent. Even though N2 is also a good feature match for the verb, by the time it is perceived, the link from N1 to the verb is strong enough to prevent the link from N2 to the verb from growing. The situation is reversed in the case of a Measure Phrase. Here, the N1 is a relatively poor feature match for the verb, so its link to the verb can only grow slowly. When the N2 is perceived, its good feature match to the verb allows it to quickly overcome the weak competition from N1 and form the N2-headed structure. In the −N2 conditions, the elided *of* and N2 are a weaker feature match, slowing their growth. In addition, they receive no boost in strength since they are not present in the input, which makes it even harder for the links needed for the N2-headed parse to grow.

We note that the model produces a less than optimal fit to the human data for Containers and Quantifiers especially. For the Quantifiers, the model can only produce plural parses, so there is no way that it could choose a plural verb, even "by accident" as a human participant occasionally might despite having built a plural-preferring structure. For the Containers, the model produced fewer plural parses than the humans in both +/− N2 conditions. We believe that this is because the strong feature match between the N1 and the verb allows that link to reach a high enough activation level to be nearly impervious to any competition from the N2, which is perceived after the N1.

Despite these limitations, this model illustrates how a self-organizing parser selects between the two possible parses in pseudopartitive constructions, showing how the pattern of agreement in our Experiment 2 data might have occurred. Although we only considered a simplified version of the full SOSP framework here, we have included all the links that plausibly exert a significant influence on the verb selection, and we have approximated feature dynamics within treelets where it is relevant (the treatment of the *of* treelet). Thus, the current simulation of pseudopartitive structures provides some evidence that the SOSP approach will scale up to include a greater variety of constructions, a point we take up in General Discussion.

6. General discussion

We explored subject–verb number agreement with pseudopartitive constructions, delving into the two factors that the Marking and Morphing theory identifies as relevant to number determination: the Marking component, which the theory takes to be the source of notional plurality effects on number agreement, and the Morphing component, which the theory takes to be the source of morphosyntactic feature effects on number agreement, including attraction. Under Marking and Morphing, these are independent factors that are combined together additively to produce agreement choices.

In pseudopartitive constructions of the form *a N1*[sg] *of N2*[pl], we first varied N1 Types across the subclasses Containers, Collections, and Measure Phrases and Quantifiers (as a control). The results of Experiment 1 showed that this manipulation progressively increased notional plurality, which affects the Marking stage in Marking and Morphing. Next, we compared the presence versus absence of N2, varying the influence of the N2's morphosyntactic features during the Morphing stage in Marking and Morphing. In keeping with the predictions of Marking and Morphing, the increasing notional plurality across N1 Types increased the rate of plural verbs selected, while removing the plural N2 from the sentence decreased it in Experiment 2.

Noting that the notional plurality part of Marking and Morphing lacks a systematic theoretical basis and that it is desirable, if possible, to consolidate a two-mechanism theory into a one-mechanism theory, we turned to SOSP. In SOSP, linguistic tree-representations form via continuous feedback interactions among treelets that are guided by vectors of syntactic and semantic features. Experiment 3 provided evidence for a set of semantic features that systematically distinguished our three N1 Types (+/− container, +/− spatial configuration, +/− abstract N2). Incorporating these featural specifications into a dynamical model, we showed how the observed effects of N1 Type and N2-presence occur in a single-mechanism under the SOSP account. As the featural specifications of N1 vary across Containers, Collections, and Measure Phrases, the N1 becomes less like a typical subject, and the first three words of the subject NP (*a N1 of*) become progressively less like a noun-preposition modification structure and more like a quantificational determiner, increasing the probability that an N2-headed noun phrase rather than an N1-headed noun phrase will form during the self-organizing parse building. Since N1 was always singular and N2 always plural in our stimuli, this change in N1 shifted the rate of plural usage. Finally, when we removed *of N2* from the input, the bottom-up support for the N2 as head was decreased, so the model showed fewer N2-headed parses in the −N2 condition as well.

We see our account not as contradicting Marking and Morphing, but as delving into the causes of several phenomena highlighted by the theory. SOSP derives the effects of both Marking and Morphing from a core, independently needed process, syntactic structure building. It improves on Marking and Morphing by replacing two sequentially ordered mechanisms with a single mechanism. In addition, SOSP incorporates both production and comprehension, while Marking and Morphing only covers production. Still, our empirical findings also vindicate Marking and Morphing's focus on semantic and syntactic factors in agreement and extend it to structures somewhat different from the typical cases studied in the literature.

6.1. SOSP and other cases of agreement attraction

Given that the pseudopartitives are arguably syntactically and semantically different from the types of structures usually considered in studies of agreement processing, it is important to ask what SOSP has to say about the typical cases. Here, we consider two central ones: canonical prepositional modification structures (e.g., *the key to the cabinets...*) and, within those, distributivity manipulations (e.g., *the signature on the checks...* vs. *the cage with the gorillas...*). While we do not model these constructions here, we argue that SOSP can plausibly account for these classic findings.

The canonical prepositional modification result is that structures of the form [Det N1 [sg] Prep Det N2[pl]] exhibit more plural completions in production than control cases with both nouns singular (e.g., Barker, Nicol, & Garrett, 2001; Bock & Miller, 1991; Brehm & Bock, 2013). The full SOSP framework predicts this effect due to a temporary interaction between the N2 treelet and the treelet of the upcoming verb (see Fig. 1). First, after production of the determiner and the N1, the N1 treelet begins to attach to the subject attachment site of the verb (for which a specific verb has not yet been specified). Then, after the production of the preposition and the second determiner, the N2 is produced. Because the N2 also fits the specifications for being a subject of the verb, it competes with N1 to attach as the subject of the verb. Typically, it will not win this competition because N1 has a lead in forming this attachment, but during its temporary interaction with the verb, N2 has a chance of pushing the verb treelet to its plural state, rather than its singular state. Because the two number states of the verb treelet are attractors, the verb can get stuck in its plural state, even though the tree ultimately stabilizes with the N1 as the verb's subject and the N2 as the object of the preposition. Thus, when the participant produces the verb, he or she produces a plural verb.

This behavior of the SOSP system is consistent with results providing evidence that agreement attraction can occur in the absence of thematic role assignment aligned with the agreement choice because the plural feature on the N2 can influence the verb temporarily even if the N2 does not end up linked as the subject where it would presumably receive its thematic role feature (e.g., Lau, Wagers, Stroud, & Phillips, 2008; Schlueter, Parker, & Lau, 2017). Even if the correct N1-headed structure forms in the end, allowing for correct thematic role assignment, the plural feature on N2 can sometimes push the verb's number feature into its plural state. SOSP, in its general form, also assumes feature flexibility in the N1 treelet after the N1 has been perceived,[10] so the treatment just described is also consistent with the result of Patson and Husband (2015) who found that cases of [Det N1[sg] Prep Det N2[pl]] were interpreted as having a plural N1 (in comprehension questions) more often than in the singular-singular case. It is noteworthy that SOSP treats classical agreement attraction cases by the same mechanism (structure formation) as it treats the pseudopartitive cases. However, it predicts a qualitative difference between them. In the classical case, the agreement interference due to N2 is temporary and only influences superficial aspects of the parse, while in pseudopartitives, the agreement "interference" due to N2 often determines the form of the final parse. The difference between the two cases arises because, in classical cases, there is only one major syntactic structure at play, while in pseudopartitives, there are two.

As we noted in the Introduction, one of the findings that motivates the Marking component of Marking and Morphing is the increased rate of plural verb production in notionally plural distributive noun phrases (e.g., *the signature on the checks*) compared to notionally singular non-distributive noun phrases (e.g., *the cage with the gorillas*). Classical production models, including Marking and Morphing, posit that meaning and form are realized at two sequentially ordered stages. By contrast, SOSP assumes that meaning is always present when form is present. There is no overarching representation which could contain the "message" of a whole utterance, so the interaction of the language system with perception and action has to take place at the level of the treelets. We assume that each treelet activates and is activated by situations that it tends to be used in. When a language user is asked to reproduce the phrase *the signature on the checks*, each activated treelet contributes to the activation of a scenario of a single signature duplicated across multiple checks. But this scenario is also compatible with the phrase *the signatures on the checks*. Because of this, the activation of the scenario will tend to cause activation of the treelet for *signatures* as well as *signature*, and these will compete to form structure. In this case, the plural feature on *signatures* provides an additional force pushing the number feature on the verb treelet towards the plural attractor, increasing the chances the verb will end up with a plural marking. If the prompt is *the cage with the gorillas*, on the other hand, the treelets will activate a scenario with a single cage, which is not compatible with *the cages with the gorillas*. Thus, there will only be a single source of plurality influencing the verb, that of *gorillas*. In effect, SOSP predicts distributivity effects because distributive singular-plural NPs are nearly synonymous with the corresponding plural-plural noun phrases. This explanation goes beyond the self-organizing treelet architecture itself, but the additional mechanism is independently needed to explain how users relate language utterances to scenarios. Moreover, these assumptions naturally generate the finding that after having processed a sentence, a person can often reconstruct the semantic gist of the sentence, but he or she may diverge with respect to syntactic details (Bock & Brewer, 1974; Mehler, 1963; Miller, 1962).

6.2. Relation to other processing models

Self-organized sentence processing is closely related to ACT-R-based sentence processing (Lewis & Vasishth, 2005), which also posits that sentence processing involves the interaction of independently acting memory traces or “chunks” that jointly specify syntactic tree structures. In the domain of agreement attraction, ACT-R approaches have mostly been applied to formal feature interference cases in comprehension (Dillon, Mishler, Sloggett, & Phillips, 2013; Jäger, Engelmann, & Vasishth, 2017; Wagers, Lau, & Phillips, 2009; note also that these are cases that Marking and Morphing would have to treat via Morphing), although Badecker and Kuminiak (2007) describe an application of ACT-R to production. The current work suggests that, to handle notional plurality (Marking) effects, it would be useful for ACT-R to address the question of how multiple semantic factors are combined to select a parse when multiple parses are available.

In relation to statistical parsing approaches (e.g., Hale, 2001; Jurafsky, 1996; Levy, 2008), SOSP broadly recapitulates the effects of statistical parse biases in ambiguous structures via stochastic treelet interactions. However, it highlights the possible relevance of temporary structure formation phenomena (like the temporary interaction of the N2 treelet with the verb treelet) which are not usually considered in those theories.

An important area for further research with the SOSP framework is how to handle timing phenomena, which form an important source of data in agreement attraction in sentence comprehension (e.g., Brehm & Bock, 2013; Pearlmutter, Garnsey, & Bock, 1999; Wagers et al., 2009; for a Bayesian meta-analysis that includes these and other data, see Jäger et al., 2017). Both ACT-R and statistical parsing approaches (e.g., Smith & Levy, 2013) make quantitative predictions about reading times in comprehension, so if SOSP were not able to make such predictions, its value as a psycholinguistic theory would be severely limited. This is an important desideratum for future work in SOSP, and determining whether it can accurately model timing will allow more direct, quantitative comparisons with the more established sentence processing models.

6.3. Conclusion

In sum, we have provided evidence that notional plurality can be unpacked into a set of semantic features. Embedding these features in a self-organizing processing model allowed us to cast agreement attraction as a natural consequence of the structure formation process in which syntax and semantics work together closely, in contrast to Marking and Morphing, which requires two mechanisms beyond parsing to account for the observed effects. The model presented here provided an example of how the SOSP approach can be explored in a focused way: Using simple models of particular constructions, we can test the hypothesis that self-organization provides a broad and unifying basis for handling sentence-level psycholinguistic phenomena.

Notes

1. The assumption of a continuous range of notional plurality values receives support from rating studies which indicate gradations of notional plurality, for example, Bock et al. (2001).
2. The status of *of* is not clear in the linguistic literature. In the N1-headed structure, *of* is generally considered to be a standard preposition. But in the N2-headed structure, opinions are divided. Deevy (1999) and Stickney (2009) treat it as the head of a functional projection that can assign case. Selkirk (1977), on the other hand, has it inserted in the phonology. We assume that *of* in the N2-headed structure is the head of a determiner phrase taking the N1 as a complement, similar to the English possessive morpheme—'*s*, as proposed by Abney (1987).
3. SOSP is also guided by a well-motivated theory of grammar, but in a bottom-up fashion.
4. Bock et al. (2012) coded "one thing" responses as a 1 and "more than one thing" responses as a 2 and analyzed the difference, using ANOVA on the mean responses.
5. The model failed to converge with the default optimizer used in the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015), but using the "bobyqa" optimizer for both stages of the optimization converged without warnings, so we report that model here.
6. Reaction time analyses will be reported in future work.
7. Due to convergence issues, the correlation parameters of the random slopes were excluded from the analysis.
8. The N2s in Containers and Collections were approximately equally breakable, so any effect in this test is most likely to be driven by differences in breakability of the N1s.
9. These fillers tested other semantic and syntactic properties of pseudopartitives in support of future comparisons to similar constructions in French.
10. Once the phonological or orthographic form is no longer present, if a sufficient force influences the treelet, it may be bumped from the stable state that the perceived word left it in into a nearby stable state, for example, changing its number value from singular to plural.

References

Abney, S. P. (1987). The English noun phrase in its sentential aspect. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, *56*, 65–85.

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bock, K., & Brewer, W. F. (1974). Reconstructive recall in sentences with alternative surface structures. *Journal of Experimental Psychology*, *103*(5), 837–843.

Bock, K., Carreiras, M., & Meseguer, E. (2012). Number meaning and number grammar in English and Spanish. *Journal of Memory and Language*, *66*(1), 17–37.

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, *43*(2), 83–128.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*(1), 45–93.

Brehm, L., & Bock, K. (2013). What counts in grammatical number agreement? *Cognition*, *128*(2), 149–169.

Brems, L. (2003). Measure noun constructions: An instance of semantically-driven grammaticalization. *International Journal of Corpus Linguistics*, *8*(2), 283–312.

Bresnan, J. (Ed.) (1982). *The mental repesentation of grammatical relations*. Cambridge, MA: MIT Press.

Büring, D. (2005). *Binding theory*. Cambridge, MA: Cambridge University Press.

Champollion, L. (2009). A unified account of distributivity, for-adverbials, and measure constructions. In M. Prinzhorn, V. Schmitt, & S. Zobel (eds.), *Proceedings of the 14th Sinn und Bedeutung conference* (pp. 84–100). Vienna.

Cho, P. W., Goldrick, M., & Smolensky, P. (2017). Incremental parsing in a continuous dynamical system: Sentence processing in Gradient Symbolic Computation. *Linguistics Vanguard*, *3*(1), 1–10.

Deevy, P. L. (1999). The comprehension of English subject-verb agreement. Unpublished doctoral dissertation, University of Massachusetts, Amherst, MA.

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

Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject–verb agreement in English. *Journal of Memory and Language*, *41*(4), 560–578.

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, *112*(3), 531.

Fodor, J. D. (1998). Unambiguous triggers. *Linguistic Inquiry*, *29*(1), 1–36.

Fodor, J. D. (2017). Ambiguity, parsing, and the evaluation measure. *Language Acquisition*, *24*(2), 85–99.

Foote, R., & Bock, K. (2012). The role of morphology in subject–verb number agreement: A comparison of Mexican and Dominican Spanish. *Language and Cognitive Processes*, *27*(3), 429–461.

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, *17*(4), 371–404.

Frank, T. D. (2014). Secondary bifurcations in a Lotka-Volterra model for N competitors with applications to action selection and compulsive behaviors. *International Journal of Bifurcation and Chaos*, *24*(12), 1450156.

Fukai, T., & Tanaka, S. (1997). A simple neural network exhibiting selective activation of neuronal ensembles: From winner-take-all to winners-share-all. *Neural Computation*, *9*(1), 77–97.

Gazdar, G., Klein, E., Pullum, G., & Sag, I. (1985). *Generalized phrase structure grammar*. Oxford, UK: Basil Blackwell.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass*, *5*(8), 509–524.

Guckenheimer, J. (1983). *Holmes: Nonlinear oscillations, dynamical systems and bifurcations of vector fields*. Berlin: Springer.

Haken, H. (1983). *Synergetics: An introduction*. Berlin: Springer.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. 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.

Hale, J. T. (2011). What a rational parser would do. *Cognitive Science*, *35*(3), 399–443.

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language*, *48*(4), 760–778.

Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, *43*(3), 525–546.

Hudson, R. A. (2007). *Language networks: The new Word Grammar*. Oxford, UK: Oxford University Press.

Humphreys, K. R., & Bock, K. (2005). Notional number agreement in English. *Psychonomic Bulletin & Review*, *12*(4), 689–695.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137–194.

Keenan, E. L. (1976). Towards a universal definition of "subject". In C. N. Li (Ed.), *Subject and topic* (pp. 303–333). New York: Academic Press.

Kempen, G., & Vosse, T. (1989). Incremental syntactic tree formation in human sentence processing: A cognitive architecture based on activation decay and simulated annealing. *Connection Science*, *1*(3), 273–290.

Konieczny, L. (2005). The psychological reality of local coherences in sentence processing. In B. Bara, L. Barsalou, M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1178–1183). Stresa, Italy: Cognitive Science Society.

Koptjevskaya-Tamm, M. (2001). "A piece of the cake" and "a cup of tea:" Partitive and pseudo-partitive nominal constructions in the Circum-Baltic languages. In Ö. Dahl & M. Koptjevskaya-Tamm (eds.), *Circum-baltic languages. Volume 2: Grammar and typology* (pp. 523–568). Amsterdam: John Benjamins.

Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 326.

Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science*, *35*(6), 1009–1051.

Lau, E., Wagers, M., Stroud, C., & Phillips, C. (2008). Agreement and the subject of confusion. Paper presented at the 21st annual CUNY sentence processing conference. Chapel Hill, NC: University of North Carolina.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

Lotka, A. J. (1920). Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences*, *6*(7), 410–415.

Mahowald, K., Graff, P., Hartman, J., & Gibson, E. (2016). SNAP judgments: A small N acceptability paradigm (SNAP) for linguistic acceptability judgments. *Language*, *92*(3), 619–635.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, *88*(5), 375–407.

McDonald, R., Nivre, J., Quirmbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zhang, H., Täckström, O., Bedini, C., Castelló, N. B., & Lee, J. (2013). Universal dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (pp. 92–97). Sophia, Bulgaria: Association for Computational Linguistics.

Mehler, J. (1963). Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, *2*(4), 346–351.

Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist*, *17*(11), 748.

Paape, D., & Vasishth, S. (2016). Local coherence and preemptive digging-in effects in German. *Language and Speech*, *59*(3), 387–403.

Patson, N. D., & Husband, E. M. (2015). Misinterpretations in agreement and agreement attraction. *The Quarterly Journal of Experimental Psychology*, *69*(5), 950–971.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427–456.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago, IL: University of Chicago Press.

Reynolds, C. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, *21*(4), 25–34.

Rutkowski, P. (2007). The syntactic structure of grammaticalized partitives (pseudo-partitives). *University of Pennsylvania Working Papers in Linguistics*, *13*(1), 26.

Schlueter, Z., Parker, D., & Lau, E. (2017). (Mis)interpreting agreement attraction: Evidence from a novel dual-task paradigm. Paper presented at the 30th Annual CUNY Conference on Human Sentence Processing, Cambridge, MA.

Schneider, E., Eschman, A., & Zuccoloto, A. (2012). *E-prime user's guide*. Pittsburgh, PA: Psychological Software Tools Inc.

Selkirk, E. (1977). Some remarks on noun phrase structure. In A. Akmajian, P. Culicover, & T. Wasow (Eds.), *Studies in formal syntax* (pp. 285–316). New York: Academic Press.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237* (4820), 1317–1323.

Smith, N., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*, 302–319.

Solé, R. V. (2011). *Phase transitions*. Princeton, NJ: Princeton University Press.

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, *43*(1), 155–167.

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, *60*(2), 308–327.

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition*, *114*(3), 447–454.

Stevenson, S. (1994a). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, *23*(4), 295–322.

Stevenson, S. (1994b). A competitive attachment model for resolving syntactic ambiguities in natural language parsing. Unpublished doctoral dissertation, University of Maryland, College Park, MD.

Stickney, H. (2009). The emergence of DP in the partitive structure. Unpublished doctoral dissertation, University of Massachusetts, Amherst, MA.

Strogatz, S. H. (1994). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Boulder, CO: Westview Press.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*(4), 355–370.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology-Learning Memory and Cognition*, *30*(2), 431–449.

Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *237*(641), 37–72.

Van der Velde, F., & de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences*, *29*(1), 37–70.

Van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, *13*(2), 22–30.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, *61*(3), 261–298.

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes*, *11*(4), 407–442.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*(2), 105–143.

Vosse, T., & Kempen, G. (2009). The Unification Space implemented as a localist neural net: Predictions and error-tolerance in a constraint-based parser. *Cognitive Neurodynamics*, *3*(4), 331–346.

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

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Details of the experiments and materials.

Appendix S2. Parameters and details of the model.