Toward a Theory of Timing Effects in Self-Organized Sentence Processing

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

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

Abstract

Many theories of sentence processing are based on the idea that a discrete, symbolic grammar defines all of the structures relevant for parsing and that it is the parser's job to select from those structures the one that best fits its input. However, local coherence effects, where people's parsing behavior suggests they are entertaining locally viable but globally impossible structures, suggest that this may not always be the case. We introduce a self-organizing sentence processing (SOSP) model of local coherence effects and use it to demonstrate how predictions about timing effects (a major source of psycholinguistic data and major shortcoming of many previous dynamical parsers) can be derived directly from a harmony (well-formedness) function covering both grammatical and ungrammatical structures. The framework we describe is widely applicable to interesting psycholinguistic phenomena, which will facilitate future quantitative comparisons with more established, grammar-supervised theories like surprisal and ACT-R.

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

Introduction

The current, most fully-developed models of online sentence processing adopt an assumption which may be called *grammar supervision*. With grammar supervision, a symbolic grammar specifies the universe of structures that will be entertained. An example is surprisal theory (Hale, 2001; Levy, 2008), which claims that the parser distributes probability over all grammatical structures compatible with the current input at each point. The processing time for each word is proportional to how much change in the probability distribution is needed after incorporating a new word (as measured by the Kullback-Leibler divergence between prior and posterior distributions estimated from a large corpus). In other words, the grammar acts as a kind of overseer, ensuring that the parses maintained up to any point are compatible with the grammar. This kind of theory has been massively successful in capturing variance in reading times and other measures in both experimentally designed stimuli and natural corpora (Levy, 2008; N. J. Smith & Levy, 2013).

However, empirical studies over the past several decades have identified a number of phenomena that challenge the grammar-supervision hypothesis. We focus on local coherence effects (see Ex. (1); Tabor, Galantucci, & Richardson, 2004; Cai, Sturt, & Pickering, 2012; Paape & Vasishth, 2015; Kukona, Cho, Magnuson, & Tabor, 2014; Konieczny, Müller, Hachmann, Schwarzkopf, & Wolfer, 2009; Konieczny, 2005; Bicknell, Levy, & Demberg, 2009; Levy, Bicknell, Slattery, & Rayner, 2009). Early-arriving words make it so that, if the grammar were supervising, only one parse would be possible, but when later words are perceived, people nevertheless show evidence of entertaining a second, conflicting parse, which is motivated by the later-arriving words. For example, the reduced forms in of Ex. (1) (i.e., without *who was*) both showed slowed reading at *tossed/thrown* relative to the unreduced form, but this effect was significantly greater for (1-a) than for (1-b) (Tabor et al., 2004).

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

We can make sense of this result if we assume that the words *the player tossed...* cause the parser to construct an active clause with *the player* as its subject, even though English grammar mandates that, in this context, *tossed* be a passive verb heading a reduced relative clause modifier of *the player*.[1] Such a parsing process is inconsistent with grammar-supervision theories, but it is precisely how self-organized approaches to sentence processing explain these effects.

Self-organized sentence processing

Self-organized sentence processing (SOSP; Kempen & Vosse, 1989; Stevenson, 1994; Tabor & Hutchins, 2004; van der Velde & de Kamps, 2006; Vosse & Kempen, 2000, 2009; Cho, Goldrick, & Smolensky, 2017; G. Smith, Franck, & Tabor, 2018; Gerth & beim Graben, 2009; Villata, Tabor, & Franck, 2018)) is an approach to modeling sentence processing which has provided insight into a wide range of known phenomena and which does not assume grammar supervision. Instead, large-scale sentence structures self-organize via continuous, local, bottom-up interaction among small pieces of syntactic tree structure (treelets) activated by the words that have been perceived or are being produced. Other examples of local interactions giving rise to higher-level structure include biological morphogenesis (Turing, 1952), fluid convection (Koschmieder, 1993), and laser light (Haken, 1983). In SOSP, feedback interactions among the treelets generally

[1] Levy et al. (2009) have argued that the effects in Tabor et al. (2004) can be accounted for if the surprisal framework is combined with a *noisy channel* assumption—words may be misperceived (e.g., *at* was actually *and* in Ex. (1). But other cases of local coherence effects are not plausibly amenable to this explanation (Kukona et al., 2014; Paape & Vasishth, 2015).

Figure 1: A snap-shot of an SOSP processor parsing “The coach teased the player tossed....”

drive the formation of structure consistent with the grammar, but when two (or more) incompatible structures receive enough bottom-up support, the system can stabilize in a state of conflict, causing processing difficulty. Such models have produced plausible accounts of a range of well-known sentence processing phenomena, including center-embedding vs. right embedding garden path effects, lexical ambiguity processing, (Vosse & Kempen, 2000), length effects (Tabor & Hutchins, 2004), parse alternation in global ambiguity **citation?**, and agreement attraction (G. Smith et al., 2018).

Although SOSP models have been proposed for a variety of established phenomena, the accounts are heterogeneous, so it is not clear yet whether they constitute a unified treatment. Also, oddly, there are relatively few results on timing data, even though timing data are the most common kind of data studied by psycholinguists and even though self-organization is generally understood via dynamical systems theory, the mathematical domain specifically concerned with interactions among variables over time. In this paper, we introduce a framework for SOSP that helps address these shortcomings. Influenced by the work of Smolensky (2006) and Haken (1983), we define a harmony function (known in other domains as a potential or energy function Smolensky, 1986) that specifies the global goodness of the system state (configurations of features on attachment sites and links between attachment sites, see Fig. (1)), employing a systematic method of deriving the harmony function from ~~a parsed corpus of sentences~~ local information about structural goodness. This method can be applied to a range of psycholinguistic phenomena and is grounded in established linguistic insights about sentence structure.

The harmony function can be thought of as a hilly landscape with peaks at the activation patterns corresponding to locally well-formed structures, including both fully grammatical structures and the clash states mentioned above (Fig. (1)). The dynamics of sentence processing are significantly determined by the gradient of harmony, so the system moves uphill increase its current harmony, modulo low-magnitude noise. Our specification of the harmony function and dynamics leads to a theory of timing effects in sentence processing, in which, all other things being equal, a higher-harmony parse is built faster than a lower-harmony one. This is simply because higher peaks have steeper gradients, which causes the system to move faster toward the peak. In sentences where multiple structures are competing, the system will sometimes ascend one peak and sometimes another (due to noise), and its path will be more curved if the peaks are more equal competitors. Therefore, average processing times are a function of how the noise causes the system to choose among the peaks and how much curvature the paths exhibit.

We first present our SOSP framework and show how it

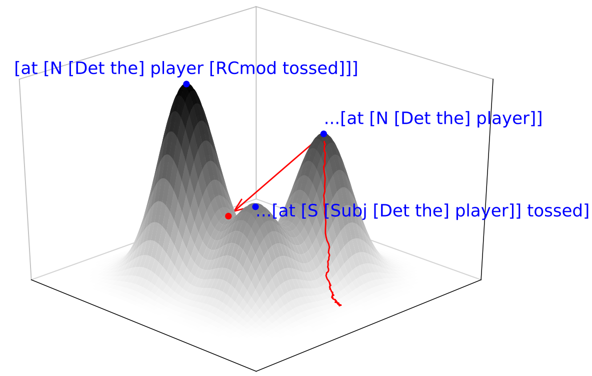

Figure 2: A partial harmony surface for English. The vertical axis is harmony, and the other two dimensions code feature/link configurations. After reading *the coach smiled at the player*, the system follows the trajectory shown in read. The system dynamics drive it to build a well-formed partial parse with *the player* attached as the nominal dependent of *at* at the peak labeled [at [N [Det the] player]. Once it has stabilized on that structure, *tossed* is read, jumping the system (red arrow) to a point intermediate between the grammatical [at [N [Det the] player [RCmod tossed]]] and the locally coherent but low-harmony [at [S [Subj [Det the] player] tossed]. From there, the system will settle again until it gets close enough to one of the peaks, then the process repeats.

makes timing predictions. We then turn to an implemented model SOSP model of local coherence effects, illustrating the principles just outlined. Finally, we conclude with a discussion of how SOSP relates to existing theories of sentence processing.

A new formalization of SOSP

In SOSP, linguistic structures are built out of lexically anchored syntactic treelets. The treelets connect with each other via graded attachment links (Fig. (1)). We assume for simplicity a dependency grammar formalism (e.g., Hudson, 2007; McDonald et al., 2013), so the only attachment sites are ones allowing a word to attach as the dependent of another word (*head* attachment sites) and ones that allow other words to attach as dependents (*dependent* attachment sites). The head and dependent attachment sites on each treelet are feature vectors encoding syntactic and semantic properties of a word and its expected dependents, respectively. Some features can change (e.g., the number marking on the determiner *the* depends on the number of its licensor), and others are fixed in the lexicon. The only constraints on which links can form are that 1) no links can form within a single treelet (e.g., a determiner dependent site on a noun cannot link to the head of that same noun), 2) links can only form between head attachment sites dependent attachment sites, i.e., no head-

head or dependent-dependent links can form.[2] All other links, grammatical and ungrammatical, are allowed to form.

To facilitate intuitive interpretation of the values on each dimension, we use the convention of placing linguistic structures at the corners of the unit hypercube (0 = feature or link off, 1 = on). In order to allow multiple tokens of the same treelet in one sentence (e.g., *the* in *the dog saw the cat*), all of a treelet's dimensions are repeated for every position in a sentence. Thus, there is a set of dimensions corresonding to *the* as the first word of a sentence, a different set of dimensions for *the* as the second word, etc. Links are therefore between sentence-position-specific instances of treelets.

Not all attachment links make equally well-formed structures, though. Structures in which all linked feature vectors are perfectly matched receive the highest possible harmony value of 1. Any feature mismatch results in a lower harmony value for that structure. In this way, SOSP implements a graded notion of well-formedness.We quantify the local harmony h_i of a (partial) lingustics structure i, i.e., degree of well-formedness for i's configuration of features and links, using Eq. 1:

$$h_i = \prod_{l \in links} \frac{dist(\mathbf{f}_{l,head})^\intercal \rho(\mathbf{f}_{l,dependent})}{len(\mathbf{f}_{l,head})} \tag{1}$$

The local harmony h_i of a structure is the product of the Hamming distance $dist(\cdot)$ between the ~~absolute value ($|\cdot|$) of the dot product of the~~ feature vectors at the head $\mathbf{f}_{l,head}$ and dependent $\mathbf{f}_{l,dependent}$ ends of each link l scaled by the length of the feature vectors ($len(\cdot)$). This definition of local harmony is valid for any combination of features and links, even those that strongly violate rules of a symbolic grammar, e.g., the clash structure [*at* [S [Subj [Det *the*] *player*] *tossed*…]]). In the simulations below, we will see that the presence of these lower-harmony structures in the mental representation of possible structures plays a key role in explaining observed timing effects.

Eq. 1 allows us to calculate the harmony of any linguistic configuration, but on their own, the h_is do not tell us how to choose a structure given the input. Thus, we need a way of relating the different structures and a mechanism for navigating among them given some input.

Defining the harmony landscape and dynamics

A simple method for defining where the peaks in our harmony function should be is to use a sum of radial basis functions (RBFs) ϕ_i (Han, Sayeh, & Zhang, 1989; Ciocoiu, 1996, 2009; Muezzinoglu & Zurada, 2006):

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

Here, $\mathbf{x}$ (a column vector) is the d-dimensional state of the system encoding values of all features and links in $\mathbb{R}^d$, each $\mathbf{c}_i$ is the location of the ith (partial) parse (encoding desired feature values and link strengths), $^\intercal$ denotes the vector transpose[3], and γ (a free parameter) sets the width of the RBFs. We then define the harmony function $H(\mathbf{x})$ as the sum of n RBFs, where n is the number of partial and full parses (harmony peaks) we wish to encode:

$$H(\mathbf{x}) = \sum_i^n h_i \phi_i(\mathbf{x}) \tag{2}$$

where the h_i give the local harmony of a (partial) parse, computed using Eq. 1. This equation creates the hilly harmony landscape mentioned above, assigning harmony values both to the $\mathbf{c}_i$ and to all states intermediate between them.[4]

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

$$\frac{d\mathbf{x}}{dt} = \nabla_\mathbf{x} H(\mathbf{x}) = -\tfrac{2}{\gamma}\sum_i^n h_i(\mathbf{x}-\mathbf{c}_i)\phi_i(\mathbf{x}) + \sqrt{2D}\, dW \tag{3}$$

D scales the magnitude of the Gaussian noise process dW. Gradient dynamical systems of this sort exhibit simple behavior: given an initial condition, the system simply settles into the nearest attractor (neglecting the effects of the noise) (Hirsch & Smale, 1974). Because we choose the locations of the (partial) parses $\mathbf{c}_i$, this setup guarantees that the system will settle into one of the pre-programmed attractors that corresponds to a symbolic structure, as long as the parameters are set so that all of the $\mathbf{c}_i$ exist as separate harmony peaks.

The approach in Eqs. 1-3 can be applied to any linguistic structure describable in terms of features and attachment links, making SOSP a general theory of sentence processing. The parsing dynamics are derived directly from the harmony function, so SOSP provides a direct mapping between the well-formedness of linguistic structures and how the system will behave when parsing that structure. We now show how this feature of SOSP leads directly to predictions about processing times.

[2] We allow links to fail to form, creating partial, fragmentary structures, however, these parses are penalized and receive lower local harmony.

[3] Note that $(\mathbf{x}-\mathbf{c}_i)^\intercal(\mathbf{x}-\mathbf{c}_i)$ is equivalent to the square of the Euclidean distance between $\mathbf{x}$ and $\mathbf{c}_i$.

[4] One question that arises with this system is what parameter settings ensure separate harmony peaks for each $\mathbf{c}_i$. Exploratory simulations and numerical bifurcation analyses (Meijer, Dercole, & Oldeman, 2009) using a one-dimensional system with peaks at $x = 0$ and $x = 1$ suggest that the harmony peaks remain separate as long as γ is small enough. When the h_i are equal, there is a pitchfork bifurcation at $\gamma = 0.5$, when the two separate harmony peaks merge into a single peak halfway between the $\mathbf{c}_i$ (Muezzinoglu & Zurada, 2006, report a similar finding). For unequal h_i, the system exhibits a cusp bifurcation, with the lower-harmony peak being absorbed into the larger harmony peak for values of γ that, in general, are lower than 0.5. As discussed below, even if the lower harmony parse does not have a separate peak, it can still affect parsing by warping the harmony surface and thereby deflecting trajectories from a more direct path to an attractor. A more systematic exploration of the parameter space is left to future work, but these explorations allow us constrain γ and the h_i somewhat in our simulations.

Predicting processing times

There are several ways to illustrate how settling times depend on the harmony of the parse that forms. To illustrate, we will first consider the simplest possible case, a one-dimensional system with a single harmony peak at $x = 0$. The harmony function is $H(x) = h\ \phi(x) = h\ \exp\left(-\frac{x^2}{\gamma}\right)$ and the dynamics are given by $\dot{x} = -\frac{2h}{\gamma}\ x\ \phi(x)$. From this equation, we can already see that the higher the harmony of the attractor, the faster system moves toward it. Another way of seeing this is to consider the time dt it takes to travel an infinitesimal distance dx,

$$dt = dx/\dot{x} = \left(-\frac{2h}{\gamma}\ x\ \phi(x)\right)^{-1}\ dx, \tag{4}$$

since time equals distance divided by velocity. To find the time t_s to settle the from an initial point x_0 at $t = 0$ to a point x_1 near the attractor at $x = 0$[5], we can simply integrate both sides of Eq. 4:

$$\begin{aligned}\int_0^{t_s} dt &= \int_{x_0}^{x_1} \left(-\frac{2h}{\gamma}\ x\ \phi(x)\right)^{-1}\ dx \\ t_s &= \frac{\gamma}{2h}\int_{x_0}^{x_1} -\frac{1}{x}\exp\left(\frac{x^2}{\gamma}\right)\ dx \\ t_s &\propto (2h)^{-1}\end{aligned} \tag{5}$$

Thus, the time it takes to settle to a point close to the attractor in this 1D system is inversely proportional to two times the harmony of the structure. (The integral on the rhs. can be calculated numerically, and since we use the same γ for all attractors, its effect on settling times is constant. The relation between well-formedness and settling times is therefore quite simple: well-formed structures are faster to build than ill-formed structures.

In general, though, an SOSP parser will have many dimensions coding multiple features and link strengths, and there will be many attractors corresponding to the different structures that can form. Does something like Eq. 5 hold in the general case? Once the system has entered the basin of attraction for a particular attractor, the ~~most relevant~~ strongest force acting on the system is the pull of that attractor; the effects of all of the other $\mathbf{c}_i$ drop off exponentially. As an approximation, we can simplify Eq. 3 so that the only element in the sum is the on for the selected attractor, similar to Eq. **??**. From there, it is simple to see that the same relation between settling time and harmony in Eq. 4 holds[6,7].

[5] It takes infinitely long to actually reach $x = 0$, so we make x_1 offset a small amount from the fixed point so the integral converges.

[6] For $d > 1$, the integral in Eq. 5 needs to be replaced with a line integral connecting the initial and final conditions.

[7] Yet another way of seeing the dependence of settling times on the local harmony is to find the characteristic time scale of the attractor (Strogatz, 1994), which involves finding the eigenvalues of its Jacobian matrix (linearization) and taking the reciprocal of the largest of them (all eigenvalues of an attractor are negative). For Eq. **??**, the linearization is $\dot{x} = -(2h/\gamma)x$, so the characteristic time scale is $\gamma/2h$, as in Eq. 5.

However, the effects of other attractors are, in general, not completely negligible. As illustrated in Fig. 3, the presence of a relatively high-harmony attractor can temporarily bow trajectories away from an attractor by warping the harmony landscape, even though the system is not in the basin of attraction of the lower-harmony competitor. So, within the basin of attraction of a particular parse, the settling time is approximately inversely proportional to double the harmony of that parse, modulo the noise and this bowing.

Word-by-word parsing works by turning on the features of a word at a particular position in the sentence. This places the state of the system away from an attractor, since attractors correspond to stable configurations of features *and* links, and the input leaves the links untouched (i.e., at zero strength). The system then settles towards one of the nearby attractors according to the harmony gradient and the noise. Over repeated trials, the noise will drive the system to settle into different parses. Thus, the theory of timing in SOSP is this: the local harmony of different structures determines approximately how fast that structure forms (modulo noise and trajectory bowing from competing attractors), and the average processing time at a given word over many trials is the weighted average of the settling times to each parse chosen. We now illustrate how this works using a simple model of local coherence effects.

A simple SOSP model of local coherence effects

A full model of the word-by-word processing of the sentences in (1) would involve incrementally turning on features of the first word *the*, allowing the system to settle to the nearest attractor, then turning on the features of *coach*, and repeating until the end of the sentence. We can model the main local coherence finding from Tabor et al. (2004), however, by assuming that the parser has already read up to *The coach smiled at the player tossed/thrown...* and that the main settling that still has to be done is choosing how to attach *player* and *tossed/thrown*, as shown in Fig. (1). We do this using a two dimensional system:

$$\begin{aligned}\phi_1(x_1,x_2) &= \exp\left(\frac{(x_1-1)^2+x_2^2}{\gamma}\right) \\ \phi_2(x_1,x_2) &= \exp\left(\frac{x_1^2+(x_2-1)^2}{\gamma}\right) \\ \dot{x_1} &= -\frac{2}{\gamma}(h_1(x_1-1)\phi_1(x_1,x_2) \\ &\quad + h_2x_1\phi_2(x_1,x_2)) + \sqrt{2D}\ dW_1 \\ \dot{x_2} &= -\frac{2}{\gamma}(h_1x_2\phi_1(x_1,x_2) \\ &\quad + h_2(x_2-1)\phi_2(x_1,x_2)) + \sqrt{2D}\ dW_2\end{aligned} \tag{6}$$

The two dimensions (x_1 and x_2) are the strength of the *tossed/thrown-player* link (fully grammatical with *player* as the head of the participle; link 1 in Fig. (1)) and the strength of the *player-tossed/thrown* link (ungrammatical, with *player* as the subject dependent of *tossed/thrown* coerced to be a main verb; link 2 in Fig. (1)). We need only two attractors in the system: one at [1, 0], which is the correct, maximal-harmony parse, and one at [0, 1], which will have different sub-maximal harmonies depending on whether *tossed* or *thrown* has been read (see Fig. 3). Because *player* is a good feature match to be the subject of *tossed* (when interpreted as an active verb), the attractor at [0, 1] is penalized only for having a missing link between *at* and *player*. For *thrown*, though, [0, 1] is additionally penalized because the features on the participle *thrown* do not match *player*'s feature that specifies that it should be dependent on a main verb. **Why else should it be penalized?**. We assume that the system starts at [0, 0], reflecting the assumption that only the links connecting the words need to develop.

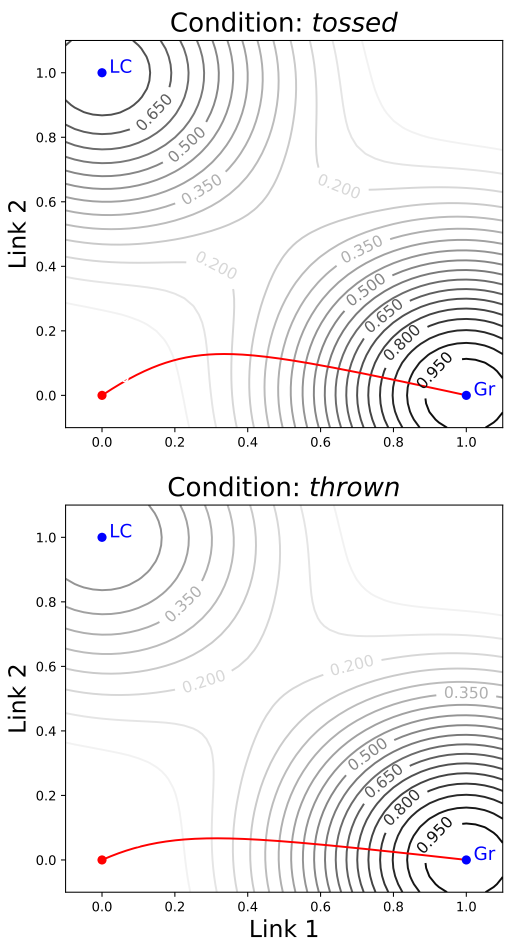

Figure 3: Contour plots of the harmony landscapes. Contour labels give the harmony at that level. Red lines show noiseless trajectories starting from [0, 0] and approaching the grammatical parse (Gr) at [1, 0]. Note how the trajectory for *tossed* is more bowed toward the locally coherent attractor (LC; causing extra slowdown) compared to *thrown*.

We simulated both conditions 2000 times using Euler forward discretization with a time step of 0.01. The noise magnitude D was set to 0.001, and γ was 0.25. The system ran until it got within a small radius of an attractor. For illustration purposes, the local harmony h_1 of the locally coherent attractor ([0, 1]) was set to 0.75 in the *tossed* condition, and in the *thrown* condition to 0.5 (see below for a more systematic investigation of h_1's effect). SOSP predicts that, for both *tossed* and *thrown*, the noise should bump the system towards the grammatical parse in most cases because its high harmony dominates the harmony landscape. When the noise does push the state towards the locally coherent attractor, it will approach it more slowly in the *thrown* condition than in the *tossed* condition because of *thrown*'s lower harmony. But because this happens so rarely, the average time will be dominated by fast approaches to the grammatical attractor. The low-harmony parse for *tossed* will be selected more often, though, due to its higher harmony, so it will pull the average settling time down more than the *thrown* condition. Thus, the presence of a relatively high-harmony competitor for the grammatical parse will cause a competition-based slowdown averaged over many trials.

As predicted, the system settled into the ungrammatical attractor in both cases, and it did so more frequently in the *tossed* condition (about 12% of runs) than in the *thrown* condition (<1% of runs). This caused the average settling time to be higher for *tossed* (in time steps: $M = 158.588, SD = 28.133$) than for *thrown* ($M = 149.459 SD = 25.576$), paralleling the results of Tabor et al. (2004).

These simulations are based on the assumption that the local harmony of the locally coherent parse is higher for *tossed* than *thrown*, making the former a stronger competitor than the latter. However, it is not immediately clear which features are relevant for calculating the h_i such that this relation holds.[8] For now, we plot mean settling times as a function of the harmony h_1 of the ungrammatical parse (Fig. 4). We used $\gamma = 0.25$ here, but this pattern holds for a wide range of γ values. This figure shows that there is broad range of h_1 values such that as long as the *tossed* condition has a higher h_1 than the *thrown* condition, we will observe local coherence effects. Thus, local coherence effects are generally predicted whenever we compare higher- and lower-harmony competitors for a grammatical parse, a result supported by the large-scale eye-tracking corpus study of Bicknell et al. (2009).

Something interesting happens when h_1 is greater than about 0.85, though. As h_1 increases beyond that point, the average settling times start to drop. The bottom panel of Fig. 4 suggests why: as the competing ungrammatical parse increases in harmony, the time it takes the system to settle to it approaches that of the grammatical parse, so it no longer pushes the overall average settling time up as much. In this range of h_1 values, there is still a slowdown due to competition, but it is not as large as the slowdowns observed for somewhat lower-harmony competitors. In effect, the model predicts that we should observe the strongest competition-induced slowdowns when the competing structure is of mod-

[8] Future work will seek to learn relevant dependency-specific features from a large corpus (Zhao, Huang, Dai, Zhang, & Chen, 2014; Bansal, 2015), which will open the door to broad-coverage analyses of reading times similar to N. J. Smith and Levy (2013) and Engelmann, Jäger, and Vasishth (2018).

erate harmony, and that we should observe smaller-magnitude slowdowns for both very low harmony competitors, and (to a lesser extent) higher harmony competitors. This behavior is, to our knowledge, unique among models of sentence processing. We speculate that this property of SOSP might provide a new explanation for the ambiguity advantage (Traxler, Pickering, & Clifton, 1998, e.g.), where certain ambiguous relative clause and adjunct attachments are read more quickly than comparable unambiguous structures. This effect has been argued to rule out competition-based theories of parsing, but if the h_i of both parses are close to 1 in the ambiguous condition but one is appreciably less than one in the unambiguous conditions, the competition-based SOSP might be able to explain this otherwise puzzling effect.

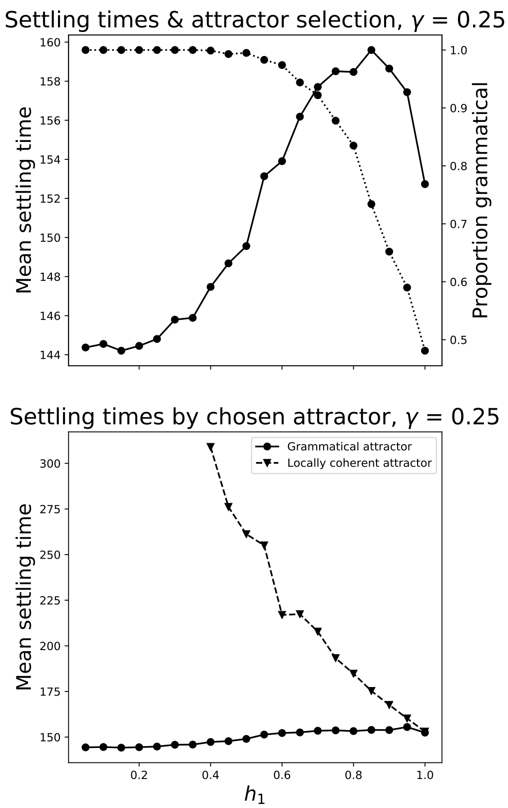

Figure 4: Top panel: mean settling times for the local coherence model as a function of the ungrammatical parse h_1 (solid line, left y-axis) and the proportion of runs in which the system selected the grammatical parse (dotted line, right y-axis). Bottom panel: mean settling time by selected parse (solid line and circles = grammatical, dashed line and triangles = locally coherent parse). For $h_1 < 0.4$, the system never settled on the ungrammatical attractor, likely because that attractor does not exist below that point (see footnote (1)).

Discussion

In this paper, we have introduced a theory of timing effects in a self-organizing sentence processing (SOSP) framework and demonstrated how it can explain the local coherence effects first presented in (Tabor et al., 2004). In SOSP, the amount of time it takes to build a structure depends on how well-formed the structure is, and the average structure-building time over many trials is the weighted average of settling times to each parse chosen.[9] We showed how this theory is derived directly from the equations of SOSP, which were designed to locally find the highest-harmony structure given the input in a harmony landscape that contains both low- and high-harmony structures as attractors.

In the local coherence model, the lower-harmony attractors in the harmony landscape played a key role: A relatively strong but ungrammatical (and therefore slow-to-build) competitor slowed settling more than a weaker ungrammatical competitor because the stronger competitor was built more often. This account differs from the noisy channel approach to local coherence (Levy et al., 2009), where it is assumed that the parser can edit its input to preserve grammaticality (Cho et al., 2017, provide a similar approach in a dynamical, harmony based model). In this way, SOSP is different from grammar supervision theories of sentence processing, where only grammatical structures are considered, e.g., surprisal theory as calculated from probabilistic symbolic grammars (Levy, 2008; Hale, 2001). Also, like surprisal but unlike a number of other sentence processing models (Kimball, 1973; Frazier & Fodor, 1978; Gibson, 1991; Eberhard, Cutting, & Bock, 2005; Lewis & Vasishth, 2005), the approach does not add anything to grammatical theory to avoid or repair parsing failure. Instead, parsing failure is an occasional, but natural outcome of the core structure building process itself—the process that discovers interpretations of word-sequence input. SOSP therefore offers a unique and parsimonious approach to sentence processing: Ungrammatical structures not play a key role in successful and unsuccessful parses, they obviate the need for processing mechanisms other than local harmony maximization.

Acknowledgments

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

References

Bansal, M. (2015). Dependency link embeddings: Continuous representations of syntactic substructures. In *Proceedings of NAACL-HLT 2015* (pp. 102–108). Denver, Colorado: Association for Computational Linguistics.

Bicknell, K., Levy, R., & Demberg, V. (2009). Correcting the incorrect: Local coherence effects modeled with prior belief update. In *Proceedings of the 35th annual meeting of the Berkeley Linguistics Society* (pp. 13–24).

[9] The weighted averaging feature of SOSP is quite similar to recent similarity-based interference approaches building on Lewis and Vasishth (2005) that model reading time effects as statistical hierarchical mixture models (Nicenboim & Vasishth, 2018; Vasishth, Jäger, & Nicenboim, 2017).

Cai, Z. G., Sturt, P., & Pickering, M. J. (2012). The effect of nonadopted analyses on sentence processing. *Language and Cognitive Processes*, *27*(9), 1286–1311.

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

Ciocoiu, I. B. (1996). Analog decoding using a gradient-type neural network. *IEEE Transactions on Neural Networks*, *7*(4), 1034–1038.

Ciocoiu, I. B. (2009). Invariant pattern recognition using analog recurrent associative memories. *Neurocomputing*, *73*, 119–126. doi: 10.1016/j.neucom.2009.02.024

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review*, *112*(3), 531–559.

Engelmann, F., Jäger, L. A., & Vasishth, S. (2018). The effect of prominence and cue association in retrieval processes: A computational account. *Cognitive Science*.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*, 291–325.

Gerth, S., & beim Graben, P. (2009). Unifying syntactic theory and sentence processing difficulty through a connectionist minimalist parser. *Cognitive neurodynamics*, *3*(4), 297–316.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Unpublished doctoral dissertation, Carnegie Mellon University.

Haken, H. (1983). *Synergetics: An introduction* (3rd ed.). Springer-Verlag.

Hale, J. T. (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. doi: 10.3115/1073336.1073357

Han, J. Y., Sayeh, M. R., & Zhang, J. (1989). Convergence and limit points of neural network and its application to pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(5), 1217–1222.

Hirsch, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra*. Academic Press.

Hudson, R. (2007). *Language networks: The new word grammar*. Oxford University 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.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*(1), 15–47.

Konieczny, L. (2005). The psychological reality of local coherences in sentence processing. In *Proceedings of the 27th annual conference of the Cognitive Science Society* (pp. 1178–1183).

Konieczny, L., Müller, D., Hachmann, W., Schwarzkopf, S., & Wolfer, S. (2009). Local syntactic coherence interpretation. evidence from a visual world study. In *Proceedings of the 31st annual conference of the Cognitive Science Society.*

Koschmieder, E. L. (1993). *Bénard cells and taylor vortices*. Cambridge University Press.

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), 326–347. doi: 10.1037/a0034903

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*, 375–419.

McDonald, R., Nivre, J., Quirmbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., ... 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).

Meijer, H., Dercole, F., & Oldeman, B. (2009). Numerical bifurcation analysis. In R. A. Meyers (Ed.), *Encyclopedia of complexity and systems science* (pp. 6329–6352). Springer New York.

Muezzinoglu, M. K., & Zurada, J. M. (2006). RBF-based neurodynamic nearest neighbor classification in real pattern space. *Pattern Recognition*, *39*, 747–760.

Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using bayesian hierarchical modeling. *Journal of Memory and Language*, *99*, 1–34.

Paape, D., & Vasishth, S. (2015). Local coherence and preemptive digging-in effects in German. *Language and Speech*, 1–17. doi: 10.1177/0023830915608410

Smith, G., Franck, J., & Tabor, W. (2018). A self-organizing approach to subject-verb number agreement. *Cognitive Science*.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*, 302–319. doi: 10.1016/j.cognition.2013.02.013

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing, explorations in the microstructure of cognition* (Vol. I: Foundations, pp. 194–281). MIT Press.

Smolensky, P. (2006). Harmony in linguistic cognition. *Cognitive Science*, *30*, 779–801.

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

Strogatz, S. H. (1994). *Nonlinear dynamics and chaos: With applications to physics, biology and chemistry*. Addison-Wesley.

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.

Traxler, M. J., Pickering, M. J., & Clifton, C. J. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*, 558–592.

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*, 37–108.

Vasishth, S., Jäger, L. A., & Nicenboim, B. (2017). Feature overwriting as a finite mixture process: Evidence from comprehension data. In *Proceedings of mathpsych/ICCM.* Warwick, UK.

Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, *9*(2), 1–16.

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

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

Zhao, Y., Huang, S., Dai, X., Zhang, J., & Chen, J. (2014, Oct). Learning word embeddings from dependency relations. In *2014 international conference on asian language processing (ialp)* (p. 123-127). doi: 10.1109/IALP.2014.6973490