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

Dynamical sentence processing models provide a natural explanation for "grammar flouting" interference effects in sentence comprehension and production, like local coherence effects, using bottom-up structure building that at least temporarily entertains ungrammatical structures. Since a major source of data on these and other phenomena comes from timing data (reading times and production latencies), it is important that such models be able to account for timing effects. However, previous dynamical parsing models have only limited coverage of reading time data, and more importantly, there is no unified theory of why different sentences should take different amounts of time to process in a dynamical parser. Here, we present the self-organized sentence processing (SOSP) framework, a dynamical sentence processing model in which the processing dynamics locally maximize a harmony function specifying the well-formedness of linguistic structures (modulo the effect of noise). We show that processing speed is inversely related to the harmony of the structure the parser chooses: higher-harmony structures (even ungrammatical ones) are processed faster than lower-harmony structures. We show that the system provides a mechanistic explanation for local coherence effects. The transparent mapping between linguistic structures and quantitative predictions about processing dynamics in SOSP will facilitate future comparisons with other well-established sentence frameworks like ACT-R and surprisal theory.

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

Introduction

Most of the time, people produce and interpret sentences according to the rules of some grammar. However, they sometimes temporarily entertain or even settle on structures that are not completely faithful to those rules. Here, we introduce self-organizing sentence processing (SOSP) (G. Smith, Franck, & Tabor, in press) as a general theory of reading time effects that we then apply to an interesting case of "grammar flouting" interference in sentence comprehension.

We consider local coherence effects (Tabor, Galantucci, & Richardson, 2004; Konieczny, 2005; Paape & Vasishth, 2015), where the parser seems to entertain a locally coherent structure that is incompatible with the rest of the sentence. For example, Tabor et al. (2004) used sentences like *The coach smiled at the player* ***tossed/thrown*** *the frisbee. . . .* The sequence *the player tossed the frisbee. . .* , on its own, is a grammatical sentence. However, the rest of the sentence in which it appears rules out this structure, since the preposition *at* cannot grammatically take a sentence as its complement (*[PP *at* [S *the player tossed the frisbee. . .*]]). Tabor and colleagues found that participants read the region containing *tossed* significantly longer than the corresponding region for *thrown*, suggesting that the locally coherent but globally illicit parse was competing with the correct parse ([PP *at* [NP *the* [N' *player* [VP *tossed the frisbee*]]]]). This result suggests that people at least temporarily entertain ungrammatical parses and motivates a theory of parsing in which suboptimal structures can influence processing (unlike strictly rational approaches to sentence processing, although see also Levy, Bicknell, Slattery, and Rayner (2009)).

Our focus on timing effects here is motivated by the fact that previous dynamical parsers like SOSP have had little success in predicting timing data, despite the fact that time plays such a central role in these models—they are all implemented as sets of differential equations or iterated maps which describe how the state of a parse changes in time. The dynamical models of Kempen and Vosse (1989); Vosse and Kempen (2009); Tabor and Hutchins (2004) derive some timing predictions for a handful of phenomena via direct simulations, however, a relatively large number of other models (van der Velde & de Kamps, 2006; Vosse & Kempen, 2000; Cho & Smolensky, 2016; Cho, Goldrick, & Smolensky, 2017; G. Smith et al., in press; Gerth & beim Graben, 2009) either do not make timing predictions or the timing predictions are not discussed. More importantly, though, there is no theory of why processing one structure should take longer than another structure in these frameworks. In the SOSP framework we present, timing effects are shown to be directly related to the well-formedness (harmony) (Smolensky, 1986) of the linguistic structure that is built: Sentence structures that are less well-formed take longer to process than more well-formed structures, and the overall average settling time (comparable to averages calculated from experiments) is average of the settling times to each parse chosen weighted by how frequently that parse is selected. While this result seems intuitive, we show below how it follows directly from the calculation of harmony and the dynamical equations that govern how the system parses. Moreover, it makes novel testable predictions that, to our knowledge, are unique among extant theories of sentence processing.

We note that one very promising related result is presented in Cho, Goldrick, Lewis, and Smolensky (2018), where Cho et al. show that, in a related but different dynamical parser (Gradient Symbolic Computation, GSC), the change in harmony from one parsing step to the next is equivalent to surprisal, which is known to predict reading times over a large range of word predictability (Hale, 2001; Levy, 2008; N. J. Smith & Levy, 2013). This effort is closely related to the present paper. In the discussion, we discuss their result in more detail and compare it with our approach.

Self-organized sentence processing

In SOSP, linguistic structures are built out of lexically anchored syntactic treelets. The treelets connect with each other via graded attachment links. We assume for simplicity a dependency grammar formalism (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, and **3) all words must be attached as the dependent of another treelet**[1]. All other links, grammatical and ungrammatical, are allowed to form, e.g., a link can form between the head attachment site of a verb to the determiner attachment site on a noun, which would amount to saying that the verb is the noun's determiner.

Not all attachments are equally well formed, though. Structures in which all linked feature vectors are perfectly match 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 in its structures. To quantify the degree of well-formedness for a particular configuration of features and links, we define a the local harmony h_i of a (partial) linguistic structure i to be:

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

For each link l that is active in the structure, the feature match is calculated as 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 the link scaled by the length of the feature vectors ($len(\cdot)$). We use the convention of placing linguistic structures at the corners of the unit hypercube, so the function $\rho(z) = 2z - 1$ maps the feature vectors element-wise from [0, 1] to [-1, 1], where +1 means a feature is "on," -1 means "off," and 0 means not specified in the lexicon.

This definition of harmony is valid for any combination of features and links, even those that strongly ([PP *at* [S *the player tossed*...]]) or weakly (*the key to the cabinets* ***are***...) violate rules of a symbolic grammar. 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.

The features on every treelet and links connecting treelets are represented as a set of dimensions in a high-dimensional continuous space, with discrete linguistic structures corresponding to corners of the unit hypercube. 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.

Eq. 1 allows us to calculate the harmony of particular linguistic configurations, but on their own, the h_is do not tell us how to choose a structure given the input. Since the h_i encode a person's knowledge of the well-formedness of different possible structures, we need a way for the parser to navigate among the different structures and locally maximize harmony given its input. To do this, we relate different parses by defining a harmony landscape on which the system navigates as it tries to find the best possible structure by local optimization. This harmony landscape was designed to assign well-formedness values for structures intermediate between discrete, symbolic linguistic structures while ensuring that structures are the only attractors of the system (i.e., points to which the system will return after a small perturbation, Strogatz (1994)).

Defining the harmony landscape and dynamics

An SOSP parser should maximize the harmony of the structure it builds given its input; in other words, it should perform hillclimbing on the harmony landscape where the hilltops correspond to discrete, symbolic structures. 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, $\mathbf{x}$ is the d-dimensional state of the system (i.e., values of all features and links) in $\mathbb{R}^d$, $\mathbf{c}_i$ is the location of the ith (partial) parse, $^\intercal$ denotes the vector transpose[2], and γ (a free parameter) sets the width of the RBF. 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:

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

[1]The matrix verb of a sentence attaches as the dependent of a special root node that is not subject to this requirement.

[2]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 $\mathbf{c}_i$.

where the h_i give the local harmony of a (partial) parse, computed using Eq. 1. Exploratory simulations and numerical bifurcation analyses (Meijer, Dercole, & Oldeman, 2009) using a one-dimensional system suggest that harmony peaks remain separate as long as γ is small enough. When the h_i are equal, there is a pitchfork bifurcation at $\gamma = 0.5$, at which point 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 undergoes 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. 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.

The SOSP system should maximize the harmony of the structure it builds given its input. Since the gradient of a scalar-valued function points in the direction of steepest ascent, we define the change in the state of the system in time simply as the gradient of the harmony function:

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

D scales the magnitude of the Gaussian noise process dW. Gradient dynamical systems of this sort exhibit quite 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. (We leave the question of learning the $\mathbf{c}_i$ to future research).

Because the parsing dynamics are derived directly from the harmony function, 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.

Predicting processing times

There are several ways to illustrate how settling times depend on the harmony of the parse that forms. To start, 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

$$\dot{x} = -\frac{2h}{\gamma}\,x\,\phi(x). \tag{4}$$

From this equation, we can already see that the higher the harmony of the attractor, the faster system moves. Since the dynamics drive the system towards the only attractor, the harmony directly affects how fast the system approaches it. Another way of seeing this is to consider the time dt it takes to travel an infinitesimal distance dx,

$$dt = dx/\dot{x} = \left(-\frac{2h}{\gamma}\,x\,\phi(x)\right)^{-1} dx, \tag{5}$$

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 $x = 0$[3], we can simply integrate both sides of Eq. 5:

$$\begin{aligned}
\int_0^{t_s} dt &= \int_{x_0}^{x_1}\left(-\frac{2h}{\gamma}\,x\,\phi(x)\right)^{-1} dx \\
t_s &= \frac{\gamma}{2h}\int_{x_0}^{x_1} -\frac{1}{x}\exp\left(\frac{x^2}{\gamma}\right) dx \\
t_s &\propto (2h)^{-1}
\end{aligned} \tag{6}$$

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 choose to keep γ is constant for all attractors, its effect on settling times is constant). The relation between well-formedness and settling times is therefore quite simple: structures with good feature matches are faster to build than structures with many feature mismatches.

In general, though, an SOSP parser will have many dimensions coding multiple features and link strengths, and there will be multiple attractors corresponding to the different structures that can form. Does something like Eq. 6 hold in the general case? If we can assume that once the system is in the basin of attraction of a particular $\mathbf{c}_i$, the effects of the other attractors is negligible, then we can show that the same relation between settling times and local harmony holds. The ϕ_i in Eq. 3 scale how strongly the system is pulled toward the i-th attractor. So, if ϕ_i becomes negligible, we can simplify Eq. 3 to have just a single element in the sum, similar to Eq. 4. For our purposes, we say that the effect of ϕ_i is negligible if $\phi_i < \theta$, $0 < \theta \ll 1$, since 1 is the minimal distance between attractors. If we denote $\sqrt{\mathbf{x}-\mathbf{c}_i)^{\intercal}(\mathbf{x}-\mathbf{c}_i)}$ (the Euclidean distance) by y, then, after inserting y into Eq. 3 and solving for y, ϕ_i will be negligible as long as $y > \sqrt{-\gamma\, ln\, \theta}$. For example, if we choose $\gamma = 0.25$ and $\theta = 0.1$, then the system should be a distance of at least $y \approx 0.759$ away from the i-th attractor for us to ignore the effects of that attractor. From there, it is simple to see that the same relation between settling time and harmony in Eq. 5 holds[4,5].

[3] It would take infinitely long to actually reach the attractor at $x = 0$, so, in order for this integral to converge, we make the stopping point offset a small amount from the fixed point.

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

[5] 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). To find the characteristic time scale, we find the

So, within the basin of attraction of a particular parse, the settling time is approximately inversely proportional to double the harmony of that parse. Word-by-word parsing works by turning on the features of a word at a particular point 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. From this initial point between attractors, the system settles towards one of the nearby attractors under the influence of the harmony gradient and the noise. Over repeated trials, the noise will drive the system to settle into different parses. The overall mean settling time at a given word is then the mean of the settling times to each attractor chosen weighted by how often the system chooses that attractor. To sum up, the theory of timing in SOSP is this: the local harmony of different structures determines how fast that structure forms, 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

We can model the local coherence effect in Tabor et al. (2004) with a two-dimensional system **(show equations?)**. We assume that the parser has already read up to *The coach smiled at the player tossed/thrown...*. The two dimensions are the strength of the *tossed/thrown-player* link (fully grammatical with *player* as the head of the participle) 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; Fig. **??**, showing the dependency structures). 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. **??**, showing harmony contour plots of the two conditions). 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 only penalized for having a missing link between *at* and *player*. For *thrown*, though, [0, 1] is also penalized because the features on the participle *thrown* do not match *player*'s feature that specifies that it should be dependent on a verb. We assume that the system starts at [0, 0], reflecting the assumption that it has already input and turned on the features of *player* and the relevant participle. The settling time of the model is taken to be the amount of time it takes for the mind to integrate these words in to a linked syntactic structure.

We simulated both conditions 1000 times using the Euler forward discretization with a time step of 0.01. The noise magnitude was set to 0.001, γ was 0.25. In the *tossed* condition, the local harmony of the ungrammatical attractor ([0, 1]) was set to 0.75 (**reason**), in the *thrown* condition to 0.5 (**reason**). The theory above predicts that the *tossed* condition should be slower than the *thrown* condition on average. In both cases, the noise should bump the system towards the grammatical parse in most cases because its high harmony causes it to dominate the harmony landscape. When the noise does push the state towards the low-harmony structure, it will approach it more slowly in the *thrown* condition than in the *tossed* condition because of *thrown*'s much 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, so its lower harmony will pull the average settling time down as much as in the *thrown* condition. Thus, the presence of a relatively high-harmony competitor for the grammatical parse will cause a competition-based slowdown.

As predicted, the system settled into the ungrammatical attractor in both cases, and it did so more frequently in the *tossed* condition (14.6% of runs) than in the *thrown* condition (0.2% of runs). This caused the average settling time to be higher for *tossed* (in time steps: $M = 159.157, SD = 28.492$) than for *thrown* ($M = 150.064, SD = 24.332$).

Discussion

In this paper, we have introduced a theory of timing effects in a self-organizing sentence processing framework and demonstrated how it can explain the local coherence effects first presented in (Tabor et al., 2004). The amount of time it takes to build a structure depends on how well-formed the structure is, and the average time over many trials is the weighted average of settling times to each parse chosen. 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 low-harmony attractors that compete with grammatical structures.

In the local coherence model, the lower-harmony attractors in the harmony landscape played a key role: when there is a relatively strong competitor for a grammatical attractor, runs in which the ungrammatical attractor is chosen increase the average settling time compared to when the competitor is even less grammatical. In this way, SOSP is different from many other theories of sentence processing that only allow the system to construct grammatical parses, e.g., surprisal (Levy, 2008; Hale, 2001), ACT-R (Lewis & Vasishth, 2005), and Gradient Symobolic Computation (GSC) (Cho & Smolensky, 2016; Cho et al., 2017, 2018). The weighted averaging feature of SOSP is quite similar to recent similarity-based interference approaches that model reading time effects as statistical hierarchical mixture models (?, ?; Vasishth, Jäger, & Nicenboim, 2017).

Add discussion of GSC

A similar SOSP model was shown to predict subject-verb number agreement patterns in English pseudopartitives (*a pile of sandwiches is/are on the table...*, and the SOSP framework has been argued to explain encoding interference effects in English and Italian (Villata, Tabor, & Franck,

---

eigenvalues of its Jacobian matrix (linearization) and take the reciprocal of the largest of them (the attractor is stable, so all of its eigenvalues are negative). For Eq. 4, the linearization is $\dot{x} = -(2h/\gamma)x$, so the characteristic time scale is $\gamma/2h$, as in Eq. 6.

2018). The present work thus extends the range of empirical phenomena that SOSP can explain and greatly generalizes the mathematical framework, facilitating further tests of the paradigm with other sentence processing phenomena.

Acknowledgments

IGERT, others?

References

Cho, P. W., Goldrick, M., Lewis, R. L., & Smolensky, P. (2018). Dynamic encoding of structural uncertainty in gradient symbols. In *Proceedings of the 8th workshop on cognitive modeling and computational linguistics.*

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

Cho, P. W., & Smolensky, P. (2016). Bifurcation analysis of a gradient symbolic computation model of incremental processing. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognititve science society* (pp. 1487–1492).

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

Gardiner, C. W. (1985). *Handbook of stochastic methods for physics chemistry and the natural sciences*. Springer.

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

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.

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).

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. (in press). 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.

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.

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.