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

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

Introduction

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

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

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

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

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

[1]Levy et al. (2009) argue that surprisal can account for Tabor et al. (2004) with a *noisy channel* assumption—words may be misperceived (e.g., *at* was actually *and* in Ex. (1-a)). Cho, Goldrick, and Smolensky (2017) present a similar approach in a dynamical model. However, not all local coherence effects are plausibly amenable to this explanation (Kukona et al., 2014; Paape & Vasishth, 2015).

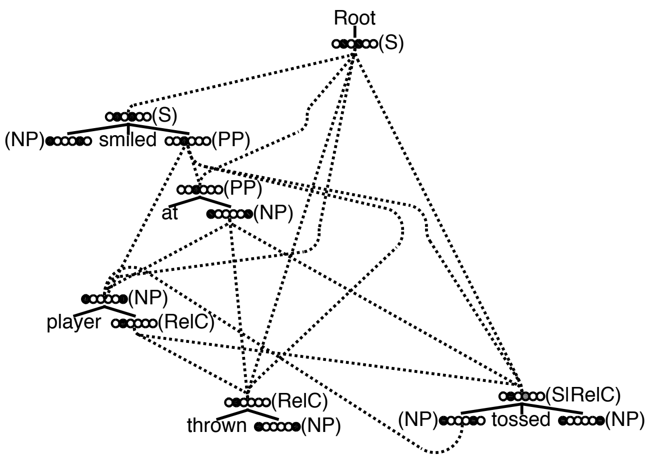

Figure 1: A snap-shot of SOSP-TH parsing a fragment of Ex. (1) showing a subset of competitive treelet interactions. Circles represent features (in order: Nominal, Verbal, Prepositional, Matrix-Clause, Agent, Patient) on attachment sites (labeled in parentheses); phonological forms are unmarked; and the dotted lines are attachment links. Note that even ill-formed structures are included, e.g., *tossed* attaching to Root as the matrix verb instead of the relative clause (RelC) head.

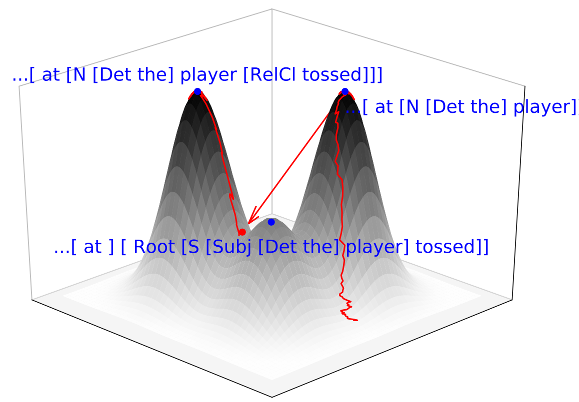

Figure 2: A partial harmony surface illustrating a sample processing path. The vertical axis is harmony, and the other dimensions code feature/link configurations. After reading *the coach smiled at the player*, the noisy dynamics push the system toward a partial parse with *the player* attached as the nominal dependent of *at* at the peak labeled [*at* [N [Det *the*] *player*]]. After stabilizing there, *tossed* is read, jumping the system (red arrow) to a point intermediate between the grammatical [*at* [N [Det *the*] *player* [RelCl *tossed*]]] and the locally coherent, low-harmony [*at*] [Root [S [Subj [Det *the*] *player*] *tossed*]] (with *at* not attached to the subsequent words). From there, the system settles again, in this case selecting the grammatical peak.

Oddly, there are relatively few SOSP results on timing data, even though timing data are the most common kind of psycholinguistic data, and even though self-organization is generally understood via dynamical systems theory, the mathematics of variables interacting in time. Our main contribution here is a novel SOSP framework that addresses this shortcoming by making the relationship between well-formedness and processing times transparent. Influenced by Cho et al. (2017), Smolensky (1986), and Haken (1983), we define a harmony function (also known as a potential or energy function) that specifies the global well-formedness of system states (configurations of features on attachment sites and attachment links, Fig. 1). We employ a systematic method of deriving the harmony function from lexical features in parsed sentences, creating a hilly landscape with peaks corresponding to both fully grammatical structures and conflict states (Fig. 2). The sentence processing dynamics noisily push the system uphill on this landscape to find local harmony maxima. This leads to a theory of timing effects in which, all other things being equal, a higher-harmony parse is built faster than a lower-harmony one. This is because higher peaks have steeper gradients, causing the system to move faster toward the peak. In ambiguous sentences, the system stochastically selects among different peaks, and its path will be more curved if competing peaks are more equally well-formed. Therefore, average processing times over many trials depend on which peaks are selected and how curved the trajectories are.

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

The SOSP-TH framework

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

[2]Links may to fail to form, making fragmentary, low-harmony parses.

root node is available to anchor the whole sentence.

Features and links that are fully "on" and "off" are coded as 1 and 0, respectively. In order to allow multiple tokens of the same treelet in one sentence (e.g., *the* in *the dog saw the cat*), all of a treelet's dimensions are repeated for every position in a sentence. Thus, there is a set of dimensions corresponding to *the* as the first word of a sentence, a different set of dimensions for *the* as the second word, etc. Links (additional dimensions of the system) are between sentence-position-specific instances of treelets.[3]

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

$$h_i = \prod_{l \in links} \left(1 - \frac{dist(\mathbf{f}_{l,head}, \mathbf{f}_{l,dependent})}{nfeat}\right) \tag{1}$$

The local harmony h_i of a structure is the product of one minus the normalized Hamming distances $dist(\cdot)$ between the head feature vectors $\mathbf{f}_{l,head}$ and dependent feature vectors $\mathbf{f}_{l,dependent}$ for each link l. $nfeat$ is the number of elements in the feature vectors. This definition of local harmony is valid for any combination of features and links, even those that strongly violate rules of a symbolic grammar, e.g., the fragmentary, locally coherent structure [at] [Root [S [Subj [Det the] player]] tossed]. In the simulations below, we will see that including these lower-harmony structures in the mental representation of possible structures plays a key role in explaining observed timing effects.

Eq. 1 allows us to calculate the harmony of any linguistic configuration, but on their own, the h_is do not tell us how to choose a structure given the input. To that end, we define a global harmony function and derive the dynamics from it.

Defining the harmony landscape and dynamics

We can define where the peaks in our harmony function are by using a sum of radial basis functions (RBFs) ϕ_i (Han, Sayeh, & Zhang, 1989; 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[4], 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 a hilly harmony landscape analogous to Fig. 2, assigning harmony values both to the $\mathbf{c}_i$ and to all states intermediate between them.

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

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

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

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

Predicting processing times

To derive predictions about processing times, we first consider the simplest possible case, a one-dimensional system with a single harmony peak at $x = 0$. The harmony function is $H(x) = h\,\phi(x) = h\,\exp\left(-\frac{x^2}{\gamma}\right)$ and the dynamics are given by $\dot{x} = -\frac{2h}{\gamma}\,x\,\phi(x)$. From this equation, we can already see that the higher the harmony of the attractor, the faster system moves toward it: Well-formed structures are faster to build than ill-formed structures.[5]

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

[3] This parallels the TRACE model of word perception (McClelland & Elman, 1986), where every position of every word is a node in the model. We agree with the critique that this is neurally implausible and may miss important generalizations. However, TRACE has been very successful at capturing phonological effects in word processing, so we feel this is a reasonable place to start.

[4] $(\mathbf{x}-\mathbf{c}_i)^\intercal(\mathbf{x}-\mathbf{c}_i)$ is the squared Euclidean distance between $\mathbf{x}$ and $\mathbf{c}_i$.

[5] There are other ways to show how settling times in a single trial depend on the harmony of the parse that forms. One is to consider the time dt it takes to travel an infinitesimal distance dx, $dt = dx/\dot{x}$, since time equals distance divided by velocity. Integrating both sides shows the settling time $t \propto (2h)^{-1}$. A third option, linear stability analysis (Strogatz, 1994) provides a similar result.

there will be many attractors corresponding to different structural alternatives. To see that higher harmony still means faster processing, we can approximate Eq. 3 near an attractor i by neglecting all terms $j \neq i$ in the sum in Eq. 3, as the effect of all other attractors drops off exponentially: $\dot{\mathbf{x}} \approx \frac{-2h_i}{\gamma}(\mathbf{x}-\mathbf{c}_i)\phi_i(\mathbf{x})$. It is clear that the same relation between settling time and harmony holds. However, the effects of other attractors are, in general, not completely negligible. Fig. 3 shows how the presence of a relatively high-harmony competitor can bow trajectories away from an attractor by warping the harmony landscape, even though the system is not in the basin of attraction of the competitor.

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

An SOSP-TH model of local coherence effects

A full model of the incremental processing of the sentences in (1) would involve incrementally turning on features of words in their sentence positions, letting the system settle to an attractor associated with a partial parse, and repeating until the sentence ends (see Fig. 1). We can model the main local coherence finding from Tabor et al. (2004) in a focused way by assuming that the parser has already read up to *The coach smiled at the player tossed/thrown...* and that it must now choose how to attach *player* and *tossed/thrown*. We need only two dimensions, one for the grammatical *player-tossed* link and one for the locally coherent *tossed*-Root link. There is thus an attractor at [1, 0] (local harmony $h_0 = 1.0$) and one at [0, 1], which will have different sub-maximal harmonies (h_1) depending on whether *tossed* or *thrown* has been read (see Fig. 3). *Player* is a good feature match to be the subject of *tossed*, and *tossed* can function as a main verb attaching to the root node, so the attractor at [0, 1] is penalized only for leaving *the coach smiled at* unattached to the rest of the structure. For *thrown*, though, [0, 1] is additionally penalized because *thrown* cannot function as a main verb, so its features do not match Root's main-verb dependent features. We start the system at [0, 0], not biased toward either attractor.

SOSP-TH predicts that the noise should bump the system toward the grammatical parse in most cases because its high harmony dominates the harmony landscape. When the noise does push the state toward the locally coherent attractor, it will approach it more slowly in the *thrown* condition than in the *tossed* condition because of *thrown*'s especially low harmony. But because this happens so rarely, the average time will be dominated by fast approaches to the grammatical attractor. The locally coherent parse for *tossed* will be selected more often due to its higher harmony, so it will in-

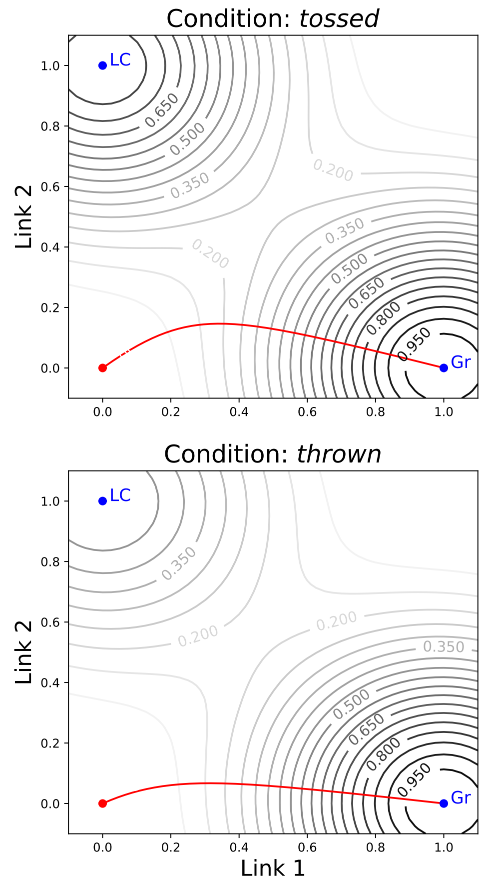

Figure 3: Contour plots of the harmony landscapes used in the local coherence simulations. Contour labels give the harmony at that level. Red lines show noiseless trajectories starting at [0, 0] and approaching the grammatical parse (Gr) at [1, 0]. Note the extra bowing toward the locally coherent attractor (LC) for *tossed*, causing extra slowing compared to *thrown*.

crease the average settling time more than *thrown*. There is also more trajectory bowing for *tossed*, which also slows processing (Fig 3). Thus, a relatively high-harmony competitor for the grammatical parse will, on average, cause a competition-based slowdown.

We simulated both conditions 2000 times using Euler forward discretization with a time step of 0.01, $D = 0.001$, and $\gamma = 0.25$. The system ran until it got within a small radius of an attractor. The local harmony h_1 of the locally coherent attractor ([0, 1]) was set to 0.8 in the *tossed* condition, and in the *thrown* condition to 0.5. As predicted, the system settled to the ungrammatical attractor in both cases, and it did so more frequently in the *tossed* condition (about 14% of runs) than in the *thrown* condition (<1% of runs). This increased the average settling time for *tossed* ($M = 159.073$ time steps, $SD = 27.692$) more than for *thrown* ($M = 149.794, SD = 24.698$), modeling Tabor et al. (2004)'s effect.

These simulations show local coherence effects for one parameter setting, but Fig. 4 shows how the same pattern holds over a wide range of parameter settings. Where it does not hold, there is possibly empirical evidence for a phenomenon that corresponds to the model, different from local coherence. Fig. 4 shows mean settling times as a function of the harmony h_1 of the ungrammatical parse. We used $\gamma = 0.25$ here, but the pattern holds for a wide range of γ values. This figure shows that we will observe local coherence effects as long as $0 <= h_{1,thrown} < h_{1,tossed} < 0.85$. This predicts that local coherence effects should be widespread, a result supported by a large-scale eye-tracking corpus study (Bicknell et al., 2009).

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

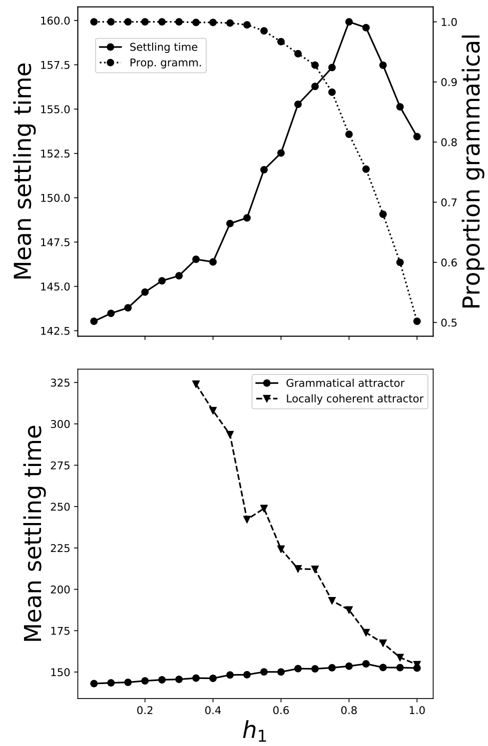

Figure 4: Top: Mean settling times for the local coherence model as a function of the ungrammatical parse h_1 (solid line, left y-axis) and the proportion of runs in which the the grammatical parse was selected (dotted line, right y-axis). Bottom: Mean settling time by selected parse (solid line, circles = grammatical; dashed line, triangles = locally coherent parse). For $h_1 < 0.4$, the system never settled on the ungrammatical attractor. Note the different y-axis ranges.

Discussion

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

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

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

Acknowledgments

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

References

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).

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).

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

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.

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., 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.*

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.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*, 1–86.

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).

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*. doi: 10.1111/cogs.12591

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.

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.

van der Velde, F., & de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences*, *29*, 37–108.

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.