Bifurcation analysis of a Gradient Symbolic Computation model of incremental processing

Pyeong Whan Cho (pcho4@jhu.edu)

Paul Smolensky (smolensky@jhu.edu)
Department of Cognitive Science, Johns Hopkins University
3400 N. Charles Street, Baltimore, MD 21218 USA

Abstract

Language is ordered in time and an incremental processing system encounters temporary ambiguity in the middle of sentence comprehension. An optimal incremental processing system must solve two computational problems: On the one hand, it has to keep multiple possible interpretations without choosing one over the others. On the other hand, it must reject interpretations inconsistent with context. We propose a recurrent neural network model of incremental processing that does stochastic optimization of a set of soft, local constraints to build a globally coherent structure successfully. Bifurcation analysis of the model makes clear when and why the model parses a sentence successfully and when and why it does not—the garden path and local coherence effects are discussed. Our model provides neurally plausible solutions of the computational problems arising in incremental processing

Keywords: Harmonic Grammar; Gradient Symbolic Computation; neural networks; incremental processing; parsing; dynamical systems theory; bifurcations

Introduction

Sentence comprehension requires building a hierarchical constituent structure from sequentially presented words. Psycholinguists agree that human language comprehension is incremental, which means the processing system builds up an interpretation with partial input before a whole sentence is presented. An optimal[1] incremental processing system must solve two closely related computational problems: (a) it must be able to keep multiple possibilities consistent with the information that has been processed, say context. (b) At the same time, it has to reject the interpretations inconsistent with context which may be consistent with bottom-up input.

For clarity, consider a finite language consisting of four sentences {"cats sleep", "cats yawn", "dogs sleep", "dogs yawn"}. A first word (e.g., "cats") of a sentence creates temporary ambiguity which can be resolved after processing a second word (e.g., "sleep") of the sentence. An optimal system must be able to consider both grammatical continuations without choosing one over the other after processing the first word. Simultaneously, the system must be able to reject impossible interpretations before reading a second word (e.g., "sleep") such that it does not take a wrong interpretation (e.g., "dogs sleep").

In this study, we propose a neural network model of incremental processing and show (a) how the proposed model can solve two computational problems in a principled way and (b) when and why the model fails to build a target structure. It will turn out that the model's failure looks like either the garden path effect (Frazier, 1987; Frazier & Rayner, 1982) or the local coherence effect (Tabor, Galantucci, & Richardson, 2004) both of which are well established in the psycholinguistics literature.

The remainder of the paper is organized as follows. First, we briefly describe a phrase structure grammar of our interest. Second, we present an overview of the model; summarize the results without technical details; and discuss the implication of the model. The technical details of the analysis are presented in "Bifurcation analysis of the GSC model", which is followed by a conclusion.

A Grammar of Interest

In the present study, we investigate the Gradient Symbolic Computation (GSC) model (Smolensky, Goldrick, & Mathis, 2014) that implements a phrase structure grammar G, consisting of four rewrite rules: G = {S1 → A B, S2 → A C, S3 → D B, S4 → D C} where S1, S2, S3, S4 are the starting symbols. Grammar G generates a finite language presented in the introduction. We designed the simple language for the following reasons: (a) the language shares the same computational problems with more complex languages. (b) The GSC model implementing a more complex grammar is more difficult to analyze. The simplicity of the grammar makes it possible to investigate the model thoroughly. (c) Parsing a sentence of the language looks trivial but can be difficult for continuous-time recurrent neural network models that utilize highly local constraints and do not explicitly monitor complete structures.

Overview of the Model and Results

The GSC model is a brain-like, continuous-time continuous-state dynamical systems model that can gradually build discrete symbolic structures. There is a transparent mapping between the symbolic structures and the activation patterns (vectors in a continuous vector space) (Smolensky, 1990) so it allows us to study the nature of the intermediate states in the middle of sentence comprehension.[2] The GSC model does stochastic optimization of constraints implementing the

[1] Human language processing may not be optimal due to performance issues. In this study, we investigate the competence of the proposed model.

[2] A popular connectionist alternative is the Simple Recurrent Network (Elman, 1990; Frank, 2009). Although it has been successful in modeling human sentence comprehension performance, the model works in discrete time and develops a rather hard-to-understand internal representation.

grammar and those implementing a force *quantization* pushing towards those neural states that encode fully discrete symbolic trees (for details, see Smolensky et al., 2014). The model maximizes Total Harmony, a measure of goodness of an activation state: $H(\mathbf{a},\gamma) = H_G(\mathbf{a}) + H_B(\mathbf{a}) + \gamma H_Q(\mathbf{a})$ where $\mathbf{a}$ is an activation state and γ (≥ 0) is quantization strength. Grammar Harmony H_G measures how much a state satisfies grammatical constraints and is defined by principles (Hale & Smolensky, 2006). Bowl Harmony H_B measures how close a state is to a baseline activation state. Quantization Harmony H_Q, roughly speaking, measures how close a state is to discrete symbolic states.[3]

In incremental processing, we assume: (a) the model reads the first word of a sentence at 0 of γ. (b) γ increases monotonically in time without assuming any specific function. (c) The model reads the second word at a certain value of γ $(= \gamma_c)$ and then γ continues to increase. The topic of the paper is a formal understanding of how γ's temporal trajectory results in correct parsing, garden path, or local coherence errors.

We performed numerical bifurcation analysis of the model (for a brief introduction, see Meijer, Dercole, & Oldeman, 2009)—the details are presented in the next section. The results suggest that there is a range of γ_c values ($\gamma_1 < \gamma_c < \gamma_2$) that guarantees accurate parsing (case 1). If the model reads the first word (e.g., 'A') of a sentence (e.g., 'A B') too long such that $\gamma_c > \gamma_2$, the model randomly chooses one structure (e.g., [$_{S2}$ A C]) over the other (e.g., [$_{S2}$ A B]) before it reads the second word, althouth both structures are consistent with the first word. It is a garden path error (case 2). If the model reads the second word too quickly ($\gamma_c < \gamma_1$), the model does not reject the structures (e.g., [$_{S3}$ D B] or [$_{S4}$ D C]) inconsistent with the first word (e.g., 'A'). Thus, when the model reads the second word at $\gamma_c (< \gamma_1)$, sometimes the model fails to build the target structure because a non-target structure (e.g., [$_{S3}$ D B]) is consistent with the second word ('B') and it is still considered by the model as a possible grammatical structure. We argue it is an local coherence error (case 3).

We argue our model has a potential as a processing model of incremental structure building. Our model provides neural/mathematical solutions of the computational problems arising in incremental processing; we know when and why the model parses a sentence correctly and when and why it does not. Unlike other constraint satisfaction models (Spivey & Tanenhaus, 1998; Tabor & Hutchins, 2004; Vosse & Kempen, 2000), the GSC model is contructed in a principled way using Harmonic Grammar (Prince & Smolensky, 1997) and tensor product representation (Smolensky, 1990). Its dynamics can be understood thoroughly not with simulations but based on principles of dynamics (see the next section)—the GSC model is not a blackbox. Unlike probabilistic symbolic models of sentence comprehension (Hale, 2001; Levy, 2008), the GSC model describes how the state changes in continuous time and proposes the solutions of computational problems at the algorithmic level. More importantly, the GSC model is not an implementation of structural probabilistic models where the structural hypothesis space is discrete. A blend state in the GSC model—an intermediate state that is located between the states encoding fully discrete structures—is not the representation of a probability distribution across discrete symbolic structures.

The GSC model proposes an interesting way of keeping past and predicting the future. In the GSC model, the present (*blend* state) contains past. Unlike the memory-based model proposed by Lewis, Vasishth, and Van Dyke (2006), there is no need of retrieval processes.[4] Given a word input, the present contains the future as well. In the next section, it will be shown that the GSC model travels along a one-dimensional manifold (a subspace of the full representation space) which can evolve to multiple grammatical structures consistent with both top-down context and bottom-up word input.

The proposed GSC model is not complete. Processing difficulty in sentence comprehension is typically measured in reading times rather than in parsing accuracy in behavioral experiments but the model does not make a prediction on word reading times directly.[5] First, we point out that psycholinguistic models such as the garden path model (Frazier, 1987) and the unrestricted race model (Traxler, Pickering, & Clifton Jr., 1998) propose longer reading times indicate the revision of the initial parse that turns out to be wrong. At this point, the GSC model does not have the ability to revise its interpretation. Parsing failures (cases 2 and 3) overviewed in this section (see also paths [2] and [3] in Figure 4) should be interpreted as the *initial* parsing failure rather than the final product. If we assume the reanalysis requires more time, there is a close relationship between the accuracy of *initial* parsing and the reading times. One way to implement the revision of the initial interpretation is to allow the model to reduce γ when it detects the present blend state is inconsistent with the bottom-up input—probably by detecting a sudden decrease in Grammar Harmony. Second, slow reading times may not necessarily suggest the revision of the initial parse. For example, in the constraint-satisfaction models (MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey & Tanenhaus, 1998; Tabor & Hutchins, 2004), processing delay is observed when multiple interpretations compete with each other without parsing failure. In the GSC model, there may

[3] $H_G(\mathbf{a}) = 0.5\mathbf{a}^{\mathrm{T}}\mathbf{W}\mathbf{a} + \mathbf{b}^{\mathrm{T}}\mathbf{a} + \mathbf{ext}^{\mathrm{T}}\mathbf{a}$; $H_B(\mathbf{a}) = -0.5\beta||\mathbf{a} - \mathbf{z}||^2$; $H_Q(\mathbf{a}) = -0.5\sum_{r \in R}(\sum_{f \in F} a_{f,r}^2 - 1)^2 - 0.5\sum_{f \in F, r \in R} a_{f,r}^2 (a_{f,r} - 1)^2$ where $\mathbf{W}$ and $\mathbf{b}$ are the weight matrix and the bias vector which are constructed by Harmonic Grammar. $\mathbf{a}$ is an activation vector, $\mathbf{ext}$ is an external input vector, $a_{f,r}$ is an acitvation value of a filler/role binding (Smolensky, 1990), F and R are the index sets of fillers and roles, $\mathbf{z}$ (= 0.5 $\mathbf{1}$) is the baseline activation vector, β is the strengh to pull a state to $\mathbf{z}$ and was set to 10. H_Q implements a constraint that only one filler must occupy a role.

[4] We do not argue retrival processes do not involve in human sentence comprehension. The GSC model suggests retrieval may not be the core process in incremental structure building.

[5] The model has a potential to make reading time predictions but at this point, we do not know with what information the model can make a decision when to read a next word. The model may be able to make such decisions by monitoring γ.

be a manifold which takes longer to travel along than other manifolds.

We conclude the GSC model can provide a different approach to incremental processing and may improve our understanding of the process.

Bifurcation analysis of the GSC model

Background

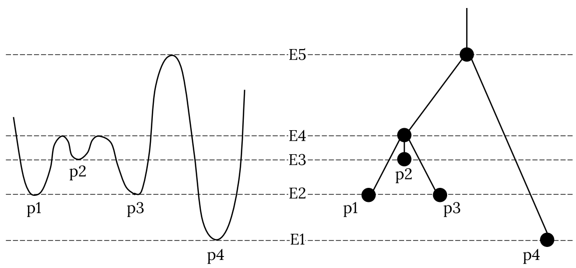

Figure 1: An arbitrary energy landscape (energy = -Total Harmony) and a disconnectivity graph constructured from the energy landscape. p1-p4 indicate the local minima and E1-E5 (E1 $< \cdots <$ E5) indicate the energy levels.

Consider an arbitray energy function $E(x, p)$ in which x is a state variable and p is a parameter. The left side of Figure 1 presents an arbitrary energy landscape (energy = -Total Harmony) when p is set to a certain value. Now consider a system that minimizes the energy function. When no noise is assumed, the system rolls down the hills of the energy landscape (i.e., gradient descent): $dx/dt = -dE(x, p)/dx = f(x, p)$. At a local minimum or a local maximum, the system will not change anymore because the gradient vanishes there. Those states x^* where $f(x^*, p) = 0$ are called *equilibrium points* (or *fixed points*). We can check what happens if a system is slightly displaced from an equilibrium point x^*. If the system approaches x^*, the state (e.g., the local minima p1-p4) is asymptotically stable and often called an *attractor*. If the system moves away from x^*, the equilibrium state (e.g., the local maxima) is unstable.

If the system is high dimensional, it is not easy to visualize the energy landscape. In that situation, we can focus on the topology of the energy landscape by constructing a disconnectivity graph (Wales, Miller, & Walsh, 1998; Becker & Karplus, 1997). The terminal nodes of the disconnectivity graph (the right side of Figure 1) correspond to the attractors (local minima of the energy landscape, or equivalently local maxima of the harmony surface). The nonterminal nodes indicate the height of a ridge (called *energy barrier*) between two local minima. The magnitude of noise required to cross over an energy barrier is determined by the barrier's height (Chiang, Hwang, & Sheu, 1987). For example, the height of the energy barrier from p1 to p2 (= E4-E2) is greater than the height of the energy barrier from p2 to p1 (= E4-E3), suggesting there is a certain level of noise at which the p2-to-p1 transition is possible while the p1-to-p2 transition is not.

The equilibrium points may be created, destroyed, or change their stability when a parameter's value changes. Such qualitative change in the dynamics of the system is called *bifurcation* (Strogatz, 1994) and the parameter values at which bifurcations happen are called *bifurcation points*. When a bifurcation occurs, the topology of the disconnectivity graph changes. In this stuy, we use the continuation method (Meijer et al., 2009) to discover the equilibrium points at different parameter values and detect bifurcations. At the same time, we construct a disconnectivity graph from the set of equilibrium points detected at a particular parameter value to visualize the toplogy of the energy landsacpe. The combination of the two techniques allows us to understand the dynamics of the GSC model thoroughly.

Method

The GSC model maximizes Total Harmony $H(\mathbf{a};\gamma,\alpha)$, or equivalently minimizes energy $E(\mathbf{a};\gamma,\alpha) = -H(\mathbf{a};\gamma,\alpha)$ in which $\mathbf{a}$ is a 27-dimensional activation state vector, γ (≥ 0) is the quantization strength parameter, and α is a variable indicating whether the model is reading the first ($\alpha = 0$) or the second ($\alpha = 1$) word of a sentence 'A B'.[6]

First, we used the continuation method to discover the one-dimensional manifold of equilibrium points in the 28 dimensional vector space (27 state variables + 1 parameter γ) when the first word 'A' was presented (i.e., $\alpha = 0$). To use the continuation method, we should know the equilibrium points at a particular γ value. In the GSC model, it is not difficult. When $\gamma = 0$, we can ignore nonlinear quantization dynamics and at the setting, the GSC model is a linear dynamical system which was constructed to have a single global optimum. The equilibrium point was discovered by the Newton's method. Alternatively, the global attractor can be discovered by integrating the differential equations $\nabla_{\mathbf{a}} H(\mathbf{a}, \gamma, \alpha)$ numerically. Once an equilibrium state was discovered, we used the continuation method to discover the one-dimensional manifold of equilibrum points.

We iterated the process with every attractor discovered when $\gamma = 200$. When γ is very large, the equilibrium points are mostly determined by Quantization Harmony. The function was designed to have the attractors at a subset of the vertices of the unit hypercube $[0,1]^{27}$. The actual attractors must be close to those vertices. By applying the Newton's method, we could find all the attractors when $\gamma = 200$. Besides the attractors, there are many unstable equilibrium points at the γ value and finding those points is much more difficult. Instead of trying to find those unstable fixed points at the paramter setting, we used the continuation method to discover the unstable equilibrium points, assuming that the unstable equilibrium points are connected to the stable equilibrium points on the manifold of the equilibrium points when γ is allowed to

[6]Given that all four sentences of the language are symmetric, we focus on the case in which the model is reading a sentence 'A B' and investigate when and how the model can build the target structure $[_{S1}$ A B] successfully.

change. If we start to follow the manifold from a stable equilibrium point, we will reach an unstable equilibrium point.

The numerical bifurcation analysis with the continuation method was performed by using a MATLAB package, the Continuation Core and Toolboxes (COCO)[7]. After performing bifurcation analysis, we chose specific γ values (γ = 20, 25, 35, 70) and constructed a disconnectivity graph at each γ value to investigate the topology of the energy landscape.

We did not perform bifurcation analysis and construct the disconnectivity graphs when the model reads the second word 'B' ($\alpha = 1$). Instead, the equilibrium points and the topology of the harmony landscape were inferred, based on symmetry, from those discovered from numerical bifurcation analysis when the model reads the first word 'A'.

Result

The continuation method discovered 795 branches of the equilibrium points including 1589 equilibrium points at 200 of γ when $\alpha = 0$.[8] Most branches are not relevant in our discussion of the model dynamics so we focus on a small number of important branches (see Figure 2). First, look at the branches (1), (2), and (5). A saddle-node bifurcation occurs at $\gamma_1 \approx 21.3$ and a subcritical pitchfork bifurcation occurs at $\gamma_2 \approx 59.6$ (for introduction, see Strogatz, 1994). Second, look at the branches (3) and (4) which evolve to the states representing $[_{\text{S3}}$ D B] and $[_{\text{S4}}$ D C] both of which are inconsistent with the first word 'A'. Those branches are disconnected from the other branches. We suspect the branches evolving to the states representing grammatical/ungrammatical structures inconsistent with the input word are disconnected from the major branch evolving to the states representing the grammatical structures consistent with the input.

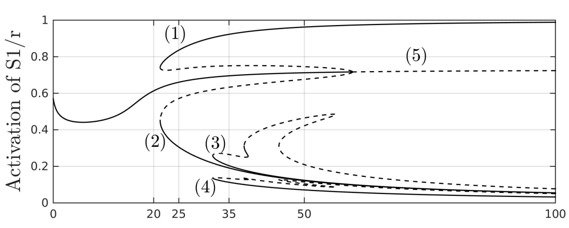

Figure 2: The bifurcation digram. The solid and dashed lines indicate stable and unstable equilibria, respectively. Only the branches evolving to (1) $[_{\text{S1}}$ A B], (2) $[_{\text{S2}}$ A C], (3) $[_{\text{S3}}$ D B], (4) $[_{\text{S4}}$ D C], and (5) a blend of S1 and S2 are presented. The vertical lines indicate the γ values at each of which a disconnectivity graph was constructued (see Figure 3).

[7] http://sourceforge.net/projects/cocotools/

[8] The COCO could not finish continuation successfully along 96 of 891 branches. All those cases occurred when the software switched to a new branch at a branch point (where multiple branches intersect each other) and tried to follow a differnet branch; the branch before switching was connected to an ungrammatical structure (e.g., $[_{\text{C}}$ D _]) at 200 of γ. Those cases need further investigation but those branches are not important in the discussion of the model.

From Figure 2, we can predict the state change while the model is reading the first word 'A'. Recall that γ is assumed to increase in time. The system has a single global attractor in the region where $\gamma < \gamma_1$. Thus, regardless of the initial state, the system will reach the global equilibrium point before γ passes γ_1. At γ_1, two more attractors emerge from a saddle-node bifurcations but they are separated from the equilibrium point along branch (5) where the system is. The system keeps following the major branch until γ passes γ_2 at which a subcritical pitchfork bifurcation occurs and the equilibrium point loses its stability. Thereafter, even with a very small noise, the system moves to either branch (1) or (2).

Figure 3 shows the topology of the energy landscape at four different γ values (see also Figure 2). When $\gamma = 20$ (Figure 3a), there is a single attractor. When $\gamma = 25$ (Figure 3b), there are three local optima on branches (1), (2), and (5). If the system has followed the major branch, the system will be at the attractor on branch (5) represented by the leftmost terminal node. The attractor is separated from other attractors on branches (1) and (2) by energy barriers. When $\gamma = 35$ (Figure 3c), the energy landscape has more attractors but all newly emerged attractors are separated from the three attractors on branches (1), (2), and (5) by high energy barriers. The bottommost part of the graph inside the square has the same local structure as Figure 3b. The blend state on branch (5) is still stable so the system will be at the equilibrium point. Figure 3d presents the disconnectivity graph when $\gamma = 70$. The graph is very complex and not easy to read. The squred region at the bottom is magnified in the third panel of the top row in Figure 4. At that time, the equilibrium point on branch (5) is not stable any more. The system moves to the branches of stable equilibrium points, either (1) or (2).

Recall that each word in the language is consistent with two grammatical structures and inconsistent with the other two grammatical structures. Given the symmetry, the equilibrium states when the model reads the second word 'B' ($\alpha = 1$) can be inferred from the equilibrium states when the model reads the first word 'A'. More specifically, when the model reads the second word 'B', the system forms a different set of branches of equilibrium points which has the same structure as shown in Figure 2. At this time, the equilibrium points along branch (2) evolves into the state representing $[_{\text{S3}}$ D B] and the equilibrium points along branch (5) evolves into the blend of the blend of $[_{\text{S1}}$ A B] and $[_{\text{S3}}$ D B]. Equilibrium points on branch (1) still evolves into the state representing $[_{\text{S1}}$ A B].

Now consider what will happen if the model reads the second word 'B' at γ_c. Figure 4 shows three qualitatively different cases in incremantal processing.

Case 1 (see path [1] in Figure 4): The model reads 'B' when $\gamma_1 < \gamma_c < \gamma_2$; γ_c = 35 in the current example. While γ increases from 0 to γ_c, the state quickly approaches a global attractor and follows the major branch (see branch (5) in Figure 2). The state change is indicated by the arrow (annotated as [1]/[2]) from panel (a) to panel (b) in Figure 4. When the model reads 'B' at γ_c = 35, the harmony landscape changes

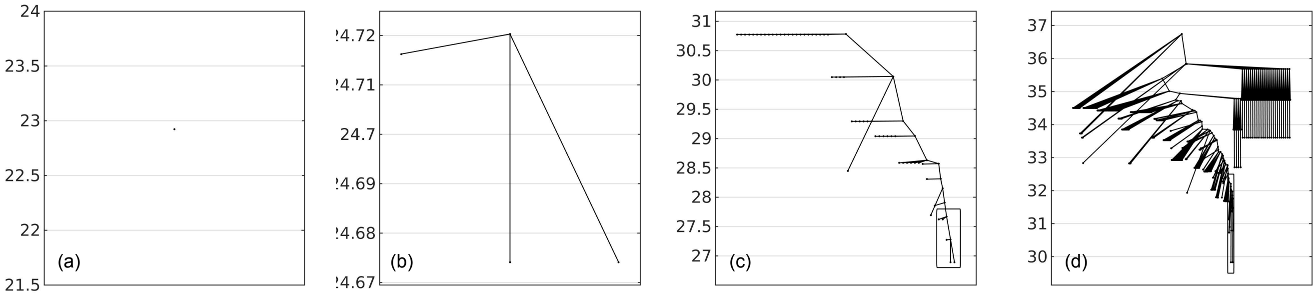

Figure 3: Topological change in the harmony landscape (given a first word 'A') as γ increases ($\gamma = 20, 25, 35, 70$)

abruptly. The previously stable equlibrium state becomes unstable so the state moves to a local optimum when $\alpha = 1$. It turns out the previous equilibrium point is located inside the basin of attraction of an equilibrium point along a branch corresponding branch (1) in Figure 2. The state change is indicated by the arrow (annotated as [1]) from panel (b) to panel (e). As γ increases further, the new equilibrium point evolves to the state representing the target structure [$_{S1}$ A B]; see the arrow (annotated as [1]) from panel (e) to panel (f). When small noise is assumed, the model builds the target structure with certainty.

Case 2 (garden path; see path [2] in Figure 4): The model reads 'B' when $\gamma_c > \gamma_2$; $\gamma_c = 70$ in the example. As in Case 1, the system reaches the global attractor and follows the manifold of the global attractors in the beginning. However, when γ passes $\gamma_2 (\approx 59.6)$, the equilibrium point on branch (5) loses its stability. With a small noise, the system moves to a new equilibrium point, either one on branch (1) or one on branch (2); see the arrows from panel (b) to panel (c). As suggested in Figure 3, the system cannot move to other attractors because they are separted by high energy barriers. When the model reads the second word at $\gamma_c = 70$, the harmony landscape changes abruptly. The attractors (1) and (2) in the top right panel are located in the basins of attraction of the new attractors (1) and (2) in the bottom right panel, respectively. The state will be at either (1) or (2) in the bottom right panel depending on where the system was (see the arrows from panel (c) to panel (f). It suggests if the system was following branch (2) while processing the first word, the system will build a non-target structure [$_{S2}$ A C] with the second word input, although it is inconsistent with the bottom-up input. This case is like the garden path effect in that the system chose one possibility over the others when it encounters ambiguity.

Case 3 (local coherence; see path [3] in Figure 4): The model reads 'B' when $\gamma_c < \gamma_1$; $\gamma_c = 20$ in the example. As in Case 1, the system reaches the global attractor and follows the manifold of the global attractors in the beginning. When the model reads 'B' at $\gamma_c = 20$, the state moves to a new global optimum; see the arrow [3] from panel (a) to panel (d). As γ increases, the system follows the major branch of equilibrium points (see the arrow from panel (d) to panel (e)) until γ passes γ_2 when the system moves to either a branch leading to the target structure or a branch leading to a non-target structure [$_{S3}$ D B] (see the arrows [3] from panel (e) to panel (f)). The latter case is an extreme version of the local coherence effect; the system failed to use top-down information (context provided by the first word 'A') and relied only on the bottom-up information to choose the locally coherent but globally incoherent structure.

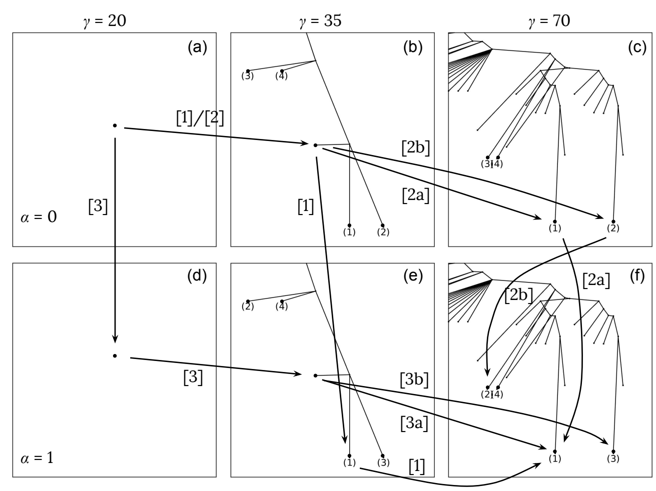

Figure 4: Parsing accuracy depends on γ_c. The arrows indicate the state change expected while the model processes a sentence 'A B' ([1] $\gamma_c = 35$, [2] $\gamma_c = 70$, [3] $\gamma_c = 20$). The panels in the second and third columns correspond to the regions included in the panels (c) and (d) in Figure 3.

Conclusion

In this study, we investigated how the GSC model handles temporary ambiguity to parse a sentence successfully in incremental processing. The model considers all structures consistent with context (the words that the model has processed) without choosing one over the others by being at a *stable blend* state which can evolve into the possible grammatical structures, suggesting the usefulness of the continuous representation space. We point out that the blend state does not represent a disjunctive set of those structures so the GSC model is not an implementation of the probabilistic models. On the other hand, the model rejects the structures inconsistent with the input by developing high energy barriers between attractors.

More importantly, the model explains when and why the model fails to parse a sentence correctly. When γ increases too quickly, the model chooses one interpretation over the other possible ones. When γ increases too slowly, the model fails to separate globally incoherent interpretations from globally coherent ones. The present study suggests that accurate parsing in the GSC model requires an optimal control of quantization strength γ.

Acknowledgments

We thank Geraldine Legendre, Akira Omaki, Kyle Rawlins, Ben Van Durme, Colin Wilson, Laurel Brehm, Nick Becker, Belinda Adam and especially Matthew Goldrick for their contributions to this work, and acknowledge the support of NSF INSPIRE grant BCS-1344269.

References

Becker, O. M., & Karplus, M. (1997). The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *The Journal of Chemical Physics*, *106*(4), 1495–1517.

Chiang, T., Hwang, C., & Sheu, S. (1987). Diffusion for global optimization in $\mathbb{R}^n$. *SIAM Journal on Control and Optimization*, *25*(3), 737–753.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Frank, S. L. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 1139–1144).

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* (pp. 559–586). Lawrence Erlbaum Associates.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies* (pp. 1–8). Stroudsburg, PA, USA: Association for Computational Linguistics.

Hale, J., & Smolensky, P. (2006). Harmonic Grammars and harmonic parsers for formal languages. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind: From neural computation to optimality-theoretic grammar. Volume I: Cognitive architecture* (pp. 393–416). The MIT Press.

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

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.

Meijer, H., Dercole, F., & Oldeman, B. (2009). Numerical bifurcation analysis. In R. A. M. Ph.D (Ed.), *Encyclopedia of Complexity and Systems Science* (pp. 6329–6352). Springer New York. (DOI: 10.1007/978-0-387-30440-3_373)

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, *275*(5306), 1604–1610.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*(1), 159–216.

Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: a framework for integrating the continuous and the discrete in cognition. *Cognitive Science*, *38*(6), 1102–1138.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1521–1543.

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

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

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 431–450.

Traxler, M. J., Pickering, M. J., & Clifton Jr., C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*(4), 558–592.

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

Wales, D. J., Miller, M. A., & Walsh, T. R. (1998). Archetypal energy landscapes. *Nature*, *394*(6695), 758–760.