Energy Minimization and Directionality in Phonological Theories

https://escholarship.org/uc/item/0pt518jn

Proceedings of the Annual Meeting of the Cognitive Science Society, 14(0)

Touretzky, David S.
Wang, Xuemei

1992

Energy Minimization and Directionality in Phonological Theories

David S. Touretzky and **Xuemei Wang**
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Goldsmith (1990, 1991) and Lakoff (in press) have both proposed phonological theories involving parallel constraint satisfaction, and making explicit reference to Smolensky's (1986) harmony theory. We show here that the most straightforward implementation of phonological constraint satisfaction models as spin glasses does not work, due to the need for directionality in constraints. Imposing directionality negates some of the advantages hoped for from such a model. We have developed a neural network that implements a subset of the operations in the Goldsmith and Lakoff phonological theories, but proper behavior requires asymmetric connections and essentially feed-forward processing. After describing the architecture of this network we will move on to the issue of whether spin glass models are really an appropriate metaphor for phonological systems.

Introduction

Goldsmith (1990, 1991) and Lakoff (in press) have both proposed phonological theories involving parallel constraint satisfaction, and making explicit reference to Smolensky's (1986) harmony theory. These proposals have been criticized by Touretzky and Wheeler (Touretzky, 1989; Touretzky & Wheeler, 1990a; Wheeler & Touretzky, in press) as computationally infeasible. They offer an alternative theory using deterministic rules, with no reapplication, that is implementable by purely feed-forward circuitry (Touretzky & Wheeler, 1991). Goldsmith (personal communication) has in turn criticized this theory on the grounds that it may not be powerful enough to handle the complexity of rule interactions found in some human languages. An answer to this challenge awaits further work by Wheeler and Touretzky.

Smolensky's harmony theory is equivalent to the Boltzmann machine model of Hinton & Sejnowski (1986), and thus falls into the category of spin glass energy minimization models (Hopfield, 1982). Here we wish to address the original criticism directed at the constraint satisfaction proposal: that it is not obvious how such models could be implemented in a spin glass architecture, and even if they could, the computation required to find a maximum harmony/minimum energy state would be excessive.

In this paper we show that the most straightforward implementation of phonological constraint satisfaction models as spin glasses does not work, due to the need for directionality in constraints. Imposing directionality negates some of the advantages hoped for from such a model. We have developed a neural network that implements a subset of the operations in the Goldsmith and Lakoff phonological theories, but proper behavior requires asymmetric connections and essentially feed-forward processing. After describing the architecture of this network we will move on to the issue of whether spin glass models are really an appropriate metaphor for phonological systems.

The Goldsmith and Lakoff Models

Goldsmith and Lakoff utilize similar three-level models, where the levels are labeled M (underlying, or morphophonemic), W (word), and P (surface, or phonetic).[1] See Figure 1. The initial underlying form of a word is denoted M_1. Free reapplication of unordered intra-level (M,M) rules produces a new string, M_n. Interaction-free, parallel application of inter-level (M,W) rules then produces the initial W-level representation, W_1. The process continues with free reapplication of unordered (W,W) rules, parallel application of (W,P) rules, and finally, free reapplication of unordered (P,P) rules.

A sample derivation in this formalism, taken from Lakoff (in press), is shown below. The data comes from Yawelmani, an American Indian dialect from California (Kenstowicz & Kisseberth, 1979). Yawelmani has an epenthesis process that inserts an /i/ after the first member of a triconsonantal cluster. It also has a vowel harmony process[2] in which round vowels cause all contiguous

---

[1] In the original version of his paper Lakoff called these levels M, P, and F, and Touretzky and Wheeler used these labels in their earlier papers. Subsequently, they and Lakoff adopted Goldsmith's notation.

[2] Not to be confused with Smolensky's *harmony theory*, an unrelated use of the word.

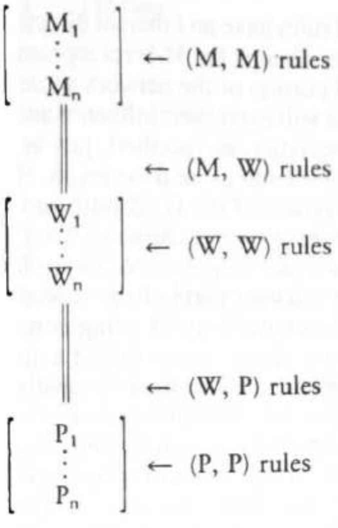

Figure 1: Structure of Goldsmith's model, taken from (Goldsmith, 1990:324).

succeeding vowels of the same height to become round, back, and non-low. The harmony rule can be stated as a (W, W) constraint as follows:

W: If [+syl,+rd,αhigh] C_0 X,
then if X = [+syl,αhigh],
then X = [+rd,+back,−low]

In /du:ll+hin/ "climbs," the (M, W) epenthesis process inserts an /i/ after the first /l/. This epenthetic vowel is not only subject to the (W, W) harmony process, in rounding it becomes the trigger for harmony applying again, to the vowel in /hin/. After an independent lowering process at (W, P), the surface form is [do:lul-hun].

M: d u: l l - h i n
W: d u: l u l h u n
P: d o: l u l h u n

Goldsmith and Lakoff see constraints as maximizing the linguistic "harmony" or well-formedness of a string. Rules act as "repair strategies" triggered by constraint violations (Sommerstein, 1974); they alter the string in language-specific ways so as to best satisfy all applicable constraints. Spin glass models also perform constraint satisfaction, and thus, spin glasses are an appealing metaphor for phonological processing. But this analogy is not without problems.

Criticisms of the Constraint Satisfaction Model

Lakoff, in introducing his theory of "cognitive phonology," argues that classical linguistic derivations are too lengthy to be taken literally as mental theories of phonological processing. However, allowing free reapplication of unordered intra-level rules leads to the same sort of lengthy derivations, at least in Goldsmith's version of the theory, where each representational level goes through a sequence of forms (e.g., $M_1, \ldots, M_n$ in Figure 1) due to serial effects of constraints. Goldsmith makes an analogy to traveling downhill in constraint violation space. (He also suggests in (Goldsmith, 1991) that an actual implementation would use units with graded activations that settle into a minimum energy state in a continuous dynamical system sort of way, which seems at odds with the notion of well-defined symbolic intermediate states.)

Our interpretation of Lakoff is that he takes a nondeterministic approach to derivations, such that there is only one form at each level rather than a sequence of forms. The two requirements of, say, a W-level form are that it must satisfy any applicable constraints at that level, and that its points of difference from M-level must all be "licensed" by some combination of (M, W) rules and (W, W) constraints. For example, rather than positing an intermediate W_1 form /du:lil+hin/ resulting from epenthesis in the Yawelmani derivation of [do:lulhun], Lakoff goes directly to the final W-level form /du:lul+hun/, which is licensed by a combination of (M, W) epenthesis, (W, W) harmony applying to the epenthetic vowel, and (W, W) harmony reapplying to the vowel of the suffix /hin/. This abstract notion of licensing of derived forms avoids adding sequentiality to the theory, but at a cost: deducing the surface form from the underlying form of a string requires either nondeterminism (not available in actual physical computing devices) or search.

Structure of Our Model

A straightforward way to encode constraints in a connectionist net is as propositional implications. Pinkas (1991) showed that any sentence in propositional logic can be encoded directly in a Hopfield net with n-ary (higher order) connections, such that the global energy minima of the network correspond exactly to those truth assignments that satisfy the sentence. For example, the implication **abc** → **d** would be encoded as shown in Figure 2 by a third-order connection among **a, b,** and **c** with weight -1, plus a fourth-order connection among **a, b, c,** and **d** with weight $+1$. All units have 0/1 states and zero thresholds. The minimum energy states of this network are those where either at least one of **a, b,** or **c** is in state 0, or all four nodes are in state 1. If **a, b,** and **c** are clamped on, then after settling, **d** will be on. If **a** is clamped on and **d** is clamped off, then after settling, at least one of **b** and **c** will be off.

Figure 3 shows the structure of a portion of our model. A segment (phoneme), depicted as a rectangle in the figure, is represented as a binary feature vector. By default, each M-level segment is mapped to an identical W-level segment; this is accomplished by weak one-to-one connections between corresponding elements. (M, W) rules may alter this mapping in one of three ways: they can alter individual features of a segment, they can cause the segment to be deleted by turning on its deletion bit at W-level, and they can cause a new segment to be inserted to the right of the current segment. The last is achieved

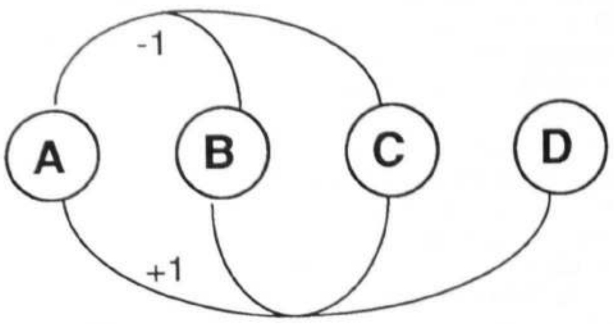

Figure 2: Encoding of **abc** → **d** in a Hopfield net with higher-order connections.

by turning on the segment's insertion bit and writing the binary feature pattern into the associated insertion slot, which appears below and to the right of the segment in the figure.

In order to permit these changes to affect the environment of subsequent rule applications, each segment is augmented with dynamically computed left and right neighbor slots, shown as diamonds in the figure. Let Del_i and Ins_i stand for the deletion and right insert bits associated with segment i. Then the left neighbor of segment i is computed as shown in Figure 4.

A similar computation is used to find the nearest vowels to the left and right of a segment (required for harmony rules, which operate on the "vowel tier".) Insert slots also have left and right neighbor and left and right vowel slots.

This rather ad hoc solution places some limits on multiple insertion and deletion processes. It is not possible to insert more than one segment at W-level between segments that were adjacent at M-level, since there is only a single insertion slot in each position. In order for left and right neighbor computations to work correctly. it is forbidden to delete adjacent segments, though this restriction could be relaxed by increasing the complexity of the neighbor computations. Finally, segments that have been inserted by (M, W) rules can be modified by (W, W) rules, but not deleted, since the insertion slot does not have its own deletion bit.

After the W-level has settled, the string is copied into a fresh buffer, removing any segments marked for deletion and making room for new segments where an insertion bit was turned on. Finally, the mechanism for mapping strings from W to P level is employed; it is identical to the M-to-W mechanism just described.

Rules in the model are replicated at each buffer position, as in (Touretzky, 1989). Intra-level rules can refer to segments, their left and right neighbors, and the nearest vowel to the left and right. Inter-level rules, since they cannot reapply, do not need to consider insertion and deletion bits; they therefore can refer to more distant neighbors with no increase in circuit complexity.

Directionality

Cross-level rules have an inherent directionality. For example, if one clamps the M-level representation and lets the W-level portion of the network settle, the applicable (M, W) rules will exert their influence and modify the W-level representation as specified, just as clamping **a, b,** and **c** in Figure 2 will cause **d** to turn on. However, a great deal of the power of the Goldsmith and Lakoff models comes from *intra*-level constraints. The (W, W) harmony constraint is a case in point. With intra-level constraints none of the relevant parts of the representation can be clamped; the entire W-level string is moving downhill in constraint violation space to find a more harmonious state. The problem is that there is usually more than one way to resolve a constraint violation. Constraints stated as implications have an implied directionality of inference, but the straightforward encoding of constraints as symmetric higher order connections does not retain this directionality. Specifically, if nodes **a, b, c,** of Figure 2 happen to be on while node **d** is off, the network can resolve the constraint violation equally well by turning **d** on or by turning one or more of **a, b,** or **c** off.

When we implemented the Yawelmani (W, W) harmony rule as a logical implication in a higher-order Hopfield net version of our model, we found that the lack of directionality could indeed produce this undesirable behavior.

It is of course possible to replace a nondeterministic search for satisfying truth assignments for propositional sentences with the deterministic application of directed inference rules. We did this by switching to a model with asymmetric connections, so that rules were implemented by essentially feed-forward circuitry. The resulting model succesfully performed the Yawelmani derivations in Lakoff's paper, including reapplication of the vowel harmony rule in the example cited above, and with a slight modification, also replicated Lakoff's example of a Mohawk derivation that involves six rule applications.

Reconsidering Constraint Satisfaction

What does the asymmetric model just described really say about constraint satisfaction? It is, after all, merely applying rules in a deterministic, feed-forward fashion, though it employs some clever tricks to make this happen in a connectionist framework. We have several observations to make about this.

First, many phonological processes appear quite compatible with a deterministic model, and do not require true constraint satisfaction. We would not claim that all of phonology fits this mold; Goldsmith and others are emphatic that it does not. But even in areas like syllabification, where one can easily imagine multiple constraints competing to mark a consonant as an onset or a coda, Touretzky & Wheeler (1990b) have shown that there are perfectly adequate parallel but deterministic accounts. So we remain skeptical about the need for constraint satisfaction as a computational mechanism.

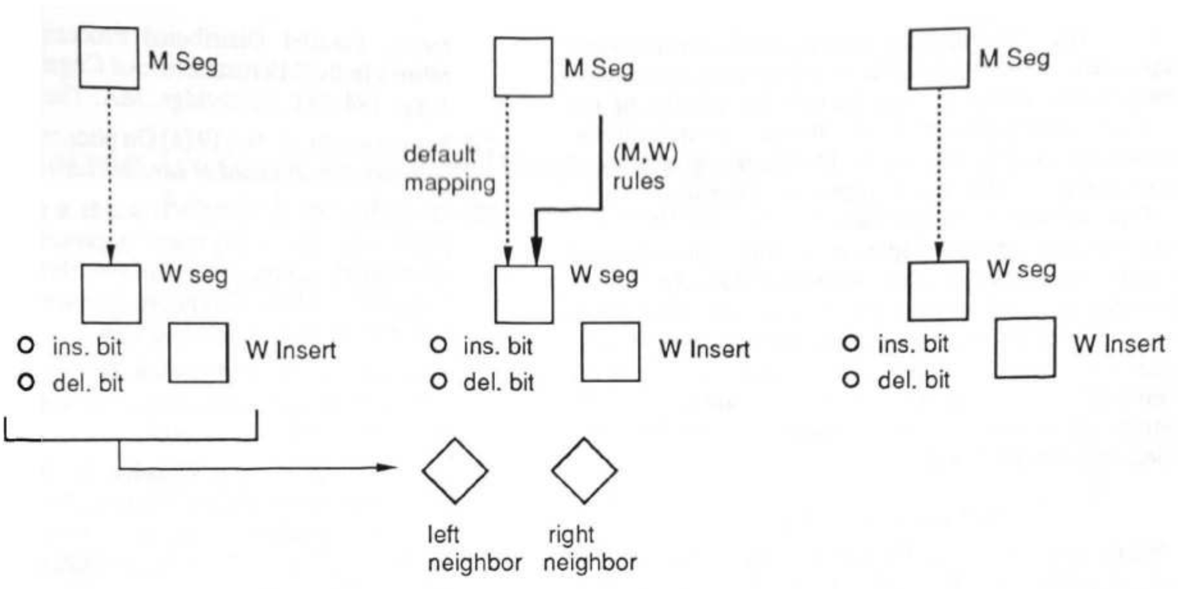

Figure 3: A portion of the circuitry of our phonological model.

$$\text{Left neighbor of } i = \begin{cases} \text{insert slot } i-1 & \text{if } \text{Ins}_{i-1} = 1 \\ \text{segment } i-1 & \text{if } \text{Ins}_{i-1} = 0 \text{ and } \text{Del}_{i-1} = 0 \\ \text{insert slot } i-2 & \text{if } \text{Ins}_{i-1} = 0 \text{ and } \text{Del}_{i-1} = 1 \text{ and } \text{Ins}_{i-2} = 0 \\ \text{segment } i-2 & \text{if } \text{Ins}_{i-1} = 0 \text{ and } \text{Del}_{i-1} = 1 \text{ and } \text{Ins}_{i-2} = 0 \end{cases}$$

Figure 4: Formula for the left neighbor of a segment.

Further investigation is required on this point.

A second point, though, is that there is a tension between the computational flavor of spin glass models (symmetric connections, nondeterministic behavior) and the directed way phonological systems resolve constraint violations. This brings us back to the traditional separation of constraints from repair strategies. The linguistic motivation for this has been parsimony: the same repair strategy may be used to fix several constraint violations, but it should only have to be stated once.[3] The computational motivation for such a separation has received less attention, perhaps because phonological theories are rarely implemented. But it now seems that if all segments are potentially subject to modification (as is the case with intra-level constraints), and rules can freely reapply (so clamping of rule antecedents is not possible), then there can be no straightforward mapping of linguistic contraints into an energy minimization model due to the loss of directionality.

Compare this situation with Touretzky & Hinton's (1988) distributed connectionist production system (DCPS), a Boltzmann machine that used simulated annealing to match production rules against working memory. Rules could reapply in DCPS, but there were two crucial differences from the phonological model. Only one rule could apply at a time, and after each rule application the state of working memory was latched to provide a steady input for the next rule match. One could of course do the same thing in the phonological model, which would bring us back once again to sequential rule application. But it is the potential for simultaneous application and interaction of multiple rules that makes the constraint satisfaction model interesting.

Conclusions

We have shown how to implement a limited version of the Goldsmith and Lakoff proposals using connectionist style hardware with asymmetric connections. However, the real contribution of this paper is a negative result, namely, that when you look closely at the details, energy minimization in spin glass models is not as apt a metaphor for linguistic constraint satisfaction as previously thought. In phonology, directionality is important.

This not to say that constraint satisfaction does not exist in phonology, but we must decide which features of the proposal are most important to preserve. If we want constraints to direct the application of repair strategies, then we may have to replace the notion of derivation as massively parallel settling into enery minima denoting well-formed strings with a more sequential form of derivation by succesive repairs. This would appear to undermine Lakoff's (and our) goal of a model fast enough to be cognitively pausible.

Alternatively, we could give up the linguist's goal of

---

[3] Sommerstein (1974), cited in Goldsmith (1990:321) shows five phonotactic constraints in Latin that share a repair strategy.

parsimony and abandon the separation of constraints and repair strategies. Constraints would become much more complicated structures that include knowledge of not just individual repair strategies, but also desired rule interactions. In this way we could maintain efficiency of derivations – at the cost of more complex rules.[4]

Can the spin glass approach be salvaged? It is conceivable that a more complex encoding of phonological constraints into a spin glass model could preserve directionality as an additional type of constraint. This would seem to require that points of application of a particular constraint be explicitly represented as part of the network's final state, resulting in a derivation that was annotated with licensing information.[5] We think this idea deserves further study.

Acknowledgements

This work was supported in part by funding from Hughes Aircraft Company. We thank Deirdre Wheeler for valuable discussions.

References

[1] Goldsmith, J. A. (1990) Autosegmental and Metrical Phonology. Oxford, UK: Basil Blackwell.

[2] Goldsmith, J. A. (1991) Phonology as an intelligent system. In D. J. Napoli and J. A. Kegl (eds.), *Bridges between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman*. Hillsdale, NJ: Erlbaum.

[3] Hinton, G. E., and Sejnowski, T. J. (1986) Learning and relearning in Boltzmann machines. In D. E. Rumelhart and J. L. McClelland (eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, pp. 282-317. Cambridge, MA: The MIT Press.

[4] Hopfield, J. J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, 81, 6871-6874.

[5] Kenstowicz, M., and Kisseberth, C. (1979) Generative Phonology: Description and Theory. San Diego, CA: Academic Press.

[6] Lakoff, G. (to appear) Cognitive Phonology. In J. Goldsmith (ed.), *The Last Phonological Rule: Reflections on Constraints and Derivations in Phonology*. Chicago: University of Chicago Press.

[7] Pinkas, G. (1991) Symmetric neural networks and propositional logic satisfiability. Neural Computation 3(2):282-291

[8] Smolensky, P. (1986) "Information processing in dynamical systems: foundations of harmony theory." In D. E. Rumelhart and J. L. McClelland (eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, volume 1, pp. 194-281. Cambridge, MA: The MIT Press.

[9] Sommerstein, A. H. (1974) On phonotactically motivated rules. *Journal of Linguistics* 10: 71-94.

[10] Touretzky, D. S. (1989) Towards a connectionist phonology: the "many maps" approach to sequence manipulation. Proceedings of the Eleventh Annual Conference of the Cognitive Science Society, pp. 188-195. Hillsdale, NJ: Erlbaum.

[11] Touretzky, D. S., and Hinton, G. E. (1988) A distributed connectionist production system. *Cognitive Science* 12:(3):423-466.

[12] Touretzky, D. S., and Wheeler, D. W. (1990a) A computational basis for phonology. In D. S. Touretzky (ed.), *Advances in Neural Information Processing Systems* 2, pp. 372-379. San Mateo, California: Morgan Kaufmann.

[13] Touretzky, D. S., and Wheeler, D. W. (1990b) Two derivations suffice: the role of syllabification in cognitive phonology. In C. Tenny (ed.), The MIT Parsing Volume, 1989-1990, pp. 21-35. MIT Center for Cognitive Science, Parsing Project Working Papers 3.

[14] Touretzky, D. S., and Wheeler, D. W. (1991) Sequence manipulation using parallel mapping networks. *Neural Computation* 3, 98-109.

[15] Wheeler, D. W., and Touretzky, D. S. (to appear) A connectionist implementation of cognitive phonology. In J. Goldsmith (ed.), *The Last Phonological Rule: Reflections on Constraints and Derivations in Phonology*. Chicago: University of Chicago Press. An ealier version is available as technical report CMU-CS-89-144, Carnegie Mellon University School of Computer Science, 1989.

---

[4] Chunking is a well-known technique for incrementally constructing efficient systems with complex rules, by observing the interactions of simple rules.

[5] Such a derivation would resemble a constructive proof.