Static Harmonic Grammar
Constraint conflict without candidate comparison

Colin Wilson
Johns Hopkins University

UBC WCCFL 38 March 7, 2020

Outline

- **Components of HG/OT**
  Global optimization is one way of resolving *constraint conflict*,
  with both advantages and disadvantages
- **Static Harmonic Grammar**
  SHG is an alternative formalization in which constraint conflict plays out
  within individual *nodes* of a representation rather than globally
  Two constraints conflict within a node if they assign it marks of opposite polarity
- **Case studies**
  Local Markedness/Faithfulness interactions, unbounded spreading,
  stress assignment with metrical grids or feet, basic syllable structure
- **Logical structure**
  Constraints in SHG can be stated as logical formulas with one free variable,
  grammars are equivalent to disjunctive normal forms in the same logic
- Further properties of and open issues in SHG

Components of Harmonic Grammar / Optimality Theory

Gen
Defines a set of universally possible **representations**, candidates for a given input
(e.g., universal feature set, syllabic constituents, foot inventory, Strict Layer Hypothesis)

Con
A set of **constraints** (originally assumed universal) that assign violations *to entire candidates*

Ex.	*I/Q	*Assign one violation to a candidate* for each [+high] vowel adjacent to a uvular
	*E	*Assign one violation to a candidate* for each [−high,−low] vowel
	Id[high]	*Assign one violation to a candidate* for each [α high] output segment that has a [$-\alpha$ high] input correspondent (where feature value $\alpha \in \{+,-\}$)

Eval
Given a weighting / ranking of the constraints, identifies the **most harmonic** or **optimal** candidate(s) for a given input (i.e., the top member(s) of the **harmonic ordering** of forms)

Ex.	/piqaj/ 'to grind'	[peqaj] $\succ$ [piqaj] $\succ \cdots$	w*I/Q $>$ w*E, wId[high]
	/misi/ 'cat'	[misi] $\succ$ [mesi], [mise] $\succ$ [mese] $\succ \cdots$	w*E $>$ wId[high]

Examples from South Bolivian Quechua (Bills et al. 1971. Laime Ajacopa 2007, Gallagher 2016)

Alternative constraint-based theories

Commitment to a set of universally-possible representations, made explicit by **Gen**, is central to generative phonology (e.g., feature geometry, autosegments, metrical structure; Goldsmith 1990) and many theories employ a univeral or language-specific set of constraints similar to **Con**

⋆ The distinctive property of HG/OT is that *global optimization over entire candidates*, as performed by **Eval**, determines the representations that are grammatical in a given language

Many alternative theories eschew optimization by assuming that constraints are *inviolable*

- **Licensing Theory** (e.g., Itô 1986, 1989, Goldsmith 1990, Lombardi 1994, Steriade 1995)
- **Declarative Phonology** (e.g. Bird et al. 1992, Sobbie, Coleman & Bird, 1996; Bird 1991, 1995, Scobbie 1991, 1993, Coleman 1995, Lodge 2003)
- **Constraints and Repair Strategies** (e.g., Paradis & LaCharité 1993, Paradis 1999)
- **Government Phonology** (e.g., Kaye, Lowenstamm & Vergnaud 1985, 1990, Charette 1990, 1991, Harris 1990, Kaye 1990)

See also (tier-based) strictly local grammars (e.g., Heinz, Rawal, & Tanner 2011, McMullin & Hansson 2016) and single-level maximum entropy models (e.g., Wilson & Gallagher 2018)

Advantages of global optimization

Optimization as in HG/OT allows many language-specific patterns and typological generalizations to be reduced to the interaction of a smaller number of *conflicting* constraints

- HG/OT factors each pattern into multiple pressures that, when weighted / ranked differently, yield a restricted typology of predicted patterns (e.g., Prince & Smolensky 1993/2004)
- HG/OT constraints are often phonetically grounded (e.g., Hayes, Kirchner, & Steriade 2004) and formally simple (e.g., Ellison 1994, Eisner 1997, Gerdemann & Van Noord 2000, Potts & Pullum 2002, McCarthy 2003, Riggle 2004)

The same type of factorization is not available to theories with inviolable constraints, in which the only mode of interaction is *conjunction* ('output must satisfy c_1 and c_2 and ...')

(examples on next slide)

Advantages of global optimization

Ex. Inviolable constraint analysis of Quechua vowel lowering (Wilson & Gallagher 2018)

- Surface-true constraints on high vowels in uvular contexts
  *Q I *I Q
- Surface-true constraints on mid vowels in exhaustive set of "nonuvular" contexts
  *# E K *K E K *V E K
  *# E # *K E # *V E #
  *# E V *K E V *V E V

Imagine Quechua' in which only a randomly selected subset of these constraints apply!
Compare HG analysis with violable constraints w*I/Q $>$ w*E $>$ wId[high]

Ex. Inviolable constraint analysis of "stress light syllable iff initial" (Hayes & Wilson 2008)

- *#[−stress]. Initial syllables must be stressed
- *#[ ][+stress,−heavy]. *Non-initial* light syllables must not be stressed

cf. HG analysis wStressInitial $>$ wStress-to-Weight (applies to all positions incl. initial!)

Disadvantages of global optimization

Optimization over entire candidates is a very powerful non-local, non-linear function

- Global optimization is computationally demanding because it depends on the entire grammar and candidate form rather than on each constraint and constituent individually
  - See Ellison 1994, Tesar 1994, 1995, 1996, Eisner 1997, and Wareham 1998 for foundational results (also Eisner 2000 on complexity of learning rankings)
  - Optimization is provably tractable *if* the set of constraints is fixed (universal) *and* all of the constraints can be 'preprocessed' into a single finite-state transducer (Riggle 2004, Heinz, Riggle, & Kobele 2009) — but the conditions under which preprocessing succeeds are uknown
- Even with simple constraints, global optimization can generate patterns that are beyond the currently known formal limits of phonology (e.g., non-rational transductions)
  - FLT results of Frank & Satta (1998), Gerdemann & Hulden (2012), Buccola (2013), Hao (2019), Koser & Jardine (2020)

Can we reap the advantages of optimization
without sowing these complexity and expressivity problems?

Outline

- **Components of HG/OT**
- **Static Harmonic Grammar**
  SHG is an alternative formalization in which constraint conflict plays out within individual *nodes* of a representation rather than globally
    Two constraints conflict within a node if they assign it marks of opposite polarity
- **Case studies**
- **Logical structure**
- Further properties of and open issues in SHG

Static Harmonic Grammar

In SHG (/ʃaɹg/), violable constraints interact *within* invididual constituents (or 'nodes') of a single representation and global optimization is replaced by a non-linear threshold function

- **Representations**
  Each representation X consists of a finite set of **nodes** ($\{1, 2, \ldots, \}$) that are **labeled** (technically 'sorted') and bear various **relations** with one another — as permitted by **Gen**
  node labels (sorts): $+$F, $-$F, tone, segment, mora (μ), syllable (σ), metrical foot (Ft), ...
  relations: association, precedence, dominance, correspondence, ...

- **Constraints**
  A constraint c_k assigns a **signed unit mark or zero** to each node n of a representation X.
  Important! Each node n is evaluated in the context of the representation X, not in isolation

$$c_k(n; X) \in \{-1, 0, +1\}$$

- **Evaluation**
  The **harmony** of node n in X is calculated by summing its weighted marks and applying an upper threshold $\epsilon \leq 0$. The **harmony** of X is equal to the total harmony of its nodes.

$$H(n; X) = \min\left(\sum_{k=1}^{K} w_k c_k(n; X)\ ,\ \epsilon\right) \qquad H(X) = \sum_{n \in X} H(n; X)$$

Static Harmonic Grammar

Special case with threshold $\epsilon = 0$

- A constraint c_k assigns a **signed unit mark or zero** to each node n of a representation X

$$c_k(n;X) \in \{-1, 0, +1\}$$

- The **harmony** of a node n in X is the sum of its weighted marks or 0, whichever is lower

$$\begin{aligned} H(n;X) &= \min\left(\textstyle\sum_{k=1}^{K} w_k c_k(n;X)\ ,\ 0\right) \\ &= \min_0 \textstyle\sum_{k=1}^{K} w_k c_k(n;X) \end{aligned}$$

*Define a node n to be **well-formed** in X iff $H(n;X) = 0$ (maximum possible harmony)*

- The **harmony** of an entire representation X is the sum of the harmonies of its nodes

$$H(X) = \textstyle\sum_{n \in X} H(n;X)$$

*Define a representation X to be **well-formed** iff all of its nodes are, equivalently iff $H(X) = 0$ (maximum possible harmony)*

Outline

- **Components of HG/OT**
- **Static Harmonic Grammar**
- **Case studies**
  Local Markedness/Faithfulness interactions, unbounded spreading, stress assignment with metrical grids or feet, basic syllable structure
- **Logical structure**
- Further properties of and open issues in SHG

Local Markedness interactions

Quechua vowel lowering (single-level analysis)

*I/Q Assign $\begin{Bmatrix} +1 \\ -1 \end{Bmatrix}$ to the $\begin{Bmatrix} [-\text{high}] \\ [+\text{high}] \end{Bmatrix}$ feature of a vowel that is adjacent to a uvular

*E Assign $\begin{Bmatrix} +1 \\ -1 \end{Bmatrix}$ to the $\begin{Bmatrix} [+\text{high}] \\ [-\text{high}] \end{Bmatrix}$ feature of a vowel that is [$-$low]

Important! A constraint assigns 0 to all nodes except as specified (implicit 'elsewhere' case)

	Representations	v*I/Q	w*E	$H([+\text{high}]_2)$ for $v > w$, $\epsilon = 0$
	$[\text{p}_1\text{e}_2\text{q}_3\text{a}_4\text{j}_5]$ 'to grind'	$+1$ $[-\text{high}]_2$	-1 $[-\text{high}]_2$	$\min(v - w, 0) = 0$
*	$[\text{p}_1\text{i}_2\text{q}_3\text{a}_4\text{j}_5]$ 'to grind'	-1 $[+\text{high}]_2$	$+1$ $[+\text{high}]_2$	$\min(-v + w, 0) < 0$
	$[\text{m}_1\text{i}_2\text{s}_3\text{i}_4]$ 'cat'		$+1$ $[+\text{high}]_2$	$\min(w, 0) = 0$
*	$[\text{m}_1\text{e}_2\text{s}_3\text{i}_4]$ 'cat'		-1 $[-\text{high}]_2$	$\min(-w, 0) < 0$

Important! Each output form is evaluated independently, not as a candidate vying against others, and all other nodes in these forms have harmonies of exactly zero

Local Markedness/Faithfulness interactions

Quechua vowel lowering (two-level analysis)

*I/Q	Assign $-u(\alpha)$ to the [α high] feature of an output vowel that is adjacent to a uvular
*E	Assign $u(\alpha)$ to the [α high] feature of an output vowel that is [−low]
Id[high]	Assign $u(\alpha)$ to the [α high] feature of an output vowel that has an [α high] input correspondent

where $u(+) = +1$, $u(-) = -1$ (and possibly $u(0) = 0$)

Representations	v*I/Q	w*E	zId[high]	H([+high]$_2$) $\epsilon = 0$
/$\mathsf{p_1 i_2 q_3 a_4 j_5}$/ → [$\mathsf{p_1 e_2 q_3 a_4 j_5}$]	$+1$ [−high]$_2$	-1 [−high]$_2$	-1 [−high]$_2$	$\min_0(v - w - z)$
* /$\mathsf{p_1 i_2 q_3 a_4 j_5}$/ → [$\mathsf{p_1 i_2 q_3 a_4 j_5}$]	-1 [+high]$_2$	$+1$ [+high]$_2$	$+1$ [+high]$_2$	$\min_0(-v + w + z)$
/$\mathsf{m_1 e_2 s_3 i_4}$/ → [$\mathsf{m_1 i_2 s_3 i_4}$]		$+1$ [+high]$_2$	-1 [+high]$_2$	$\min_0(w - z)$
* /$\mathsf{m_1 e_2 s_3 i_4}$/ → [$\mathsf{m_1 e_2 s_3 i_4}$]		-1 [−high]$_2$	$+1$ [−high]$_2$	$\min_0(-w + z)$

As long as $v > (w + z)$ and $w > z$, the harmony of node [α high]$_2$ and thus of the entire map is exactly zero for the grammatical cases and below zero for the ungrammatical cases

Local Markedness/Faithfulness interactions

Nootka rounding and unrounding (Campbell 1973, Sapir & Swadesh 1978, McCarthy 2002, 2003)

*oK	Assign $u(\alpha)$ to the [α round] feature of a dorsal in the context [+syll,+round]/_
*K^w]$_\sigma$	Assign $-u(\alpha)$ to the [α round] feature of a syllable-final dorsal
*K^w	Assign $-u(\alpha)$ to the [α round] feature of a dorsal (*ComplexSeg)
Id[round]	Assign $u(\alpha)$ to the [α round] feature of a dorsal that has an [α round] input correspondent

Rounding is contrastive on dorsals	[ɬa:kʷiqnak] 'pitiful'	yId[round] $>$ z*K^w $H(n) = \min_0(y - z) = 0$
. . . except after round vowels	[ʔo.kʷiːɬ] 'making it'	w*oK $>$ yId[round] $+$ z*K^w $H(n) = \min_0(w - y - z) = 0$
. . . and except syllable-finally	[m'oːq] 'throwing off sparks'	v*K^w]$_\sigma$ $>$ w*oK $+$ yId[round] $H(n) = \min_0(v - w - y + z) = 0$

where node n is the [α round] feature of the first dorsal in each form

Unbounded spreading

Nonlocal phonological interactions can be accounted for in SHG with enriched representations, such tiers and associated tier-adjacency relations, and by *chaining of local interactions*

Ex. Johore Malay progressive nasal spreading (Onn 1980, Walker 1988)

Agree[+nasal]$_{LR}$	Assign $u(\alpha)$ to an [α nasal] feature in the context [+nasal] ___
*NasFric	Assign $-u(\alpha)$ to the [α nasal] feature of a fricative
*NasVoc	Assign $-u(\alpha)$ to the [α nasal] feature of a vocoid (vowel or glide)

where *NasFric and *NasVoc are members of a set of nasal-affinity constraints (e.g., Pulleyblank 1989, Walker 1998, Boersma 1999, Piggot 1992, Cohn 1993)

/p$_1$ə$_2$ŋ$_3$a$_4$w$_5$a$_6$s$_7$a$_8$n$_9$/ →	v*NasFric		w*Agree[+nasal]$_{LR}$		z*NasVoc	
‘supervision’	$+1$	-1	$+1$	-1	$+1$	-1
* [p$_1$ə$_2$ŋ$_3$a$_4$w$_5$a$_6$s$_7$a$_8$n$_9$]	[−n]$_7$			[−n]$_4$!	[−n]$_{2,4,5,6,8}$	
* [p$_1$ə$_2$ŋ$_3$ã$_4$w$_5$a$_6$s$_7$a$_8$n$_9$]	[−n]$_7$		[+n]$_4$	[−n]$_5$!	[−n]$_{2,5,6,8}$	[+n]$_4$
* [p$_1$ə$_2$ŋ$_3$ã$_4$w̃$_5$a$_6$s$_7$a$_8$n$_9$]	[−n]$_7$		[+n]$_{4,5}$	[−n]$_6$!	[−n]$_{2,6,8}$	[+n]$_{4,5}$
[p$_1$ə$_2$ŋ$_3$ã$_4$w̃$_5$ã$_6$s$_7$a$_8$n$_9$]	[−n]$_7$		[+n]$_{4,5,6}$	[−n]$_7$	[−n]$_{2,8}$	[+n]$_{4,5,6}$
* [p$_1$ə$_2$ŋ$_3$ã$_4$w̃$_5$ã$_6$s̃$_7$a$_8$n$_9$]		[+n]$_7$!	[+n]$_{4,5,6,7}$	[−n]$_8$	[−n]$_{2,8}$	[+n]$_{4,5,6}$

In-class exercise: Determine which nodes in each output have $\min_0$ harmonies of zero if $v > w > z$

Metrical grids and quantity-insensitive stress

Stress is also unbounded, because the stress level of the last (resp. first) syllable can depend on a rhythmic sequence initiated at the first (resp. last) syllable, and the chaining analysis is similar

Ex. Alternate left-to-right starting with stress & leftmost main stress (e.g., Maranungku)

Alternate$_{LR}$	Assign $-u(\alpha)u(\beta)$ to a level-1 grid mark or hole β in the context α __
StressInitial	Assign $u(\alpha)$ to the level-1 grid mark or hole α above the lefmost level-0 grid mark
MainInitial	Assign $u(\alpha)$ to the level-2 grid mark or hole α above the leftmost level-1 grid mark

where $u(\mathtt{x}) = +1$, $u(\mathtt{o}) = -1$ (and possibly $u(\mathtt{<o>}) = 0$)

level 2:	x	o	o	o	o	o	o
level 1:	x	o	x	o	x	o	x
level 0:	x	x	x	x	x	x	x

level 2:	x	o	o	o	o	o	o	o
level 1:	x	o	x	o	x	o	x	o
level 0:	x	x	x	x	x	x	x	x

wAlternate$_{LR}$, wStressInitial, wMainInitial $>$ wAlternate$_{RL}$, wStressFinal, wMainFinal

Metrical grids and quantity-insensitive stress

(see empirical typologies of Hayes 1995, Gordon 2002, Goedemans, Heinz & van der Hulst 2014)

Ex. Alternate right-to-left starting with stress & rightmost main stress (e.g., Urubu Kapor)

wAlternate$_{RL}$, wStressFinal, wMainFinal $>$ wAlternate$_{LR}$, wStressInitial, wMainInitial

Ex. Alternate right-to-left starting with stress *except* do not stress the final syllable *unless* it is the only syllable of the word & leftmost main stress (e.g., Pintupi)

NonFinality	Assign $-u(\alpha)$ to the level-1 grid element α above the rightmost level-0 grid mark
Culminativity$_1$	Assign $u(\alpha)$ to a level-1 grid element α above a solitary level-0 grid mark

wCulminativity$_1$ $>_{\text{(universal)}}$ wStressInitial, wMainInitial, wNonfinality $>$ wAlternate$_{LR}$

level 2:	x	o	o	o	o	o	o
level 1:	x	o	x	o	x	o	o
level 0:	x	x	x	x	x	x	x

level 2:	x
level 1:	x
level 0:	x

Metrical grids and quantity-insensitive stress

Sketch of implementation in Python with parented trees from NLTK (Bird et al., 2009)

```
def u(n): # Unitizer function for metrical grids
  return +1.0 if n.label() == 'x' else -1.0 if n.label() == 'o' else 0.0

def alternate_LR(t): # Alternate-LR constraint
    # In the context x __ assign +1 to o and -1 to x,
    #                 o __ assign +1 to x and -1 to o
    for beta in t.subtrees(): # Examine each grid mark or hole
        alpha = n1.left_sibling() # and its immediately preceding sibling
        if alpha is not None: # Initial elements are unconstrained
          beta.marks['alternate_LR'] = -u(alpha)*u(beta)

def H(n, w, epsilon=0.0): # Harmony of a node
  H_n = 0.0
  for c in n.marks:
    H_n += w[c] * n.marks[c]
  return H_n if H_n < epsilon else epsilon
```

Basic syllable structure

Onset	Assign $\left\{\begin{matrix}+1\\-1\end{matrix}\right\}$ to a $\left\{\begin{matrix}\text{Onset}\\\text{Nucleus}\end{matrix}\right\}$ node at the beginning of a syllable
NoCoda	Assign $\left\{\begin{matrix}+1\\-1\end{matrix}\right\}$ to a $\left\{\begin{matrix}\text{Nucleus}\\\text{Coda}\end{matrix}\right\}$ node at the end of a syllable
Interior-Coda	Assign $\left\{\begin{matrix}+1\\-1\end{matrix}\right\}$ to a Coda node that $\left\{\begin{matrix}\text{is}\\\text{is not}\end{matrix}\right\}$ followed by an Onset
Final-Coda	Assign $\left\{\begin{matrix}+1\\-1\end{matrix}\right\}$ to a Coda node that $\left\{\begin{matrix}\text{is}\\\text{is not}\end{matrix}\right\}$ at the end of a word

The dependency between a Coda and a following Onset imposed by Interior-Coda parallels one type of **transconstituent government** in Government Phonology

Variable weighting of these and Faith predicts that (i) Onsets are preferred in all languages, (ii) universally *VC.V, and (iii) Codas may be disallowed (Hawaiian) *or* allowed only internally (Diyari) *or* allowed only word-finally (Luo) *or* allowed in both positions (Manam)

Broselow 2003: "lack of universal isomorphism between word margins and syllable margins"

Outline

- **Components of HG/OT**
- **Static Harmonic Grammar**
- **Case studies**
- **Logical structure**
  Constraints in SHG can be stated as logical formulas with one free variable, grammars are equivalent to disjunctive normal forms in the same logic
- Further properties of and open issues in SHG

Logical structure

In logical terms, a phonological representation X is a **model** consisting of a finite domain D of nodes and a set of functions and relations on those nodes

Mark assignment by an SHG constraint c_k can be formalized logically by three open formulas with a single unbound variable x that stands in for the receiving node

Assign $+1$ *to node* n *in* X iff $X, n \vDash \phi_k^+(x)$

Assign -1 *to node* n *in* X iff $X, n \vDash \phi_k^-(x)$

Assign $\;0$ *to node* n *in* X iff $X, n \vDash \phi_k^0(x) = \neg(\phi_k^+(x) \vee \phi_k^-(x))$

where $X, n \vDash \psi$ iff ψ is `True` in model X when n is substituted for x

Ex. *I/Q$^-(x) = \texttt{high_ftr}(x) \wedge (\texttt{uvular}(\texttt{prec}(\texttt{seg}(x))) \vee \texttt{uvular}(\texttt{succ}(\texttt{seg}(x))))$

(see Bird & Blackburn 1991, Bird & Klein 1990, 1994, Bird 1995, Potts & Pullum 2002, Jardine 2014, 2017, Chandlee & Jardine 2019, a.o. for logical characterizations of representations and constraints)

Logical structure

Let Φ be the set of all **conjunctions** of $\{\phi_k^+(x), \phi_k^-(x), \phi_k^0(x)\}_{k=1}^K$ that contain exactly one formula for each constraint c_k. Each conjunction is a possible **total marking** of a node

For a fixed weighting $\mathbf{w}$ of the constraints and $\phi \in \Phi$, we can determine the harmony of a node n in any model X such that $X, n \vDash \phi$. There is exactly one such ϕ for each node

Let $G_{\mathbf{w}}(\Phi)$ be the subset of Φ such that the harmony of a node satistifying $\phi \in G_{\mathbf{w}}(\Phi)$ is exactly 0 under weighting $\mathbf{w}$. This subset specifies all local configurations of well-formedness

Then model X has exactly zero harmony under weighting $\mathbf{w}$ iff $\forall n \in X \, (X, n \vDash \bigvee G_{\mathbf{w}}(\Phi))$

$\star$ As all logics of interest are closed under conjunction and disjunction, the logic needed to state the entire grammar is *no more expressive* than that needed to formalize the individual constraints

Outline

- **Components of HG/OT**
- **Static Harmonic Grammar**
- **Case studies**
- **Logical structure**
- Further properties of and open issues in SHG

Further properties

- Is there a version of SHG that uses strict domination instead of weighting?
  Can we have Optimality Theory without optimization?!

  SOT: *A node n is **well-formed** in X iff every constraint that assigns a negative mark to n is dominated by some constraint that assigns a positive mark to n*

- What predictions about constraint interaction does SHG/SOT inherit from HG/OT?
  - Transitivity of weighting/ranking (see McCarthy 1997 on "process specific constraints")

- Aren't *positive constraints* known to be harmful because they prefer *infinite words*?
  - Positive marks in SHG/SOT are insulated within nodes: application of $\min_\epsilon$ node-wise ensures positive values cannot 'leak out' and cancel negative marks elsewhere in the representation

- In SHG/SOT, what ensures that each input (UR) has a *single* well-formed output (SR)?
  - Nothing! But excessive output variation would have low communicative fitness
  - Output variation is observed, see the large literature on partially or variably ranked constraints

Further properties

- Is there a **probabilistic** version of SHG like MaxEnt grammars (Hayes & Wilson 2008)?
  - Yes! For threshold $\epsilon = 0$, $p(X) \propto 1$ if X is grammatical, < 1 otherwise
  - For threshold $\epsilon < 0$, $p(X)$ declines with $\#$ of nodes all other things being equal
  - Conditional version $p(Y|X) = \frac{1}{Z(X)} \exp(H(Y|X))$ where X is input part of Y
- Is there a **serial** version of SHG/SOT like serial HG/OT (e.g., Prince & Smolensky 1993/2004, McCarthy 2008, 2011, Tessier 2013, Jarosz 2014, McCarthy & Pater 2015)?
  - Maybe! Define step $X \rightarrow Y$ as grammatical iff $H(Y|X) = 0$
    and allow serial derivation to halt when reach fixed point $H(X|X) = 0$
- Probabilistic serial version of SHG is also possible by combining these definitions

Further properties

- How is SHG/SOT related to **targeted constraints** (e.g., Wilson 2000, 2001, 2003, Hansson 2001, McCarthy 2002, Pulleyblank 2006)?
  - Some of the properties of targeted constraints are shared: for example, the pressure to assimilate or dissimilate can be concentrated on just one of two interacting elements
    Ex. Assign $-u(\alpha)$ to an [α lateral] feature in the context [α lateral] __ on the liquid tier
  - Otherwise the proposals are quite different, in particular the theory of targeted constraints assumed that grammaticality was determined by harmonic ordering of forms

- Is SHG/SOT equivalent to performing optimization **locally**, within each constituent or other domain, instead of globally? (e.g., Frank & Satta 1998, Heck & Mueller 2007)?
  - No! The harmony of each node can depend on any other node *as mediated by the relations* that are part of the representations (logical models) permitted by **Gen**
  - *There is no domain smaller than the prosodic word* that can delimit phonological interactions
    segment? no: vowel lowering before/after uvulars
    syllable? no: place assimilation of Coda nasal to Onset stop
    foot? no: assimilation across Ft boundaries, assignment of main stress

Further properties

2. A FRAMEWORK FOR PDP 47

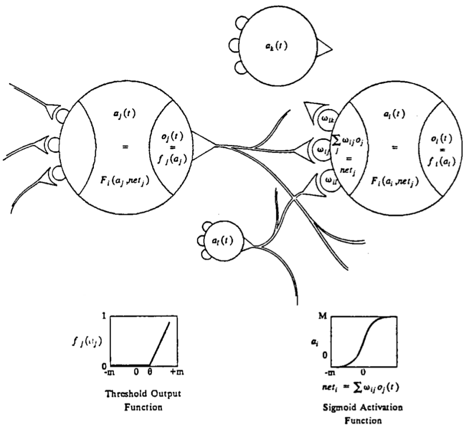

FIGURE 1. The basic components of a parallel distributed processing system.

SHG has a close conceptual relationship to idealized models of neural computation (e.g., Rumelhart, Hinton & McClelland 1986)

- Nodes send weighted excitatory $(+1)$ and inhibitory (-1) signals to one another through highly structured patterns of synaptic connection
- Each node is a threshold logic unit, summing up its incoming signals and activating according to a non-linearity
  n is **active** in X iff $H(n; X) = 0$

SHG was directly inspired by the neural network model of context-free tree grammars and parsing in Hale & Smolensky 2001, but differs in applying $\min_\epsilon$ non-linearity at each node

Remaining issues

- How can SHG/SOT account for interactions that do not appear localizable to a single node, such as vowel epenthesis to split marked clusters or blocking of vowel deletion by the OCP?
  - One approach to epenthesis, adapting an idea from directional evaluation (Eisner 2000), is to pool marks from the epenthetic and immediately preceding/following segments
  - More generally, some mechanism of locally pooling marks *or* directly rewarding 'buffer' segments that break up marked sequences will have to be incorporated into the theory
- What principles determine the type of node that receives marks from a given constraint?
  - Generally place marks as 'low' as possible, thus maximally restricting constraint interaction
  - Forces the issue of **locus of violation** that has arisen for **directional evaluation** (Eisner 2000), targeted constraints, (non-)gradient evaluation (McCarthy 2003), and elsewhere
- Does SHG/SOT predict plausible **factorial typologies** under reweighting/reranking?
  - Depends on constraint set but initial results (e.g., grid-based stress) are promising
  - Help! No current resuts on finding typological predictions except by trying various weightings/rankings

Remaining issues

- What is the computational complexity of SHG/SOT?
  - Depends on constraint set and close relationship between logics and machines (automata)
  - *Verifying* grammaticality of X involves computation proportional to $\#$ of nodes in X
  - *Constructing* a grammatical X may be costly and require search (even optimization!)

- How is constraint weighting/ranking learned from positive evidence in SHG/OT?
  - *Implicit negative evidence* is available as in HG/OT: unobserved reps allowed by **Gen**
  - For gradient-based learning, try applying **straight-through estimator** to $\min_\epsilon$ non-linearity

$$\frac{\partial}{\partial w_k} \sum_n \min_\epsilon \left(\sum_{k=1}^{K} w_k c_k(n;X) \right) \approx \frac{\partial}{\partial w_k} \sum_n \sum_{k=1}^{K} w_k c_k(n;X)$$

(On straight-through gradient estimator see Hinton 2012, Bengio et al. 2013, Yin et al. 2019)

Why *Static* Harmonic Grammar?

Representations are evaluated in isolation, as fixed objects, rather than as points along a dynamic path of harmonic improvement (cf. Smolensky 1986, Legendre, Miyata & Smolensky 1990, Prince 1990, Goldsmith & Larson 1990, Smolensky, Goldrick & Mathis 2014)

Each node in a representation, and the representation itself, is in *static equilibrium* if the weighted positive and negative marks balance out in the $\min_0$ sense

statics *(plural noun, usually treated as singular)*
the branch of mechanics concerned with bodies at rest and forces in equilibrium

In classical mechanics, a particle is in mechanical equilibrium if the net force on that particle is zero. By extension, a physical system made up of many parts is in mechanical equilibrium if the net force on each of its individual parts is zero.

(credit Oxford Dictionary and Wikipedia)

Thank you!

And thanks to Eric Bakovic, Marina Bedny, Gillian Gallagher, Jonathan Rawski, Paul Smolensky for discussing this proposal without starting too much static