Connectionism and Harmony Theory in Linguistics

Alan Prince, Paul Smolensky

CU-CS-533-91 July 1991

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

Connectionism and Harmony Theory in Linguistics

Alan Prince, Paul Smolensky

CU-CS-533-91 July 1991

Department of Computer Science
University of Colorado at Boulder
Campus Box 430
Boulder, Colorado 80309-0430 USA

(303) 492-7514
(303) 492-2844 Fax

Class notes from a 4-week course of the same name taught at the Linquistics Institute, University of California at Santa Cruz, sponsored by the Linquistics Society of America

Connectionism and Harmony Theory in Linguistics

LSA Linguistic Institute
University of California, Santa Cruz
LINS 247
July 1991

Topics

Spreading Activation & Harmony Maximization

Spreading activation

PDP-I:1,2

Harmony maximization

PDP-III:3[49-53]; PDP-II:14[7-38]; PDP-I:6

Goldsmith-Larson Dynamic Linear Model

Goldsmith, John:

Local modeling in phonology.

A dynamical computational theory of accent systems.

Prince, Alan:

Convergence of the Goldsmith-Larson Dynamic Linear Model of sonority and stress structure.

Closed-form solution of the Goldsmith-Larson Dynamic Linear Model of syllable and stress structure and some properties thereof

Inductive Learning & Error Minimization

Hebb rule

PDP-I:1[31-40]; PDP-III:4[83-86],5[90-93]

Delta rule

PDP-III:[86-89,93-96,121-130]; PDP-I:11

Past Tense Acquisition Model

PDP-II:18

Pinker, Steven & Prince, Alan.

Rules and connections in human language.

On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. [Pinker & Mehler.]

Back Propagation

PDP-III:5[130-137]; PDP-I:8

NetTalk

Sejnowski, Terrence J. & Rosenberg, Charles R. Parallel networks that learn to pronounce English text.

Recurrent back propagation

PDP-I:8[354-361]

Elman, Jeffrey. Finding structure in time.

Continuity & Similarity

Hungarian Vowel Harmony Model

Hare, Mary. The role of similarity in Hungarian vowel harmony: A connectionist account.

Harmonic Grammar

Legendre, Geraldine, Miyata, Yoshiro, & Smolensky, Paul:

Can connectionism contribute to syntax? Harmonic Grammar, with an application.

Unifying syntactic and semantic approaches to unaccusativity: A connectionist approach.

Harmonic Grammar -- A formal mutli-level connectionist theory of linguistic well-formedness: An application.

Harmonic Grammar -- A formal mutli-level connectionist theory of linguistic well-formedness: Theoretical foundations.

Harmonic Phonology

Lakoff, George. A suggestion for a linguistics with connectionist foundations.

Touretzky, David S., Wheeler, Deirdre W, & Elvgren, Gillette. Rules and maps III: Further progress in connectionist phonology.

Goldsmith, John. Phonology as an intelligent system.

Prince, Alan & Smolensky, Paul. Class Notes

Sources

Unless otherwise indicated, all articles are in the Reader.

'PDP-X:Y[a-b]' denotes 'PDP books, volume X, chapter Y, pages a-b'

On reserve in library & available through bookstore:

PDP-I: Rumelhart, David E., McClelland, James L., & the PDP Research Group. *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press/Bradford Books.

PDP-II: McClelland, James L., Rumelhart, David E., & the PDP Research Group. *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Cambridge, MA: MIT Press/Bradford Books.

PDP-III: McClelland, James L. & Rumelhart, David. *Explorations in parallel distributed processing: A handbook of models, programs and exercises.* Cambridge, MA: MIT Press/Bradford Books.

Pinker & Mehler: Pinker, Steven, & Mehler, Jacques (Eds.) (1988). *Connections and symbols.* Cambridge, MA: MIT Press/Bradford Books.

Available through Institute:

Reader: Published articles and reports

Class Notes:

To be cited as:

Prince, A. & Smolensky, P. (1991). *Notes on Connectionism and Harmony Theory in Linguistics.* Technical Report CU-CS-533-91, Department of Computer Science, University of Colorado, Boulder.

Linguistic Background *(incomplete)*

LINS 247
Connectionism and Harmony Theory in Linguistics
P. Smolensky/ A. Prince
July, 1991

Archangeli, D. and D. Pulleyblank (July, 1991) "African Tongue Root Harmony: Optimal Systems," paper given at NSF/LSA Conference on Phonological Feature Organization, UCSC, Santa Cruz.

Batistella, E. (1990) *Markedness: the Evaluative Superstructure of Language*, SUNY Press, Albany.

Cairns, C. and M. Feinstein (1982) "Markedness and the Theory of Syllable Structure," *Linguistic Inquiry* 13, 193-226.

Chomsky, N. (1951) *The Morphophonemics of Modern Hebrew*, MA dissertation, U. Penn., Philadelphia.

Chomsky, N. and M. Halle (1968) *The Sound Pattern of English*, Chapter 9, Harper & Row, New York.

Chomsky, N. (1981) *Lectures on Government and Binding*, Foris: Dordrecht.

Clements, G. N. (1988) "The Role of the Sonority Cycle in Core Syllabification," *Working Papers of the Cornell Phonetics Laboratory* 2, 1-68.

Goldsmith, J. (1990) *Autosegmental and Metrical Phonology*, Basil Blackwell, Oxford, pp. 319-332.

Goldsmith, J. (in press) "Phonology as an Intelligent System," in Lea, ed., *Bridges between Psychology and Linguistics*, Falco & Falco, Inc., East Brunswick, NJ.

Hale, K. (1973) "Deep-Surface Canonical Disparities in Relation to Analysis and Change: An Australian Example," *Current Trends in Linguistics* 11, 401-458.

Hayes, B. (1980/81) *A Metrical Theory of Stress Rules*, Ph. D. Dissertation, MIT, Cambridge.

Hayes, B. (1991) *Metrical Stress Theory: Principles and Cases Studies*, ms. UCLA, Los Angeles.

Hyman, L. (July, 1991) "Imbrication in Čibemba," presentation at UCSC, Santa Cruz.

Itô, J. (1986) *Syllable Theory in Prosodic Phonology*, Ph. D. dissertation, UMass, Amherst.

Itô, J. (1989) "A Prosodic Theory of Epenthesis," *Natural Language & Linguistic Theory* 7, 217-260.

Itô, J. (1991) "Prosodic Minimality in Japanese," in K. Deaton, M. Noske, and M. Ziolkowski, eds. *CLS-26 II*, University of Chicago, Chicago.

Itô, J. and R.-A. Mester (1986) "The Phonology of Voicing in Japanese: Theoretical Consequences of Morphological Accessibility," *Linguistic Inquiry* 17, 49-73.

Itô, J. and R.-A. Mester (1991a) *Coursebook*, LSA Summer Institute, UCSC, Santa Cruz.

Itô, J. and R.-A. Mester (1991b) "Melodic and Prosodic Licensing," paper given at LSA Institute Feature Workshop, UCSC, Santa Cruz.

Kaye, J., J. Lowenstamm (1984) "De la syllabicité," in F. Dell, D. Hirst, and J.-R. Vergnaud, eds., *Forme sonore du langage*, Hermann, Paris, pp. 123-159.

Kaye, J., J. Lowenstamm, and J.-R. Vergnaud (1985) "The Internal Structure of Phonological Elements: a Theory of Charm and Government," *Phonology Yearbook* 2, 305-328.

Kenstowicz, M. and C. Kisseberth (1979) *Generative Phonology: Description and Theory*, Academic Press, Inc., New York.

Kenstowicz, M. and C. Kisseberth (1979) *Topics in Phonological Theory*, Academic Press, Inc., New York.

Kiparsky, P. (ca. 1980) "Vowel Harmony, " ms. MIT.

Kiparsky, P. (ca. 1974) "Abstractness, Opacity, and Global Rules," in O. Fujimura, ed., *Three Dimensions of Linguistic Theory*.

Kiparsky, P. (1982) "Lexical Phonology and Morphology," in I. S. Yang, ed., *Linguistics in the Morning Calm*, Linguistic Society of Korea, Hanshin, Seoul.

Kiparsky. P. (1983) "Some Consequences of Lexical Phonology," ms., MIT, Cambridge.

Kisseberth, C. (1970) "On the Functional Unity of Phonological Rules," *Linguistic Inquiry* 1, 291-306.

Lakoff, G. (in press) "Cognitive Phonology," in J. Goldsmith, ed., *The Last Phonological Rule* [sic], University of Chicago Press, Chicago.

Levin, J. (1985) *A Metrical Theory of Syllabicity*, Ph. D. dissertation, MIT, Cambridge.

Liberman, M. and A. Prince (1977) "On Stress and Linguistic Rhythm," *Linguistic Inquiry* 8.2, 249-336.

LaPointe, S. and M. Feinstein (1982) "The Role of Vowel Deletion and Epenthesis in the Assignment of Syllable Structure," in H. v. d. Hulst and N. Smith, eds., *The Structure of Phonological Representations (Part II)*, Foris, Dordrecht, pp. 69-120.

McCarthy, J. (1986) "OCP Effects: Gemination and Antigemination," *Linguistic Inquiry* 7, 207-264.

McCarthy, J. and A. Prince (1986) *Prosodic Morphology*, ms. UMass Amherst and Brandeis University, Waltham.

McCarthy, J. and A. Prince (1990) "Foot and Word in Prosodic Morphology: the Arabic Broken Plural," *Natural Language & Linguistic Theory* 8, 209-283.

Mester, R.-A. (1986) *Studies in Tier Structure*, Ph. D. dissertation, UMass, Amherst.

Myers, S. (1991) "Phonological Well-formedness," to appear in *Linguistic Inquiry*.

Noske, R. (1982) "Syllabification and Syllable Changing Rules in French," in H. v. d. Hulst and N. Smith, eds., *The Structure of Phonological Representations (Part II)*, Foris, Dordrecht, pp. 257-310.

Paradis, C. (1988) "On Constraints and Repair Strategies," *The Linguistic Review* 6, 71-97.

Paradis, C. and J.-F. Prunet (in prep.) "Constraints in Conflict and the Phonological Level Hierarchy," ms. Université de Laval and Université du Québec à Montréal.

Perlmutter, D. (1971) *Deep and Surface Structure Constraints in Syntax*, Holt, Rinehart, and Winston, New York.

Prince, A. (1991) "Quantitative Consequences of Rhythmic Organization," in K. Deaton, M. Noske, and M. Ziolkowski, eds. *CLS-26 II*, University of Chicago, Chicago.

Prince, A. and P. Smolensky (April, 1991) "Optimality," paper given at Arizona Phonology Conference, UAz, Tucson.

Ross, J. R. (early 1970's) Various unpublished mss. re conspiracies and output constraints, MIT, Cambridge.

Selkirk, E. O. (1983) *Phonology and Syntax: the Relation between Sound and Structure*, MIT Press, Cambridge.

Selkirk, E. O. (1981) "Epenthesis and Degenerate Syllables in Cairene Arabic," in H. Borer and J. Aoun, eds., *Theoretical Issues in the Grammar of the Semitic Languages*, MIT, Cambridge.

Singh, R. (1987) "Well-formedness Conditions and Phonological Theory," in W. Dressler et. al, eds, *Phonologica 1984*.

Smolensky, P. (1986) "Information Processing in Dynamical Systems: Foundations of Harmony Theory," in D. Rumelhart, J. McClelland, and P.D.P.R. Group, eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, I: Foundations*, pp. 104-281.

Sommerstein, A. (1974) "On Phonotactically Motivated Rules," *Journal of Linguistics* 10, 71-94.

Stampe, D. (1973/9) *A Dissertation on Natural Phonology*, Ph. D. dissertation, University of Chicago, Garland Publishing Co., New York.

Steriade, D. (1982) *Greek Prosodies and the Nature of Syllabification*, Ph. D. Dissertation, MIT.

Wilkinson, K. (1986) "Syllable Structure and Lardil Phonology," ms. UMass, Amherst.

Wilkinson, K. (1988) "Prosodic Structure and Lardil Phonology," *Linguistic Inquiry* 19, 325-334.

Yip, M. (1988) "The Obligatory Contour Principle and Phonological Rules: a Loss of Identity," *Linguistic Inquiry* 19, 65-100.

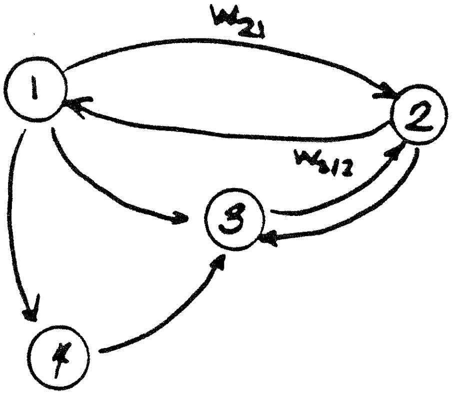
Optimization in Connectionism & Linguistics

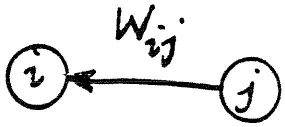
I.	Spreading activation & Harmony maximization	Phonological prominence: stress, syllable structure
II–III	Inductive learning & Error minimization	Past tense inflection (English) Pronunciation (English) Induction of grammatical categori through distributional analysis (Eng
IV	Continuity & similarity	Vowel harmony (Hungarian)
	Harmonic Grammar	Unaccusativity (French)
V–VIII	Harmonic Phonology	• Universal phonology • Typologies • Syllable structure • Stress systems • Phonologies of particular languages

Assignment:
- 5-10pp. paper due last class
- Outline due one week before last class

Materials:
- Reader (published papers/reports)
- Class Notes (unpubl. drafts by P&S; handouts; overheads
- Bookstore
- Library

Summary of 7/8





nodes – activation

connections – weights

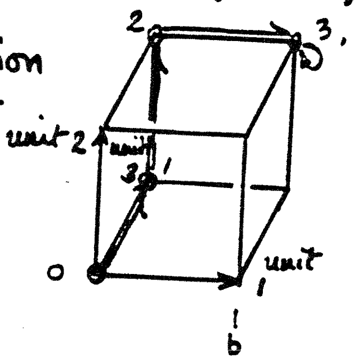
$$Net_i = \sum_j W_{ij} a_j$$

$$a_i = f(Net_i)$$

$$\begin{array}{rc} & \text{units} \\ \text{Iteration} & \begin{array}{cccc} & 1 & 2 & 3 \\ 0 & (0 & 0 & 0) \\ 1 & (0 & 0 & b) \\ 2 & (0 & b & b) \\ 3 & (b & b & b) \\ 4 & (b & b & b) \\ \vdots & \vdots & \vdots & \vdots \end{array} \end{array}$$

time ↓

activation vector $\vec{a}$



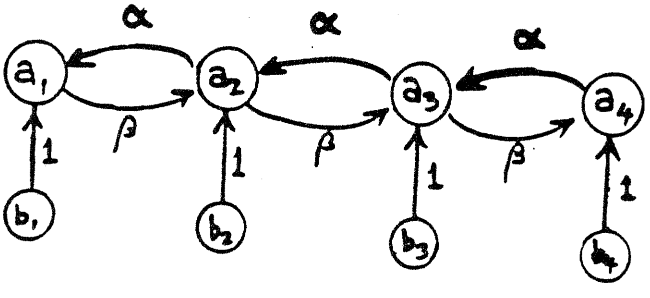
'state space'

'dynamical system'

Analogy: spreading activation ~ derivation (e.g. autosegmental spreading)

→ 'structure building'

Goldsmith – Larson D.L.M.



prominence

- stress (over syllables)
- sonority (over segments)

	α	β	final $\vec{a}$	$\vec{b} = (0\ 0\ 0\ b)$
Examples:				
	1	0	(b b b b)	leftward spreading
	-1	0	(-b b -b b)	leftward alternation
	$-\frac{1}{2}$	0	$(-\frac{1}{8}\ \frac{1}{4}b\ -\frac{1}{2}b\ b)$	damped " "
(1 1 1 1)	$-\frac{1}{2}$	$-\frac{1}{2}$	$(\frac{1}{2}\ \frac{3}{2}\ \frac{3}{2}\ \frac{1}{2})$	interdicitated rises [illegible]

* Connectionist computation is optimization.

I. The basic (low-level) computational mechanism of conn. is spreading activation.

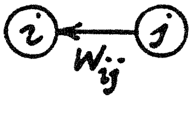
1. Interesting things happen – in new & different ways – due to this novel mechanism.
    1. Goldsmith & Larson's Dynamic Linear Model (DLM)

    [2... to come]

2. High level understanding of spreading activation – global analysis – is possible through ...
    1. ... Linear Analysis
        1. The activation pattern is a vector in a high-dimensional state space evolving through time – a dynamical syste
        2. The connection strengths form a weight matrix which can be analyzed using linear algebra; e.g., the DLM can be solved in closed form.

        [3. Related methods allow us to use vectors to represent and matrices to process structured data...]
    2. ... Harmonic Analysis
        1. In certain nets, activation spread steadily increases a numerical well-formedness measure on activation vectors –

$$\text{Harmony} = \sum_i \sum_j a_i W_{ij} a_j$$



        2. Activation spread in these nets is an algorithm for

3. H maximization can be interpreted as (soft) constraint satisfaction

4. These constraints can be used to declaratively define grammar.
   Global H maxima define the grammatical structures (competenc
   Local H maxima are computed by connectionist nets (performan

3. Activation has several interpretations
   1. (Graded) truth value of a hypothesis.
   2. Value of a (graded) linguistic quantity
      1. DLM: (continuous) context-dependent prominence

[3. None – meaningless internal computational data]

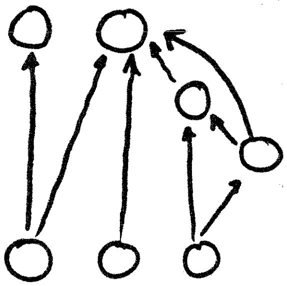
Harmonic nets:

1. activation rule: 'If Net>0, excite, else inhibit'

2. connectivity:

   either

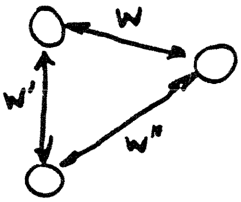
   - feedforward



'all connections go up'

   or

   - symmetric



'all connections go both ways'

3. updating activations

   either

   - one-at-a-time ('asynchronous')

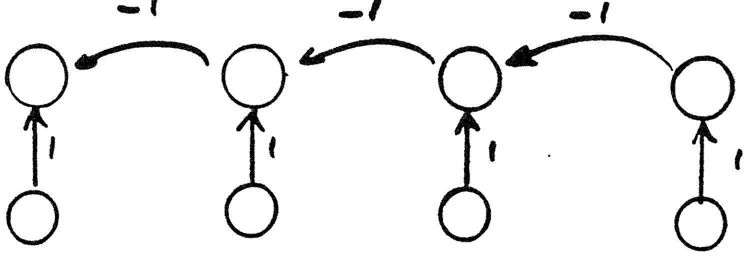
   or

   - smooth (unit activations change very little at each step)

Notes:

1. Ensures $\Delta a_i \Rightarrow \Delta h_i > 0$ 'each update separately locally harmoni [1,2]
2. Ensures $\Delta a_i \Rightarrow \Delta H > 0$ 'each update separately globally harmoni [3-6]
3. Ensures $\{\Delta a_i\}$ [illegible] $\Rightarrow \Delta H > 0$ 'all updates [illegible]'

GL D.LM.

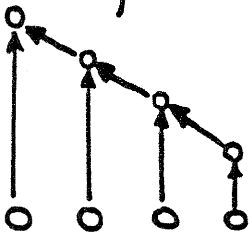


$\vec{b}$ =(0 0 0 b)

$\underline{H} = \sum_i \sum_j a_i w_{ij} a$

time ↓ $\vec{a}$=(	0	0	0	0)	0
	0	0	0	b	0
	0	0	-b	b	0 + (-b)(-1)(b) =
	0	b	-b	b	b^2 + (b)(-1)(-b) =
	-b	b	-b	b	$2b^2$ + (-b)(-1)(b) =
	-b	b	-b	b	$3b^2$
	⋮	⋮	⋮	⋮	⋮

H monotonically increases ... why?



1. Net is feedforward
   'all arrows go up'

2. Only one activation value changes at a time

Ex: Look at other examples of G-L D.L.M. from Lecture 1

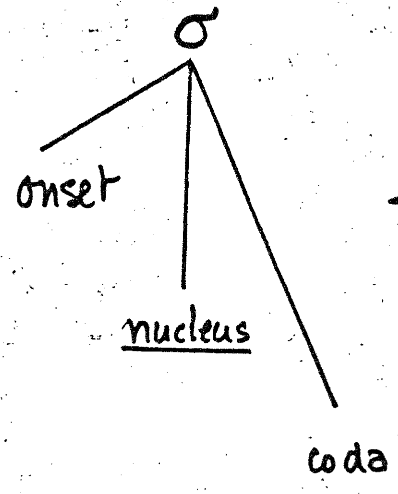
Constraint satisfaction view:

If i+1 is strong, i should be weak;
" " " weak " " " strong.

Every 'rule application' in this 'derivation' is harmonic:

Syllabification in Berber

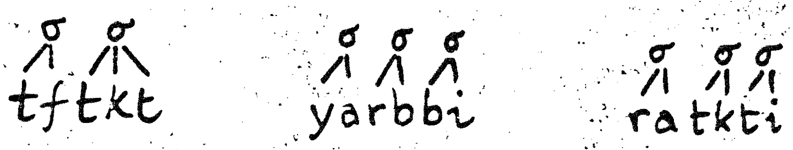
Berber: F. Dell & M. El-Medlaoui, 1987, 1988



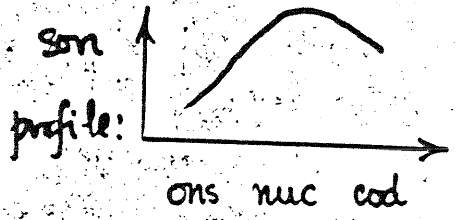
← = 1 segment (except phrase-initially)

← = 1 segment NOT NEC. A VOWEL prism: Br-br

← ≤ 1 segment (except phrase finally)



What governs syllabification? 'Ideal' sonority profile:



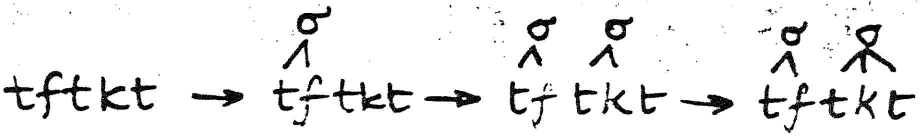
Berber's 8-level sonority heirarchy:

a >	u, i >	r, l >	m, n >	z >	f, s, x >	b, d, g >	p, t, K
8	7	6	5	4	3	2	1

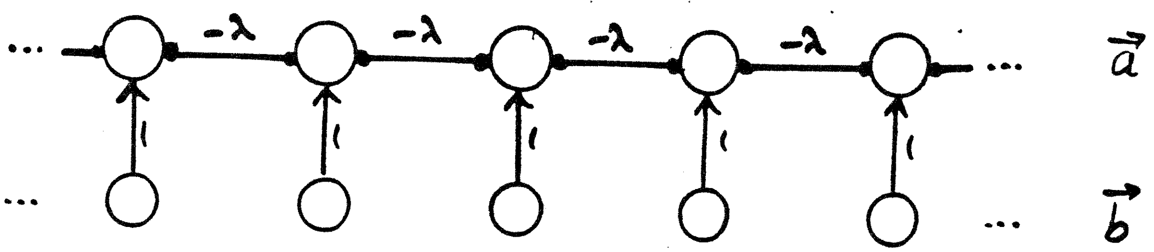
Dell-El Medlaoui Algorithm:

```
FOR s = 8, 7, ..., 1
    SCAN LEFT-TO-RIGHT
        IF FIND $S_i$ $S_{i+1}$ ($\text{son} = s$) THEN BUILD $\overset{\sigma}{S_i\ S_{i+1}}$
ATTACH any remaining unsyllabified segments as CODAs
```

Eg.:



Harmonic BrbrNet



Each $a_i \in [0,1]$ during computation; finally,

each $a_i = 0$ ('non peak') or $a_i = 1$ ('peak')

'$a_i = 1\{\text{segment } i \text{ is a peak}\}$'

Each b_i is a measure of the sonority of segment i;

(In G-L, $b_i = \text{son}(s_i)$; here, $b_i = 2^{\text{son}(s_i)} - 1$; $\text{son}(s_i) \in \{1, 2,$

'exponential sonority scale in Berber')

Activation rule:

To a_i add $c \cdot \text{Net}_i$; if this is > 1, set $a_i = 1$
" " " < 0, set $a_i = 0$

c small
consta

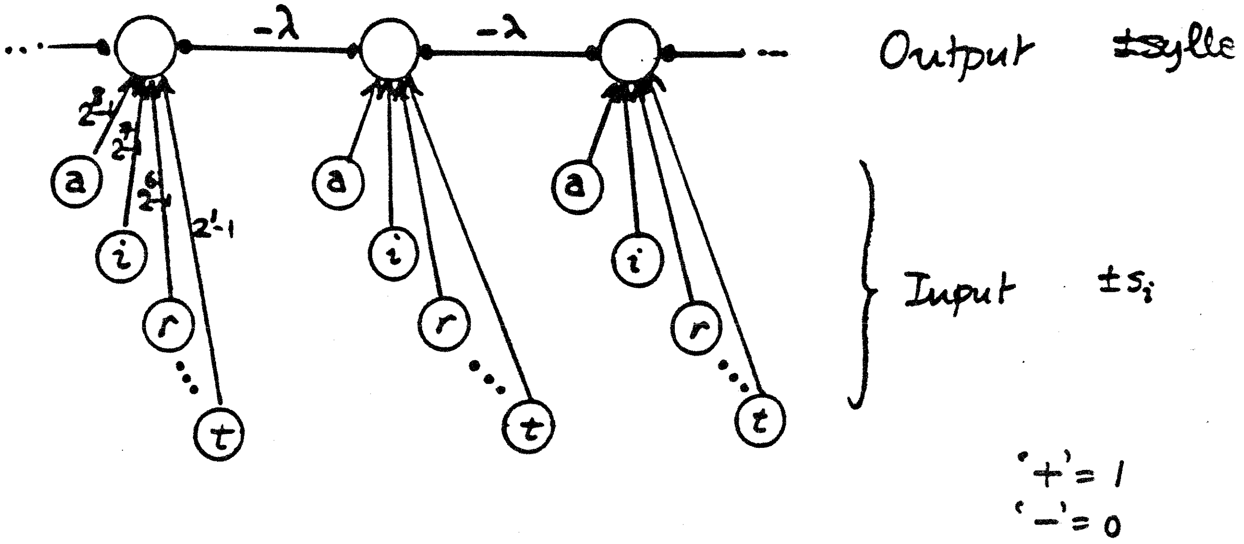
Note: 1. 'If Net > 0, excite, else, inhibit'
2. Connections are symmetric.
3. Activation updating is smooth.

∴ Harmonic

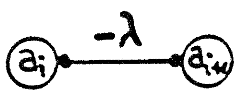
Theorem: 1. For sufficiently large λ, $\vec{a}$ is a <u>local</u> H maximum iff $\forall i\, a_i \in \{0,1\}$ and the syllabification satisfies the Berber syllabification constraints: *VV; *CCC

2. For sufficiently large λ, $\vec{a}$ is a <u>global</u> H maximum iff $\forall i\; a_i \in \{0,1\}$ and corresponds to the correct Berber

Harmonic Gammar View of BrbrNet



Grammar of soft rules:

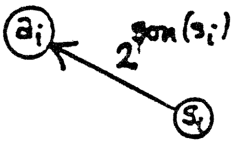


$*(+syl_i)\&(+syl_{i+1})$ strength: λ

↖subtract 'strength' from H if violate

↙add 'strength' to H if satisfy

$\surd(+s_i)\ \&\ (+syl_i)$ strength: $2^{son(s_i)}-1$



Principle of Harmonic Structural Assignment:

Assign the structure that maximizes H

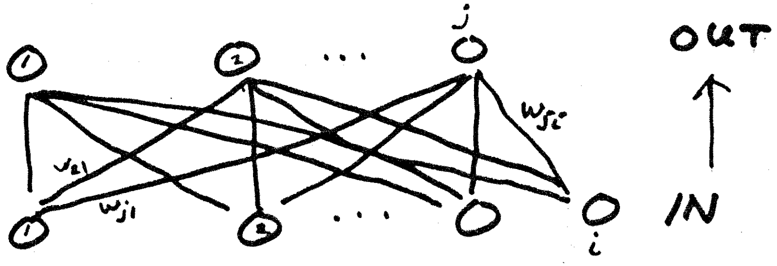
Learning from Experience

The problem: suppose we have a collection of InPut Patterns $P_1, \ldots, P_n$,

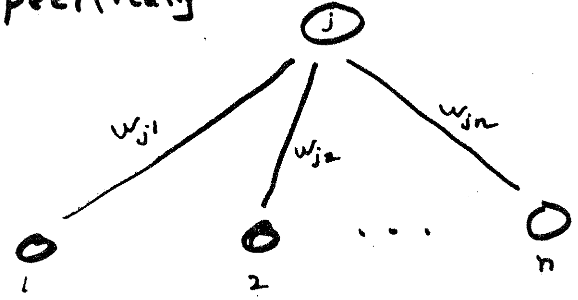
each matched with an output Target Pattern $T_1, \ldots, T_n$.

- Can we get a NETWORK to LEARN to produce T_i given P_i ??

Concretely, find w_{ji} for



Specifically



We have

$$net_j = a_1 w_{j1} + a_2 w_{j2} + \cdots a_n w_{jn}$$

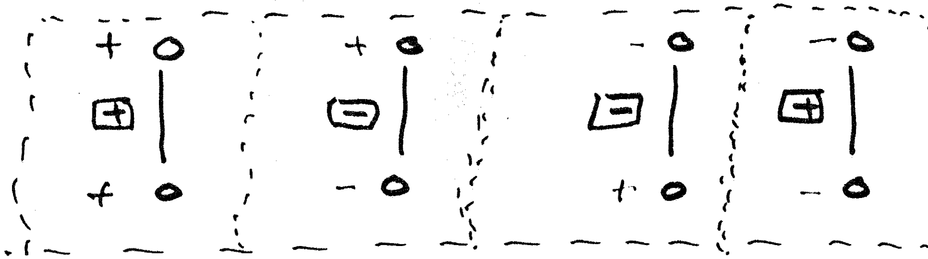
$$= \sum_i a_i w_{ji}$$

AND:
- excite out_j if $f(Net_j) > 0$
- inhibit out_j if $f(Net_j) \not< 0$

We therefore manipulate Net_j to achieve desired goal.

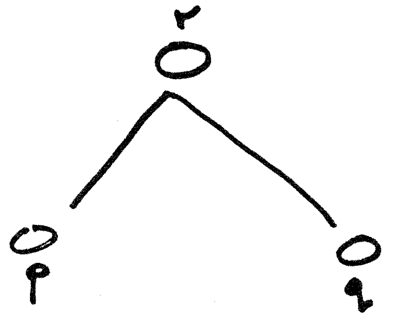
The inputs are fixed So we work on the WEIGHTS.

First Idea: Hebb Rule: reward node-wise correlation.



- Set $w_{ji} = a_j a_i$
- Sum over all input patterns

Example of Hebbian Learning



	p	q	$p \wedge q$
$\vec{p}_1$	1	1	1
$\vec{p}_2$	1	-1	-1
$\vec{p}_3$	-1	1	-1
$\vec{p}_4$	-1	-1	-1

	w_{rp}	w_{rq}
p_1	1	1
p_2	-1	1
p_3	1	-1
p_4	1	1
Σ	2	2

Technical Background

vector = $(x_1, x_2, \cdots, x_n) = \vec{x}$

Helps us to speak of chunks as significant:

"Pattern" = vector

$\vec{p}_1 = (1, 1)$ $\vec{p}_2 = (1, -1)$

dot product $\vec{a} \cdot \vec{x}$

a **number** not another vector

$$(a_1, a_2, \ldots) \cdot (x_1, x_2, \ldots)$$

$$= a_1 x_1 + a_2 x_2 + \cdots a_n x_n$$

$$= \sum_i a_i x_i$$

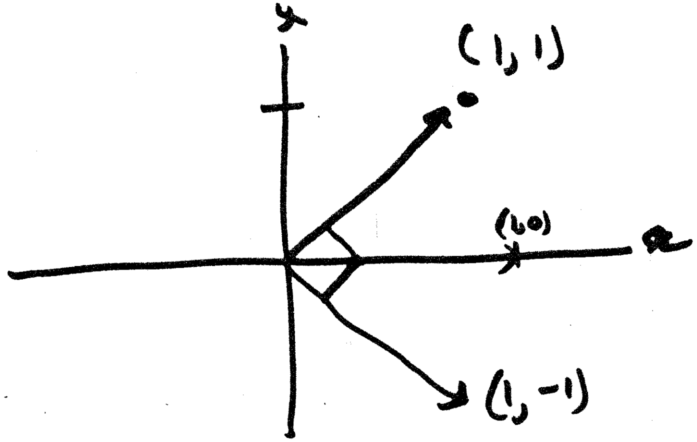
Dot Product measures angle between vectors of unit length.

$$\vec{a} \cdot \vec{x} = |a|\,|x| \cos\Theta$$

Perpendicular (Orthogonal) **Vectors**

$$\vec{a} \cdot \vec{x} = 0$$

Dot Product, dividing out length, gives **correlation**.



Matrix - rectangular array

$$\begin{array}{cc} & \begin{array}{ccc}\text{col 1} & \text{col 2} & \text{col 3}\end{array} \\ \begin{array}{c}\text{row}\\ \\ \text{row2}\\ \\ \text{row3}\end{array} & \left(\begin{array}{c|c|c} 11 & 12 & 13 \\ \hline 21 & 22 & 23 \\ \hline 31 & 32 & 33 \end{array}\right)\end{array}$$

EXAMPLE

$$\begin{array}{cc} & \text{in} \rightarrow \\ \begin{array}{c}\text{out}\\ \downarrow\end{array} & \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}\end{array}$$

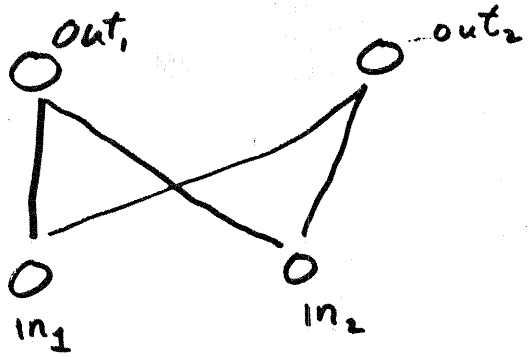
Two Interesting Cases

$$(a \quad b \quad c)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = (ax + by + cz)$$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}(x \quad y \quad z) = \begin{pmatrix} ax & ay & az \\ bx & by & bz \\ cx & cy & cz \end{pmatrix}$$

Hebb en bref.

$$\vec{w}_{j\bullet} = \begin{pmatrix} o \\ u \\ t \\ p \\ u \\ t \end{pmatrix}(i\,n\,p\,u\,t)$$



$$W = \begin{array}{cc} & \text{in} \rightarrow \\ \overset{\text{out}}{\downarrow} & \begin{pmatrix} out_1 \cdot in_1 & out_1 \cdot in_2 \\ out_2 \cdot in_1 & out_2 \cdot in_2 \end{pmatrix} \end{array}$$

$$= \begin{pmatrix} out_1 \\ out_2 \end{pmatrix}(in_1 \quad in_2)$$

Adding Vectors
& Adding Matrices

Rule: add corresponding elements.

$(a \quad b) + (c \quad d) = (a+c \quad b+d)$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} x & y \\ z & w \end{pmatrix}$$

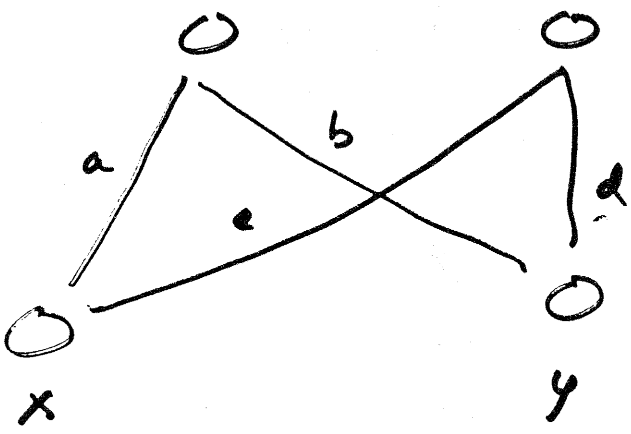
$$= \begin{pmatrix} a+x & b+y \\ c+z & d+w \end{pmatrix}$$

Multiplying Matrices

$A \cdot B = C$

$c_{ij} = \text{row } i(A) \cdot \text{col } j(B)$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$$



$W \cdot \vec{in} = \vec{net}$

Total Hebb Experience

- start off $w_{ji} = 0$
- $W = \sum_p \begin{pmatrix} o \\ u \\ t \\ p \\ u \\ t \\ p \end{pmatrix} (\text{input-}p)$ add up over all patterns

<u>Q</u>: What if a trained net faces a new input?

<u>A</u>: Compute!

$$\vec{out}_{New} = W \cdot \vec{input}_{New}$$

$$= \left[\sum \begin{pmatrix} o \\ u \\ t \\ p \\ u \\ t \\ s \end{pmatrix} (\text{input } s) \right] \cdot \begin{pmatrix} i \\ n \\ p \\ u \\ t \\ New \end{pmatrix}$$

DOT PRODUCT!

How the trained net behaves

① If <u>input$_{New}$</u> $\perp$ all items in training set, output is 0.

② If <u>input$_{New}$</u> $\perp$ all but <u>one</u>, get a scaled version of the output for that one

③ If <u>input$_{new}$</u> is non orthog (simil to several, output is a weighted <u>sum</u> of outputs proper to them. weighted by degree of similarity. (dot. prod)

$$\vec{out}_{New} = \sum_k \begin{pmatrix} o \\ u \\ t \\ p \\ u \\ t \\ k \end{pmatrix} \vec{input}_k \cdot \vec{input}_{New}$$

NON HEBB - LEARNABLE

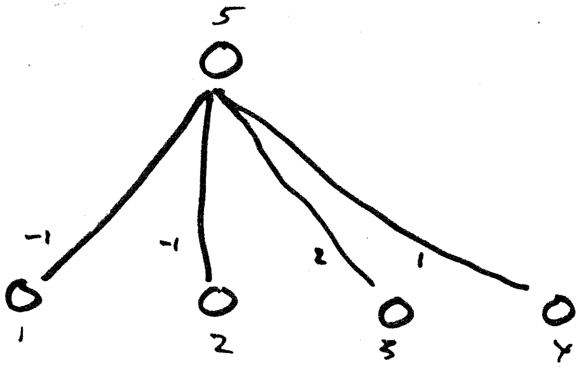
	NODE 1	2	3	4	OUT 5
P_1	+	−	+	−	+
P_2	+	+	+	+	+
P_3	+	+	+	−	−
P_4	+	−	−	+	−

Hebb Run:

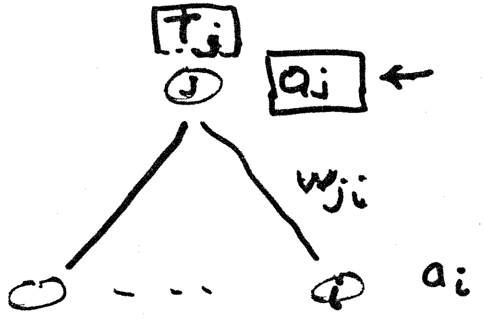
	w_{51}	w_{52}	w_{53}	w_{54}
P_1				
P_2				
P_3				
P_4				
Σ				

Hebb:

correct



Delta Rule



$$\Delta w_{ji} = \varepsilon \underbrace{(t_j - a_j)}_{\substack{\text{error} \\ \delta_j}} a_i$$

$$\Delta W = \varepsilon \sum_{k}^{\#patt.} \begin{pmatrix} \text{errors-}k \end{pmatrix} (\text{input-}k)$$

Using the Delta Rule

INPUTS = $\{P_1, \ldots, P_n\}$

OUTPUT TARGETS = $\{T_1, \ldots, T_n\}$

Procedure:

- Present Input Set (1 by 1)
- Compare Output Produced by net to desired target
- Make step proportional to error (ε) in right direction

Do it again

And again

⋮

$$W \leftarrow W + \Delta W$$

What is Error?

$$E = \frac{1}{2}\left(err_1^2 + err_2^2 + \dots err_n^2 \right)$$

$$= \frac{1}{2} \sum (t_j - o_j)^2$$

"sum squared error"

$$\Delta w_{ji} = -\varepsilon \frac{\partial E}{\partial w_{ji}}$$

No good -

$$\text{"}E\text{"} = \frac{1}{2}\left(err_1 + err_2 \dots err_n \right)$$

"signed error"

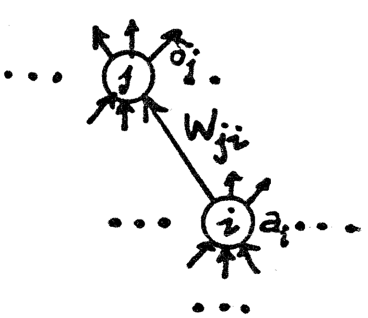
Perceptron Convergence Thm.

This procedure will find a set of weights that works if such a set exists.

BACK-PROP

Historical remarks

- * Learning internal representations – 'hard' vs 'easy' learning
  ('Boltzmann learning; Harmony T
  'hidden units'
- * Technical breakthrough

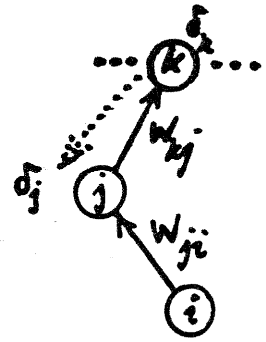


$$\Delta W_{ji} = \varepsilon \delta_j a_i$$

δ_j = 'error at j'

= ??? for 'hidden' unit?

Define recursively: $\delta_j = \left(\sum_k \delta_k w_{kj} \right) f'(net_j)$ interior

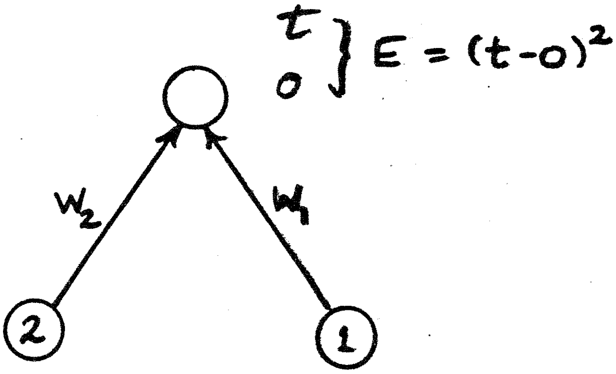


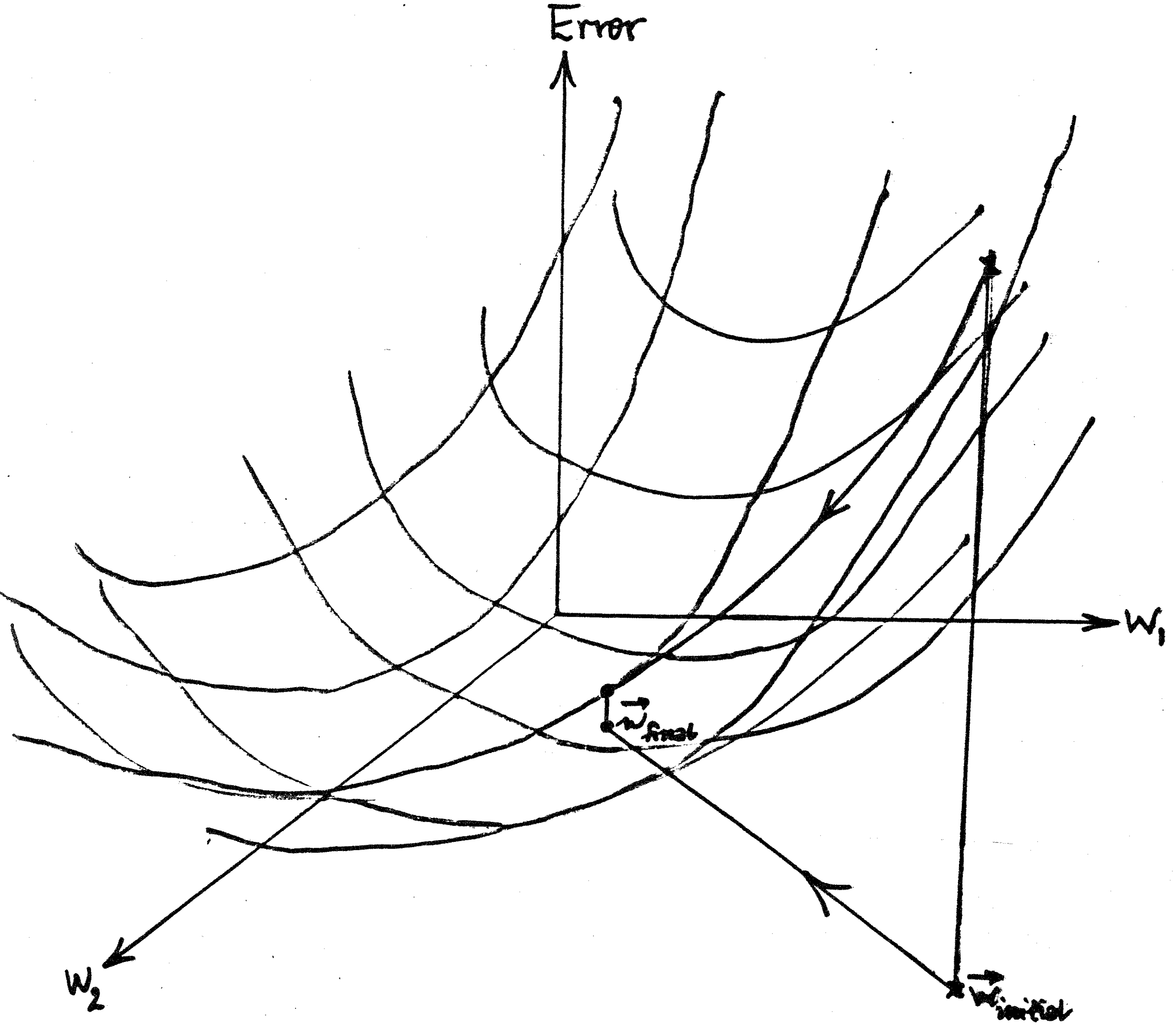
$= (t_j - a_j) f'(net_j)$ output

Gradient descent: $\vec{\Delta W}$ = direction of steepest descent in Error

$$\left.\begin{array}{l} E \\ + \\ \text{optimization technique} \end{array}\right\} = \begin{array}{l} \text{conn.} \\ \text{learning} \\ \text{algorithm} \end{array}$$

$$\Delta w_{ij} = -\varepsilon \frac{\partial E}{\partial w_{ji}}$$


t
o
$E = (t-o)^2$
W_2
W_1
2
1

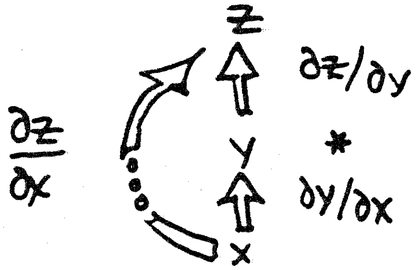

Error
W_1
W_2
$\vec{w}_{final}$
$\vec{w}_{initial}$

'$\Delta x \Rightarrow \Delta y$' := a change in x of amount Δx leads to a change in y of amount Δy

$$\frac{\Delta y}{\Delta x} := \text{rate of change of } y \text{ w.r.t. } x$$

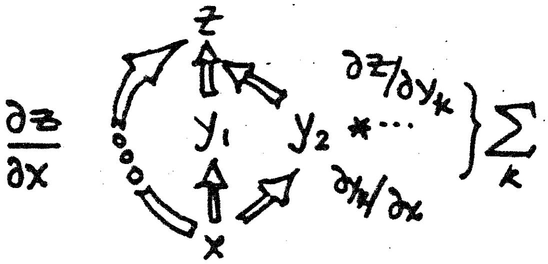
$$\rightarrow \frac{\partial y}{\partial x} \quad \text{for very small } \Delta x, \Delta y$$

Chain rule:

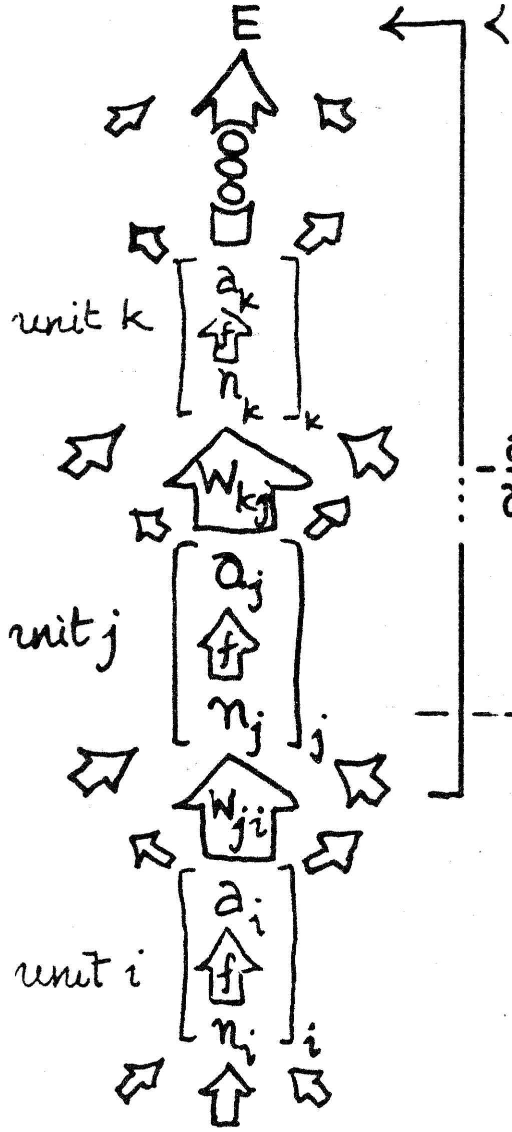


$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x}$$

Additive property:



$$\frac{\partial z}{\partial x} = \sum_k \frac{\partial z}{\partial y_k} \cdot \frac{\partial y_k}{\partial x}$$



$$-\frac{\partial E}{\partial w_{ji}} = -\frac{\partial E}{\partial n_j} \cdot \frac{\partial n_j}{\partial w_{ji}}$$

$$= \delta_j \cdot a_i$$

$$\delta_j = -\frac{\partial E}{\partial n_j} = \sum_k \frac{\partial a_j}{\partial n_j} \frac{\partial n_k}{\partial a_j} \frac{\partial E}{\partial n_k}$$

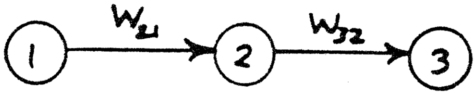
$$= \sum_k f'(n_j) w_{kj} \delta_k$$

Definitions:

$$a_j = f(n_j) \qquad \delta_j = -\frac{\partial E}{\partial n_j}$$

$$n_j = \sum_i w_{ji} a_i$$

propagation has no convergence theorem – why?



1	+1	1	+1	1
1	−1	−1	−1	1
1	+2	2	$+\frac{1}{2}$	1

$w_{21} \cdot w_{32} = 1$

$a_i = net_i$ 'linear net'

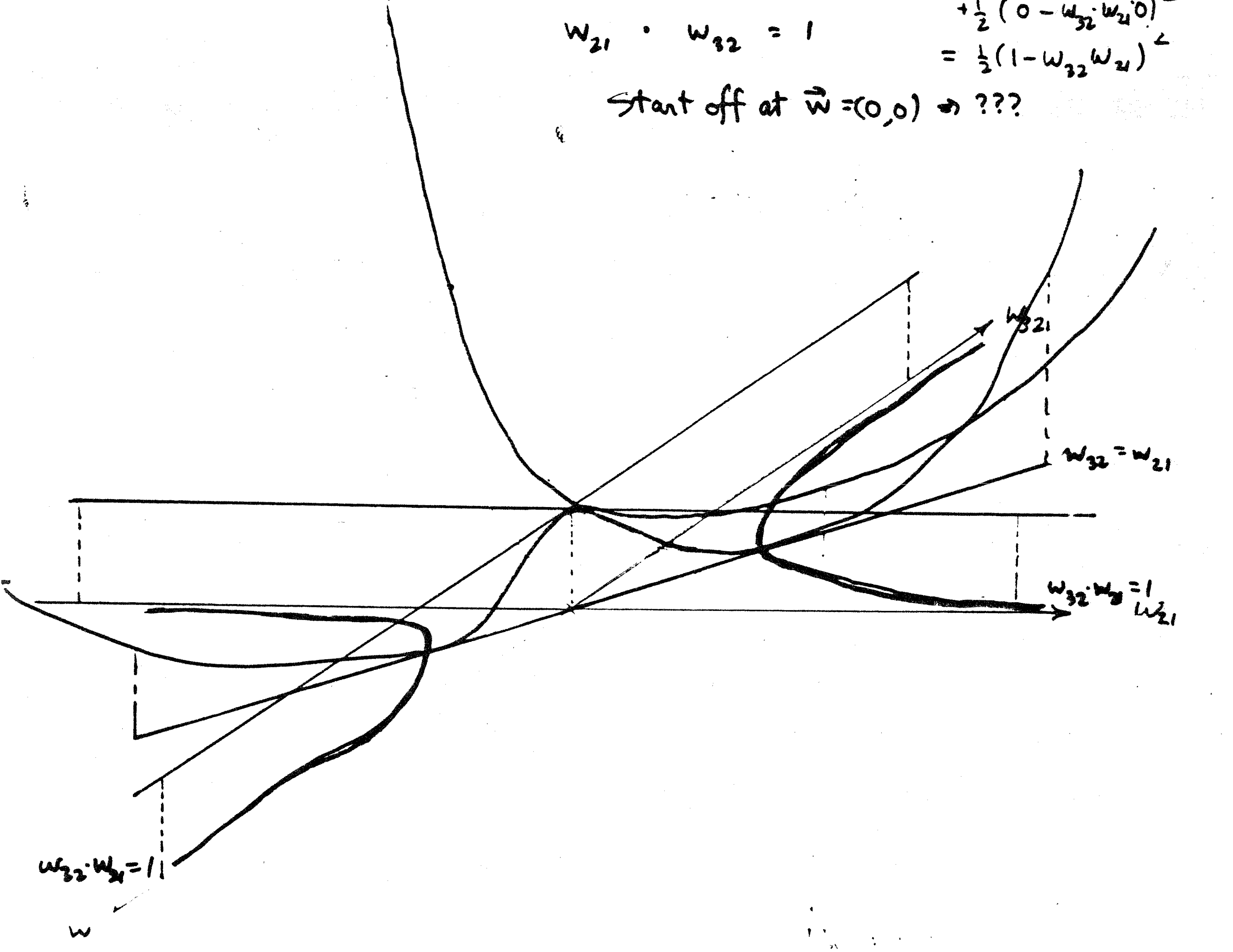
target = input 'auto-associal

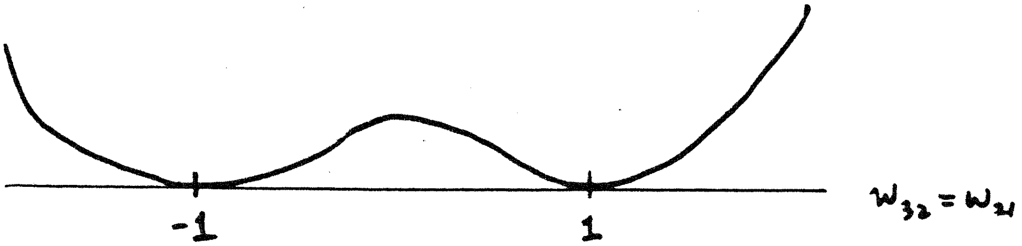
$t_3 = a_1$ inputs: $\{0, 1\}$

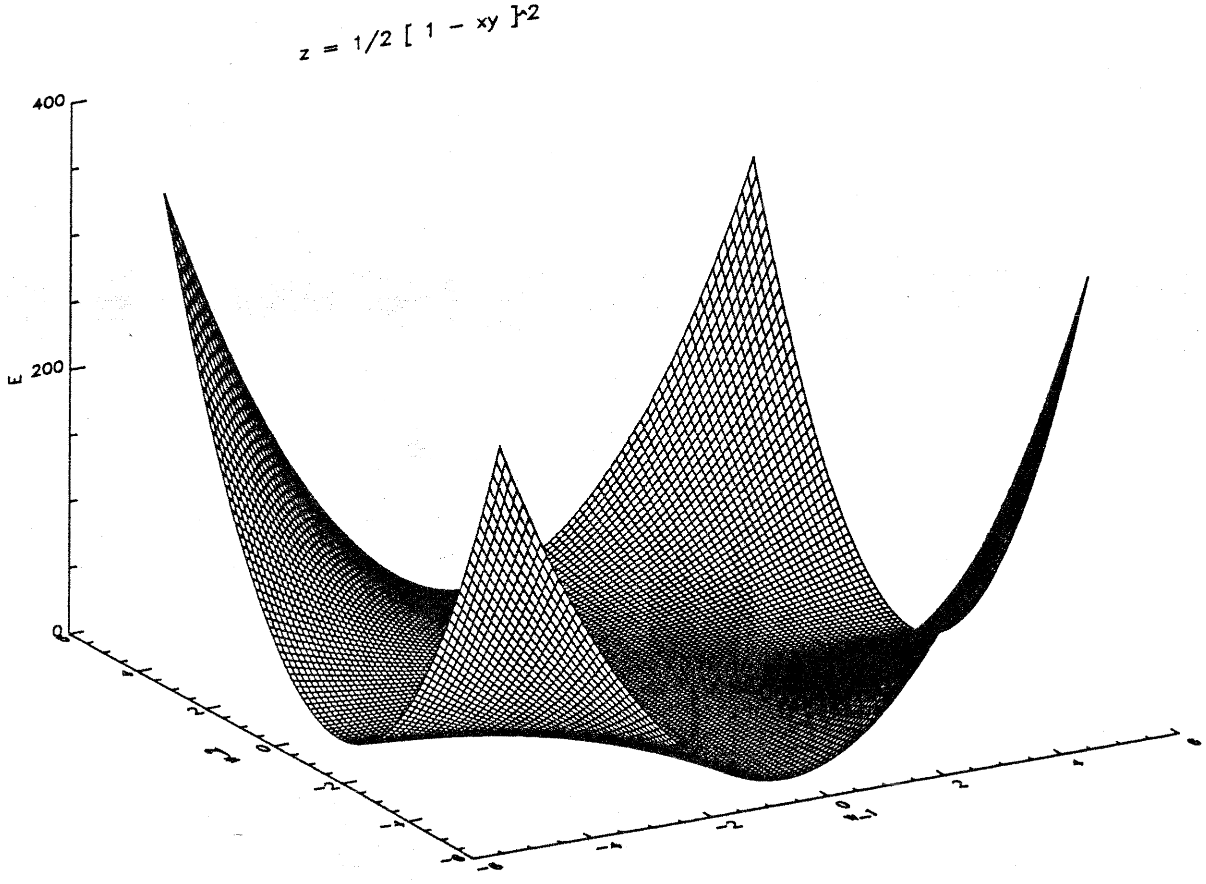
$a_3 = w_{32} a_2 = w_{32} \cdot w_{21} \cdot a_1$

$$E = \frac{1}{2}(1 - w_{32} \cdot w_{21} \cdot 1)^2 + \frac{1}{2}(0 - w_{32} \cdot w_{21} \cdot 0)^2 = \frac{1}{2}(1 - w_{32} w_{21})^2$$

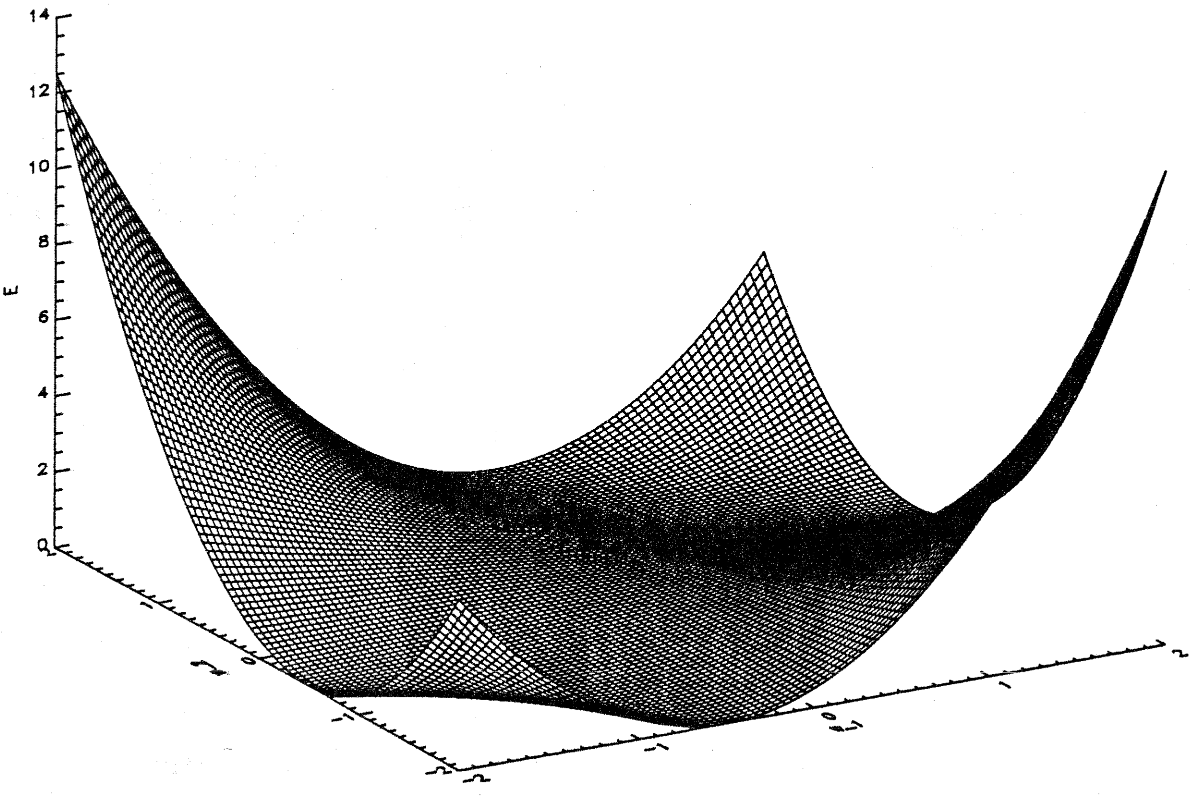
Start off at $\vec{w} = (0,0) \Rightarrow$???

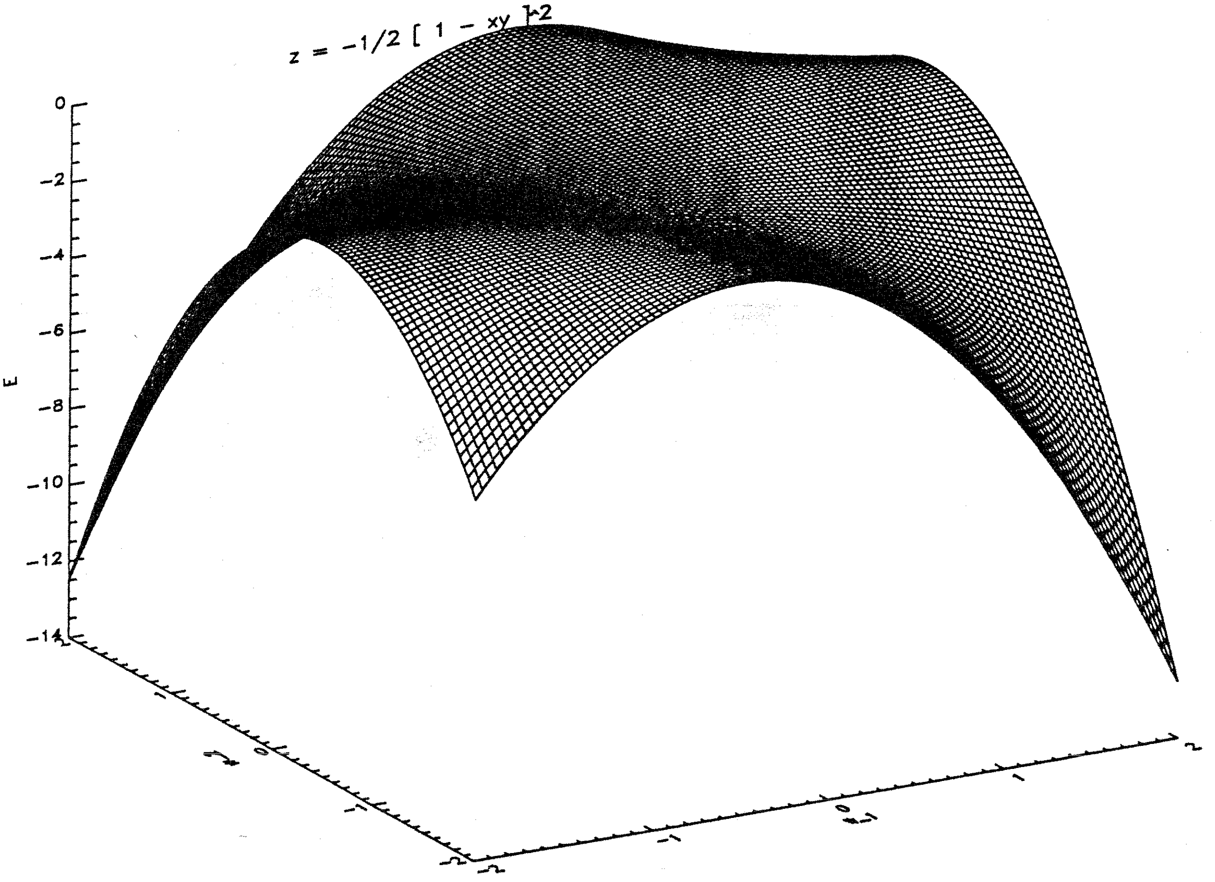


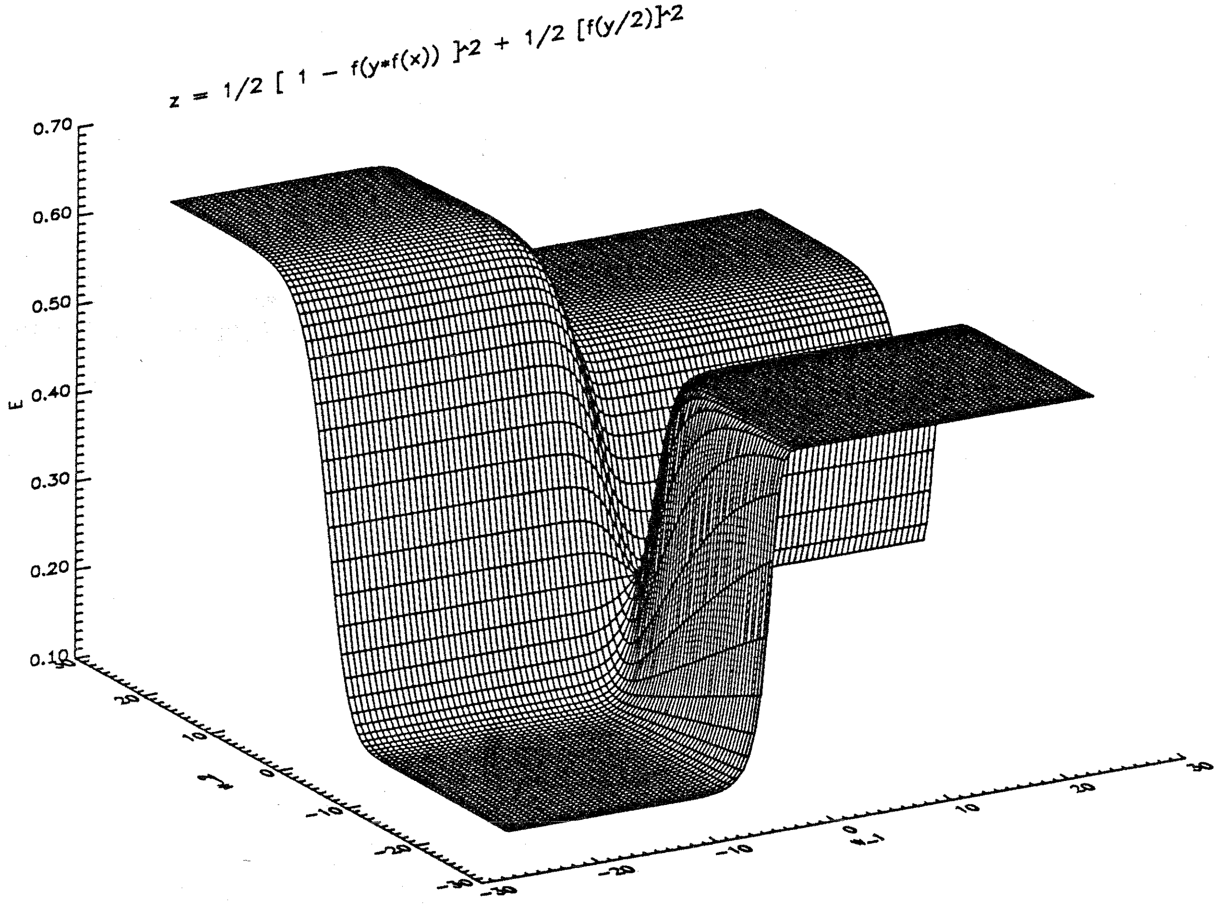


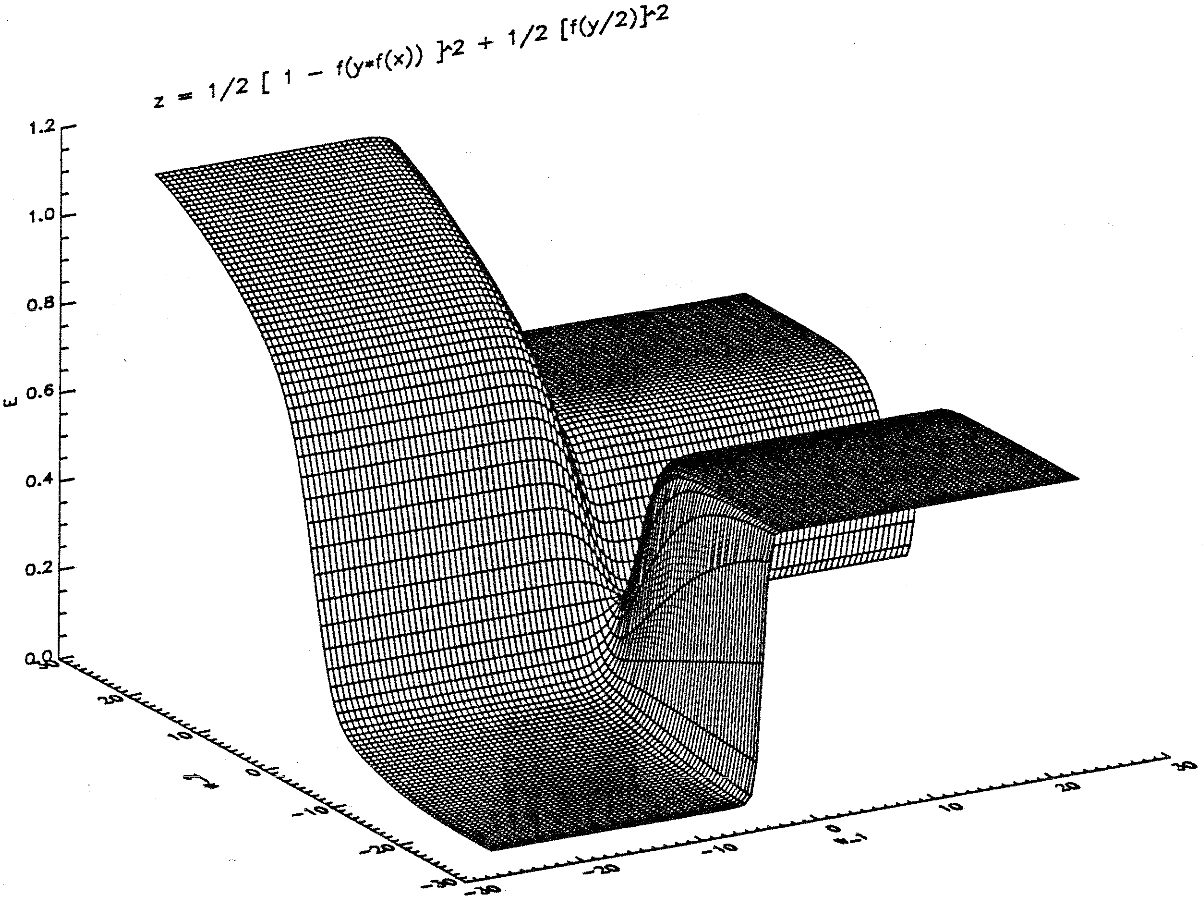

z = 1/2 [1 - xy]^2
400
E 200
0
6
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6

z = 1/2 [1 - xy]^2

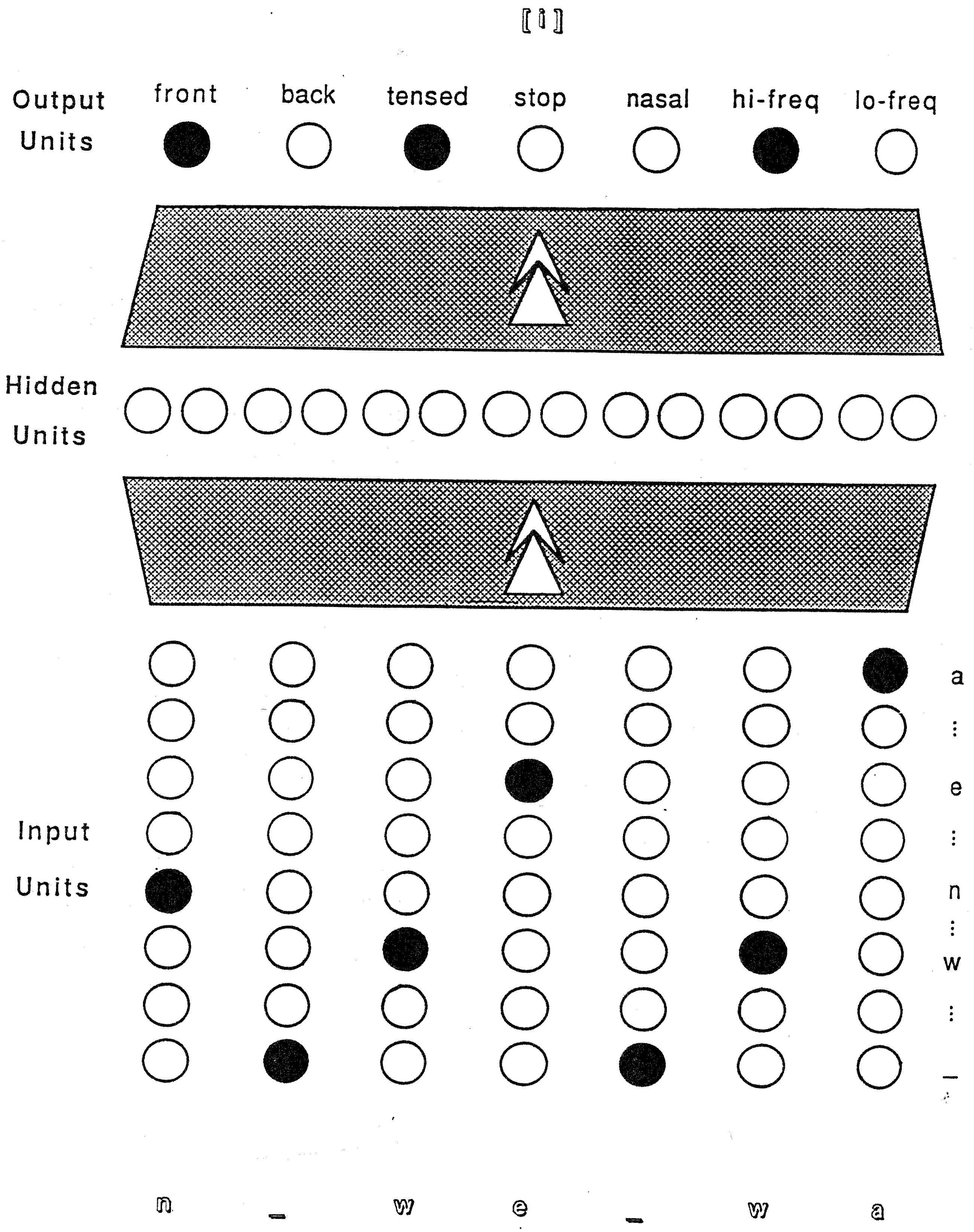



z = -1/2 [1 - xy]^2
E
0
-2
-4
-6
-8
-10
-12
-14
-2
-1
0
1
2
-2
-1
0
1
2


z = 1/2 [1 - f(y*f(x))]^2 + 1/2 [f(y/2)]^2
E
0.70
0.60
0.50
0.40
0.30
0.20
0.10
30
20
10
φ_2
0
-10
-20
-30
-30
-20
-10
0
φ_1
10
20
30


z = 1/2 [1 - f(y*f(x))]^2 + 1/2 [f(y/2)]^2
E
0.0
0.2
0.4
0.6
0.8
1.0
1.2
w_2
30
20
10
0
-10
-20
-30
w_1
-30
-20
-10
0
10
20
30

NetTalk: Sejnowski, T.J. & Rosenberg, C.R. '87



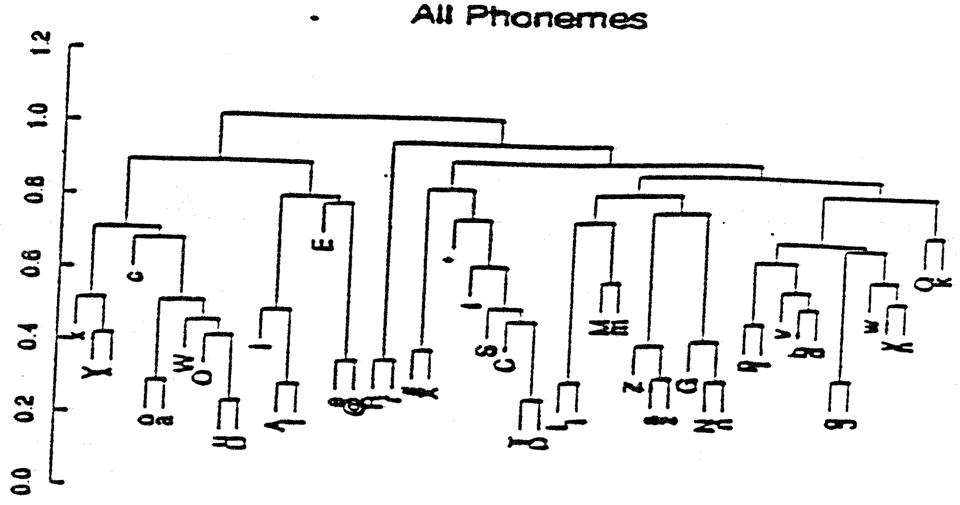

All Phonemes

Figure 2: Hierarchical clustering results for 48 phonemes.

FIGURE 6

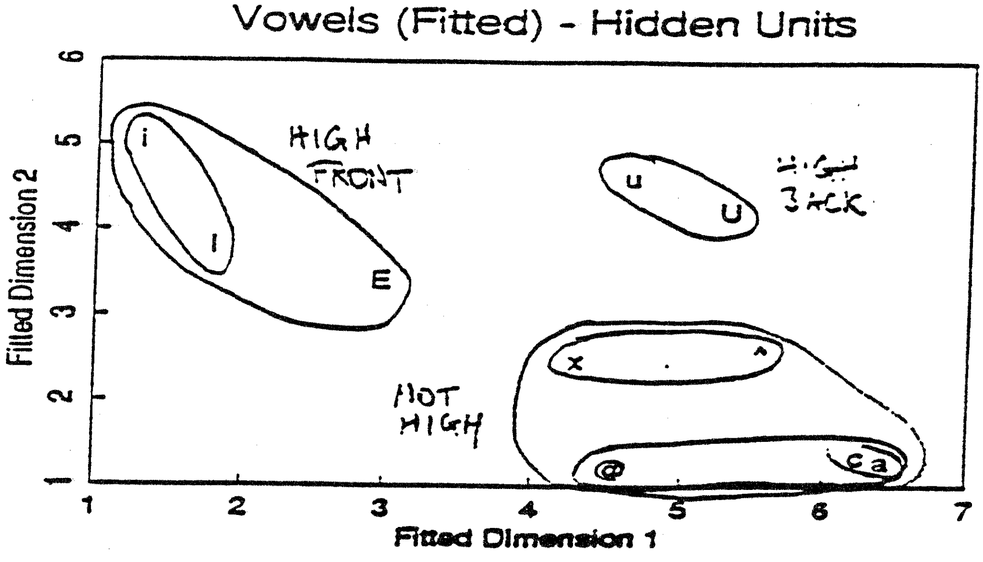

Vowels (Fitted) - Hidden Units

Figure 4: Fits based on linear regression of the three factor solution on the attribute dimensions "place of articulation" and "vowel height", plotted in Fig. 3. Data based on hidden unit activations. The groupings are based on a centroid hierarchical cluster analysis of the original correlation matrix.

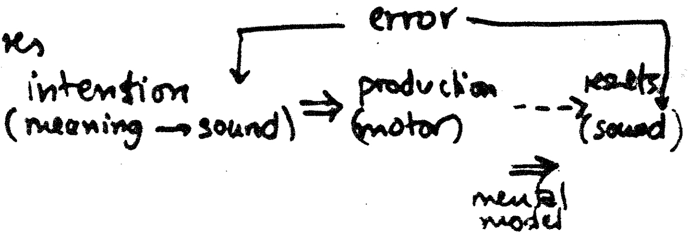
Rosenberg
Cog. Sci. '87

Connectionist Learning

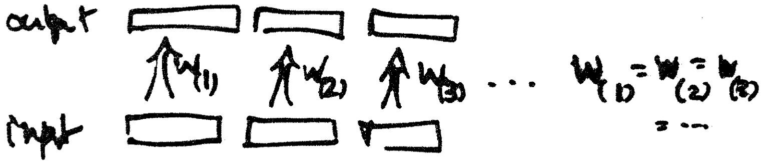
processing architecture
\+
E function
\+
minimization technique
} = connectionist learning algorithm

Extensions of back prop:

- More general processing architectures
  - distal targets Jordan:



- * constrained weights eg: copies for processing strings



- * recurrent networks
- Different E
  - Occam's razor - simplicity vs accuracy
  - particular kinds of target patterns eg: (0001000) for classificat

Alternatives to supervised learning: no 'targets', new E

- Reinforcement learning
- Unsupervised learning
- * 'Completion task': observe & predict, extracting structure

  No negative evidence

* → Elman

Strategies & Principles of Connectionist Applications in Linguistics: A summary

Low-level strategy

1. **Connectionist principle:**
Learning is some kind of statistical analysis of co-occurence.

Linguistic interpretation:
Temporal sequence of connectionist patterns → temporal sequence of linguistic units

Linguistic principle:
Linguistic categories can be induced from positive examples thru the prediction of the next item.

Example:
Inducing word boundaries, lexical categories: Elman

2. **Connectionist principle:**
Processing is activation-driven: The [continuous] activation of a unit is increased/decreased over time in proportion to the activation of its neighbors in the network.

Linguistic interpretation:
activation → feature value (e.g. accentual prominence, sonority, tone)
network neighbors → tier neighbors

Linguistic principle:
The [continuous] derived value of a feature is increased/decreased over time in proportion to the derived value of that feature for its left- and right-neighbors in the tier.

Example:
Goldsmith & Larson theory of stress & sonority

3. **Connectionist principle:**
Processing is similarity-driven.

Linguistic interpretation:
temporal sequences of activity patterns → segmental strings
connectionist unit → distinctive feature
similarity → sharing of distinctive features

Linguistic principle:
The tendency for a missing feature in segment *s* to be filled with a value consistent with that of a preceding segment *s'* is increased by
similarity between *s* and *s'*
similarity between *s'* and other segments near it
proximity in string/time of *s'* and *s*

Example:

Transparency in Hungarian vowel harmony: Hare.

Methodological aside:

Low-level strategy for application of connectionism to linguistics (radical):

Low-level connectionist networks [principles] → specific linguistic account/'model' [→ higher-level analysis (possibly)]

High-level strategy (conservative):

Low-level connectionist networks → high-level connectionist computational principles → linguistic principles defining a grammatical formalism → specific linguistic accounts

Note: This strategy is made possible only through technical innovations in the high-level analysis of connectionist computaiton.

High-level strategy

4. Connectionist principle:

Connectionist representations can be analyzed on a higher level as symbolic structures.

Linguistic interpretation:

Henceforth, all grammatical representations will be symbolic: traditional, unrevised. The rules too will be traditional in form, but their interpretation and interaction will be new. (Conservatism.)

5. Connectionist principle:

Processing (at both lower- and higher level) is harmony-driven.

Linguistic interpretation:

connectionist input → linguistic 'input' to grammar (string of words, phonemes)
connectionist well-formedness → linguistic well-formedness
pattern of activity arising from an input → structural description of the input
pattern of connections → grammar

Linguistic principle A:

Grammatical well-formedness is quantitatively measured by Harmony. The grammar assigns to any input the structural description that maximizes Harmony. This maximal Harmony value is the well-formedness of the structure: high for well-formed and low for ill-formed.

The terms in the Harmony function define a Harmonic Grammar, and may be interpreted as *soft rules* of the form:

If structural configuration X occurs in the structure, add H_X to the Harmony of the structure.

The numbers H_X can be automatically determined from the data via back propagation.

Note: Harmonic Grammar permits (but does not require) the study of **very complex interactions of multiple factors.**

Example:

Unaccusativity phenomena in French: Legendre, Miyata, Smolensky

Problem: Understanding the *grammar* is almost as difficult as understanding the original *data*...

Linguistic principle B:

> **Optimality.** A grammar is a *preference relation* $\succ$ among structural descriptions; it assigns to any input that structure which is *optimal* (prefered by $\succ$ to all other structures).
>
> The preference relation $\succ$ among structures is built up compositionally from preference relations among the substructures or subdimensions from which structures are built.
>
> The basic preference relations and their means of combination are drawn from a universal repertoire.

6. Connectionist principle:

Learning is error minimization.

Linguistic interpretation: ???
Example applications: ???

We chose four diagnostic contexts:

Object Raising (OR)	**La glace** est facile à faire *fondre* .	**Ice** *is easy to make* **melt.**
Croire "believe" (CR)	Je croyais **Marie** déjà *sortie.*	*I believed* **Marie** *to have already gone out.*
Participial Absolute (PA)	*Parti* avant l'aube, **Pierre** est arrivé à destination le jour même.	*Gone before dawn,* **Pierre** *arrived at his destination on the same day.*
Reduced Relatives (RR)	**La neige** *fondue* a formé de la boue.	*The* ***melted*** **snow** *formed mud.*

Specifying HNet' based on linguistic analysis ⟹

Data: 760 French sentences generated from the four diagnostic contexts, 143 intransitive verbs, and 3 types of arguments with their acceptabilities judged by human informants in 5 graded levels.

Choice of constituents:

Inputs: We chose the following to code each of the 760 French sentences:

(a) Argument: two semantic features - AN (animate) and VO (volitional).
(b) Verb: two aspectual features - TE (telic) and PR (progressive); 143 individuated verbs as lexicon.
(c) Context: four diagnostic contexts (OR, CR, PA, and RR)

Hidden: two hidden constituents, motivated linguistically by the unergative/unaccusative distinction.

Choice of constituent interactions:

175 independent quadratic terms representing pairwise constituent interactions. (143 lexical, 32 non-lexical)

E.g., we assume each diagnostic context can interact with (have preferences for) semantic features of the argument and aspectual features of the verb (include the terms $H_{OR,AN}$, $H_{CR,TE}$, etc.).

E.g., we assume each diagnostic context *cannot* interact with (have preferences for) individual verbs (eliminate terms $H_{OR,dire}$, $H_{CR,fondre}$, etc.)

Treatment of hidden constituents:

We assumed that the two hidden constituents are mutually exclusive: turn on one of the hidden units that maximizes **H** and turn off the other.

➡

Results

- **Examples of soft rules:**

  After training the **HNet'** on these data, we interpret the weights in **HNet'** as a set of soft rules:

  - verb *s'asseoir* "walk" prefers hidden constituent 2 - if hidden constituent 1 is chosen, decrease **H** (well-formedness) by 4.5.
  - verb *aller "go"* prefers hidden constituent 1 - if hidden constituent 1 is chosen, increase **H** by 3.4.

- **The performance:**

  As a collection of these soft rules, the grammar embodied in **HNet'** could account for the signs [+ or -] of acceptabilities for **all** of the 760 sentences, and for the graded acceptabilities [5 levels] for all but 30 of them.

- **Strong interactions among syntactic and semantic components:**

  The network achieved this performance with strong interactions among various syntactic and semantic components: almost every rule is strong enough to flip a decision between acceptable & unacceptable.

- **Extendability of the grammar:**

  We tested how well the grammar generated can extend to new verbs. After training on 536 sentences with 87 verbs, we gave the network 80 sentences with 20 new verbs. The network could account for the signs of acceptabilities of all but 2 of the 80 sentences by adding an appropriate new lexical entry for each verb.

- **Predictions:**

  The network was able to make certain predictions about acceptability patterns allowed for verbs and arguments with certain feature combinations. We are in the process of evaluating these predictions.

- Goal: Account for the acceptability judgements of 8393 Ss (3608 intransitives 4785 transitives)

  [an exploration in D.I.C.F.L.; for an exploration in T.I.C.F.L, see Prince & Smolensky, 1991]

Data:

- · Constructions (11): OR, CR, PA, PE, RR; ON, PCL; PRQ, FOR, PR, IMP
  - · Targets: Subj, D.O.
- · Embedded Predicate
  - · Identity 408 = 183 transitives + 225 intransitives
  - · Features Given aspectual features – TE, PR
    Learned (arbitrary) features – n = 6
- · Embedded Argument Structure
  - · Transitives : Subj, D.O. } semantic features: VO, AN, DEF
  - · Intransitives: Arg }
    deep G.R. Subj, D.O.
  - · Acceptability judgements: +, +?, ?, -?, -

Account:

- · Implemented as a local connectionist network
  - · Trained using backpropagation on judgements
  - · Current best account: of 8393 Ss:

    104 errors (1.2%) = 90 transitive (2.5%) + 14 intransitive (.4%)

Account:

- As a set of Harmonic Grammar rules:

  If construction = CR & target = D.O. & deepG.R. = D.O.
  then add .55 to Harmony = well-formedness

  If predicate = mourir 'die' & deep G.R. = Subj
  then subtract 2.6 from Harmony

  ⋮

  If construction = CR & target = D.O. & Subj = + VO
  then subtract 1.5 from Harmony

  · interacting via the

  Harmonic Structural Assignment Principle:
  assign the structure that maximizes Harmony

- As the higher-level description of a lower-level connectionist network obeying certain general principles of connectionist representation & processing (in particular, pattern completion via Harmony maximization)

  Basic linking assumption:

  linguistic well-formedness = connectionist well-formedness
  = Harmony

Az Phonology Conference
April, 1991
Tucson

OPTIMALITY

Alan Prince
Center for Complex Systems
Brandeis University

Paul Smolensky
Center for Opto-electronic Computing
University of Colorado

(1) **Optimization in Grammar Choice** (Chomsky, 1951,1965)

UG(D) => G_1,G2,...... All G_i's agree on D.
Eval { G^i } −> G* G* contains the real rules of L.

Chomsky's idea was to find a coding for the operations and structures of grammar such that the length (i.e. shortness) of the encoding of G was the measure of G's value.
—For phonology, we must determine Eval(Lexicon)+ Eval (Rules).
Otherwise, we can drive Eval(Rules) to optimum by no Rules.

(2) **Optimization as a grammatical mechanism.**
Simple Case:
Do-α (IN) −> OUT_1, OUT_2,.....
WF { OUT_i } −> OUT* The real output.
More generally,
B(IN, Do-α, OUT*) > B (IN,Do-α, OUT_j) "Greatest Benefit."

(3) **The rap against optimization.**
a. Have to use numbers — parameter city; and use counting.
b. Many messy tradeoffs if t > k , a > i, what then of ti vs. ka?
c. Optimization is computationally intractable, in general.

(4) **Counter-rap.**
c'. It is not incumbent upon grammar to compute. Well-definition is all; not efficiency or even algorithmicity (Chomsky).
a'. Actually ORDER, not counting, is the key.
b'. We will argue that many cases of apparent tradeoff can be handled without tradeoff calculation, by a principle of combining orders.

(5) **Optimization in a grammatical mechanism**: the case of Berber Syllabification (Dell & El-Medlaoui 1987, 1988)

(6) ANY segment may be syllabic:

rA.tK.tI yA.rB.bI tF.tKt A.tXk txZ.nAkkw tzMt
tM.zh tR.gLt tI.dI tR.bA

(7) ONSETS: syllables must have onsets, except phrase-initially.

(8) SONORITY: syllabification is sensitive to relative sonority.
a. tzMt — *tZmt vs. tM.zh — *tmZh
b. rat.lult — *ra.tL.wLt ie. lul — *LwL

c. D&E establish the following 8-point hierarchy:

a > ui > rl > mn > z > fsx > bdg > ptk

low V > high V > liq > nasal > voiced fric > v'less fric > voiced stop > v'less stop

(9) **D&E's Algorithm** (DEA)

Form core syllables (X)Y ("CV"), where X is any segment and Y is successively replaced by the features describing the steps of the sonority scale; in descending order.
Iterate from Left to Right for each fixing of the nuclear variable Y.

(10) Secondary phenomena (to be ignored):
a. obstruents de-syllabified #_ and _#.
b. sonorant C's optionally desyllabify _#.
c. tautomorphemic geminates never begin an onset. # + m m

(11) EXAMPLES:

/ratlult/	−>	{rA}tlult	−>	{rA}t{lU}lt	−> {rA}t{lU}{lT}
/t-xzn-t/	−>	tx{zN}t	−>	{tX}{zN}t	'thou stored'
/t-xzn-a-s/	−>	txz{nA}t	−>	t{xZ}{nA}t	'she stored for him'

(12) DEA serves to *optimize* the quality of syllables: high sonority items like to be nuclei.

(13) But DEA *per se* makes no contact with this. E.g. Why *descend* the scale?

(14) **Harmonic Syllabification** (HS).
Form the best core syllable from free material in the domain, good (σ) = good (nuc).
Iterate until you can do more, H$ (Σ) = Σ .

(15) **This is a fully 'Harmonic' process**: ('Harmony' from Smolensky, 1986.)
Go to the most well-formed available state.

(16) Note that HS is different from DEA. HS has no directional bias, does not sensibly run in a direction (LR) to deal with sonority plateaux.

(17) Evidence for directionality weak anyway.

a. /kiut/	kiwt, *kyut	but better onset (DE note)
b. /rksx/	R.kSx, *Rk.sX	but no final syllabic obstruent.
c. /bayn-n/	ba.yNn, *bay.nN	but no geminate qua CV.
d. /ugmn/	u.gMn, ug.mN	both ok. DEA predicts only the first.
e./itbdrin/	*i.tBd.rin, only it.bD.rin	*DEA fails*. Nb. closed σ bad.

(18) Conclusion: optimization, with grades of well-formedness (representational Harmony), is a mechanism of grammar.

(19) To make this approach fly, we need two things:
a. Ways of evaluating combinations of phonological dimensions where the dimensions themselves needn't have an internal wf'dness structure, but may be scalar. Harmony scale on dim. comb. [post-Praguean contextual markedness: SPE, Kean,....,Archangeli & Pulleyblank.]

b. Ways of combining different relevant dimensional scales. to make Harmony scale.
Ways of combining Harmony Scales along different dimensions.

(20) **Dimensional Combination.** Example: weight and stress. (Prince, 1990)
a. Separate prominence (not markedness) scales: H>L, stress > unstress.
b. Combine implicatively: if H, then stressed. (WSP = =Hayes's 'quantity sensitivity').
Good are: H', L^. Bad is: H^. [unstressed heavy]
c. Weaker implication: if stressed, then Heavy.
Good are H', L^. Bad is: L'. [stressed light]
d. Combination respecting strength of implications:
H',L^ > L' > H^ . This is a Harmonic ordering, expressing degree of concord of the two dimensions.

(21) **Combination of Harmony Scales.** Example: 'unbounded' stress systems like: Stress the rightmost heavy syllable, if no heavy then stress the rightmost syllable (R/R). Prince (1983) and Hayes (1991) see these as enhancements of prominence rather than 'unbounded foot' phenomena. We extend this result, thereby limiting feet to the authentic binary rhythmic units.

(22) **The pattern R/R.** Two factors:
a. WEIGHT. Heavy syllables should be stressed.
b. POSITION. Edgemostness: Stress should be near the end.

(23) WEIGHT dominates POSITION. WEIGHT >> POSN. (Nb. One stress.)

(24) WEIGHT must be satisfied, if possible. But there can be several choices:
a. cv.CVV.cv.cv.cvv.cv.cv.
b. cv.cvv.cv.cv.CVV.cv.cv.

(25) Indeterminacies are resolved then by POSN. (24b) is chosen. ↔ bar.

(26) No indeterminacy in (cv)* = ...cv.cv.CV.

(27) This is **LEXICOGRAPHIC ORDER** — a generalization of alphabetic order. Alphabetically, strings are ordered first by initial character (= =WEIGHT), then ties are broken by the ordering on the second character (= =POSN), etc. ptrs > qaaaaaaa. NB:NO TRADEOFF

(28) **LEXICOGRAPHIC PRODUCT** of two orders,
P *>> Q,
total order on the the pairs (p,q) constructed from the orders on P and on Q; first ordering by P-dimension, then by Q-dimension.

(29) Example: alphabetic ordering itself: POSN *>> ALPHA.
(1,2,3,4,....) *>> (a,b,c,...)
1a > 1b > 1c > ....2a > 2b > 2c....

(30) **Claim:** Lexicographic Order is the essential mode of combining Harmony scales.

(31) Berber Syllabification is also a kind of Lexicographic Order: analyses in which /m/'s are nuclei are absolutely preferred to competitors in which /m/'s are not but /z/'s are.
We can say:
x's should be peaks >>y's should be peaks, for x > y in sonority.
With suitable technical development, this might replace the serial (iterative) HS.

(32) **Prominence-driven Stress.** Claim: works by Lexicographic ordering PROMINENCE >>POSN, with development of notions of prominence and position.

(33) **Hindi** (modern) (Kelkar, 1968; Hayes, 1991)
"Stress falls on the heaviest available syllable, and in the event of a tie, the rightmost nonfinal candidate wins." (Hayes)

(34) Heaviness: CvvC, CvCC ($\mu\mu\mu$) > Cvv, CvC ($\mu\mu$) > Cv (μ)

(35) Heaviest is best:
kiDHAR jaNAAB as.BAAB ru.pi.AA REEZ.gaa.rii

(36) Among equals, last nonfinal is best:
[μ]⊢ sa.MI.ti
[$\mu\mu$]⊢ ru.KAA.yaa PUS.ta.kee roo.ZAA.naa
[$\mu\mu\mu$]⊢ AAS.maa.jaah aas.MAAN.jaah

(37) Analyis: WEIGHT >>POSN
But POSN is constructed from NONFINALITY, and RTMOSTNESS.

(38) NONFINALITY >>RTMOSTNESS

(39) WEIGHT >>NONFINALITY >>RTMOSTNESS.

(40) **Piraha.** (Everett & Everett,84; Everett.,88; Levin, 85; Davis, 88; Hammond, 90; Hayes, 91).

"Stress the rightmost token of the heaviest syllable type in the last 3 syllables of the word."
(41) Parallels: 'rightmost heavy or rightmost syllable'. 'HEAVINESS' >>EDGEMOSTNESS

(42) What is 'Heaviest'? C = voiceless, G = voiced cons.

a. ?A.ba.gi	'toucan'	Cv>Gv
b. ?a.ba.PA	'Amapá'	rtmost Cv wins
c. so.AI.pi	'nose'	vv > Cv
d. KAA.gai	'word'	Cvv>Gvv
e. poo.GAI.hi.ai	'banana'	Gvv>vv ; in 3-σ window; nb poo.
f. ka.pii.ga.ka.ga.kai.ka.bao.BAO		rtmost: in window: nb. kai

(43) **Full 'Heaviness' Hierarchy:**

CVV > GVV > VV > CV > GV > V , where C=voiceless, G = voiced.

(44) Evidence that the stress is there:
a. Some speakers treat it as a beat wrt hand-gestures.
b. Some speakers devoice everything after stressed syllable.
c. Some speakers delete everything after stressed syllable.
d. May be some tropism of tone to stress.

(44) ISSUES:
A) How is 3 syllable limitation established?
B) How is 'heaviness' hierarchy established?

(45) Ans. to (A). *Domain of Stress is Minimal Prosodic Word, minWd.*
(Cf. Prosodic Circumscription (McCarthy & Prince, 1990), whereby a prosodically delimited subdomain serves, in lieu of the expected morphological category, as the domain of a process.)

(46) *What's minimal about 3?* Prosodic Hierarchy says Wd contains F contains σ. So, hierarchically-locally, minWd contains ONE F. Maximal such Wd = F+σ (or indeed σ + F).

(47) Piraha stress must lie in maximal minWd, for F = $\sigma\sigma$.

(48) Evidence for involvement of Wd category:
Closely-knit phrases show ONLY ONE STRESS when last word is 2 σ.

bao.sai # bii.SAI	'cloth # red'
?a.ba.pa # go.GI	'city-name # where-at'
BUT:	
?i.sii.HOA. # ai.BAI.?i	'liquid fuel # much'
?a.pa.PAI # ?II.ta.ha	'head # hurts'
ka.H(A)I # ?o.ga.ba.GAI	'arrow # want'

Analysis: PWd = max minWd is established at]-edge of morph-word.
(Cf. Chen, Selkirk, Hewitt). Note RL orientation: final wins.

(49) PWd Parses. —Stress falls inside PWd. Must have }# .

bao.{sai # bii.SAI}#
{?a.pa.**PAI**}# {?II.ta.ha}#
{ka.H(A)I}# ?o.{ga.ba.GAI}#

(50) ANS. to (44B). 'Heaviness' Hierarchy.
CVV > GVV > VV > CV > GV > V
a. xVV > yV in every case.
b. C > G > nil This is just onset goodness! (R. Shafer).

Prominence Order = Weight >>Onset

(51) Full Story. IN-PWD>> WEIGHT >>ONSET >>RIGHTMOSTNESS.

(52) **Maithili** (Jhaa, 1958; Hayes, 1986, 1991)
Stress penultimate, final, antepenultimate, dependent on foot structure AND VV location.

(53) In words with only SHORT VOWELS, stress is penultimate:

SAta	pa.TIta dhana.HEra	bhina.SEra
ja.han.NEma	sun.NARA	

This is actually a Moraic-Trochaic (RL) pattern (Hayes)
— weak vowels in open syllables reduce; those in closed do not; nb. no closed final syllables.

(54) LONG VOWEL FINAL: Stress is FINAL for SHORT penult V—

a.dina.TAA gora.min.TII eka.hat.TAA

LONG V in antepenultimate stressed if VV followed by two LIGHT:

ji.muta.BAA.hana kaka.ROO.hari BAA.sana

PENULT LONG V is stressed:

pu.RUU.kha SAA.bhaa ba.raha.MAA.saa

(55) ANALYSIS: Main Stress falls on MOST PROMINENT in final circumscribed window; when there is equality, penult is favored (Non-finality).

(56) Window: 2 units = FF, Fσ, σF. = 2F, if headless feet exist (H86)

(57) **Elementary Prominence Orders:**
FOOT: Foot-heads > Nonheads;
VV: VV > V

(58) Relation: VV >>FOOT

(59) As usual, PROMINENCE >> POSN,
for POSN = NONFINALITY >> RIGHTMOSTNESS

(60) Cases:

...|CVcv.CVcv|	= =prom/rightmost Fhead.	dhana.HEra
...|CVC.CVcv|	ditto	ja.han.NEma
cvv.|cv.CVV|	prom wins.	ku(u).ra.NAA
...|CVV.CVcv|	prom wins.	BAA.sana
....|CVV.cvv|	= =prom/nonfinality wins.	SAA.bhaa

(61) Fact Note: Following Hayes, we assume ...cvv.CVC.cv#
galaa + benna —> galaBENna but CAA.rillo 'fourth'
VV shortens in nonfinal, non-mainstressed syllables. If wrong, then PROM = heavy.

(62) **Awadhi** (Saksena, 1971; Hayes, 1991). Exercise for auditor: Stress is penultimate, except that in light-heavy#, heavy is stressed.

(63) **Broad Vista.** To run grammar on optimization, we need to be able to say WHEN to perform an operation. I.e. when it is better to act than to do nothing:
B(X—>Y) > B (X —> X)

(64) **Basic Harmonic Condition:** H(Y) > H(X). Well-formedness increasing: up Harmony Hill. But not sufficient: else Wd —> Tata.

(65) From universal perspective, need to order Transitions and combinations of Transitions and contexts to obtain markedness implications about processes.

(66) Example. **LANGO** Vowel Harmony — one property thereof. (Pulleyblank, Yest.)
a. ATR Harmony spreads forward from ALL vowels over a SINGLE CONSONANT: in V_1-C-V_2 , V_2 harmonizes with 1st vowel.

b. But over TWO intervening C's, the trigger V must be high:
only in hi-CC-V do we get spread of ATR.

(67) Appears to involve complex **TRADE-OFF** in virtues of spreading from various sources, and difficultly of doing same in various environments. But there is NO TRADE-OFF.

(68) Approach. Suppose we know the following:
a. Spread ATR from HI > Spread ATR from NonHI. "HEIGHT"
b. Spread over single C > Spread over CC. "DISTANCE"

(69) Form the Lexicographic Product ~~[illegible]~~ HEIGHT *>> DISTANCE: (or vice versa)

Spread from HI over C > Spread from H over CC
> Spread from NonHi over C > Spread from Non-Hi over CC.
Lango: do nothing.

(70) Now, an individual language may insert DO-NOTHING anywhere in the order.
All things preferable will be done, then.

(71) **Insert DO-Nothing** at bottom: any action is better than inaction: no constraint, full ATR harmony.

(72) **Lango.** Insert Do-Nothing above last element in order.
Spreading-from-NonHi-over-CC is worse than Nothing:
But everything else is better. So act accordingly.

###

Constraint Interaction and Harmony Maximization in Grammar

Alan Prince
Brandeis University

Paul Smolensky
University of Colorado

I. Basic Idea -- General

A. Picks up two themes in modern linguistic thought (see attached bibliography; most of the entries are concerned with these):

1. Constraints or processes that give way or are conditional.

EG.

a. Do such-and-such EXCEPT when the result conflicts with some other constraint. Eg. Make final σ extrametrical EXCEPT in monosyllables. Eg. Foot form.

b. Do such-and-such ONLY when necessary to achieve wellformedness. Eg. Epenthesize a glottal stop ONLY to provide syllable with onset. Epenthesize a Vowel ONLY to make word big enough to hold foot.

2. Direct application of universal principles of markedness in the operation of individual grammars.

EG. Prefer syllables to have onsets. Prefer the collocation of +Hi and +Atr. In the standard theory (eg. SPE, ch. 1-8, et seq.), these are no more than observations made about what grammatical processes happen to end up doing. But they can form the foundation for principled theory of grammar itself; can be directly appealed to in specific grammars.

B. Rapport with mode of functioning of central species of connectionist networks: harmony increase to maximum.

C. Fundamental Harmonic Imperative: among the possibilities, choose the one that maximizes harmony (relative wf'edness).

II. What we need to carry out this project.

-Notion of preferences (approx = markedness)

-Ways of combining orders. (Lexico. Direct.)

-Way of assessing harmony of structures, as determined by constraint order. (LOCS. Tableaux)

III. Constraint Domination. Some familiar basic types of constraint interactions.

A. Do something EXCEPT. A positively-regarded option is embedded lowdown.

1. Extrametricality: méns, *<mens> LexWD=PrWD >> EXM.

LexWD=PrWD. Lexical Words should be Prosodic Words, therefore contain feet.

EXM. Final syllables should be foot-loose.

2. Imbrication in Cibemba (Hyman, 1991):

Short Base	Long Base
fik-il -e	*matik-il-e
*fi-l-t -e	mati-l-k -e

Generalization. Stems CvC suffix il; Longer bases insert l after last vowel in bases.
Analysis. Take l to be default C. Then suffix /l/ to PrWd.

FB >> PrWdWF

FB. Feet should be binary at some level of analysis (σσ, μμ).
PrWdWF. PrWd words should be consist exhaustively of well-formed units.

3. Foot WFedness. (#égo:#, *#é#, *#e#)

FB >> LexWD=PrWD >> EXM >> PARSE >> RH

RH. Rhythmic Harmony. Trochaic feet LL, H are better than LH.
PARSE. All material should be parsed into prosodic structure. (= Itô's Prosodic Licensing).

Say Stray Erasure en route to the Phonetics.

4. Infixation 1: gr-um-adwet. *umgradwet. *gumradwet.
Goodfirstsyll >> Leftmostness of prefix

Goodfirstsyll. No Coda.
Leftmostness of Prefix. 'Prefix' means affixal content is Leftmost in Affix + Stem collocation.

5. Infixation 2: metgo-go-t, *metgot-go.
Edge-preserve >> Rightmostness of suffix (light syll).

Edge-preserve. Segment at edge of base should appear at edge of reduplicated base (pa-prath)

Rightmostness of suff. 'Suffix' means affixal content is Rightmost in Affix + Stem collocation.

B. Do something ONLY. A rejectionist option is lowdown & overruled by a higher-order cstrt — that is, the high-order constraint can force the disfavored option.

HAVE-ONSET >> FILL-NODES

HAVE-ONSET = σ should dominate ON constituent.
'Epenthesis' = Insert empty syllable structure (Selkirk, Itô).
FILL-NODES = nodes should dominate segments = AVOID epenthesis.

Universal Typology in Harmonic Phonology: 0^{th}-order Syllable Structure

Paul Smolensky & Alan Prince

July 25, 1991

Fundamental Definitions

LOCS:

Suppose given:

- a *role-filler decomposition* of S specifying for each $s \in S$ a description of s as a multi-set M_s of role/filler pairs
- an ordering $\sqsupset$ of fillers
- a function *predom* specifying a *predominant element* of any multi-set M_s of role/filler pairs; define:

$$residue(M) = M \setminus predom(M)$$

Then the corresponding *Lexical Order on Composite Structures S* is the ordering on S defined recursively by:

$M_s \succ M_{s'}$ iff

$$filler(predom(M_s)) \sqsupset filler(predom(M_{s'}))$$

or

$$filler(predom(M_s)) = filler(predom(M_{s'})) \text{ and } residue(M_s) \succ residue(M_{s'})$$

Two cases of interest:

Filler-driven LOCS: $predom(M_s)$ = any role/filler pair in M_s containing a maximal filler (w.r.t. $\sqsupset$).

Role-driven LOCS: $predom(M)$ = a role/filler pair in M_s containing a maximal role, w.r.t. an ordering on roles $\supset$ which must be further supplied

SDG: The **Strict Domination Grammar**

$$p_1 \gg p_2 \gg \ldots \gg p_n$$

is the preference relation $\succ$ defined from a sequence $p_1, \ldots, p_n$ of Boolean predicates over possible structural descriptions, defined as the following role-driven LOCS:

- the 'roles' are the predicates $p_1, \ldots, p_n$ and the 'filler' of each role is the truth value of each predicate: $M_s = \{\pm p_1, \pm p_2, \ldots.\}$ where the sign of p_k indicates whether s satisfies predicate p_k (+) or not (−). [We sometimes write M_s more mnemonically by suppressing the $+p_k$ and writing the $-p_k$ as $*\,p_k$]
- the ordering on fillers is $+ \sqsupset -$ (T $\sqsupset$ F; 'preferences')
- the ordering on roles is $p_1 \supset \ldots \supset p_n$

[Aside: If in the definition of SDG we change the LOCS from role- to filler-driven, then the order we get is 'Prefer fewer marks', where a 'mark' is a violated preference.]

FUT(*P*): The *Factorial Universal Typology* determined by a set P of universal preferences is the typology arising from the strict domination grammars determined all possible permutations of the preferences in P.

[That is, as far as the grammatical module governed by the set P is concerned, FUT(P) characterizes the set of possible languages as exactly those arising from all possible orderings of the preferences in P.]

0^{th}-order (C/V) Syllable Structure

Goal:

To explain the typology:

onsets: required/optional/ * forbidden;
codas: * required/optional/forbidden

[nuclei: assume required by definition of syllable]

		onsets	
		required	optional
codas	forbidden	Σ_{CV}	$\Sigma_{(C)V}$
	optional	$\Sigma_{CV(C)}$	$\Sigma_{(C)V(C)}$

Main Claim:

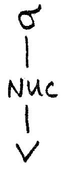
The universal typology of C/V syllable structures is FUT(*P*), where the set of preferences *P* is:

ONS: Prefer syllables with onsets.
-COD: Prefer syllables with no coda.
PARSE: Prefer parsed segments.
FILL: Prefer filled syllable positions.

Note: Best syllable is CV., which satisfies all preferences. Thus /CV/ is assigned CV. So it is not possible to prohibit onsets or to require codas.

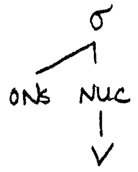
Onsets (ignore -COD for now):

Look at underlying form with preferred coda (i.e., none) and an onset wanting repair (i.e., none): /V/; Possible structures:



V.
{ * ONS}

⟦V⟧
{ * PARSE}



^V.
{ * FILL}

Phonetic assumptions: unparsed segments are not realized (deletion); unfilled syllable positions are somehow realized (epenthesis).

The preferred structure is the one with the lowest mark; given an order

$p_1 \gg p_2 \gg p_3$,

the preferred structure is the one having { * p_3}; the order of p_1 & p_2 don't matter. So rather than 3! = 6 cases there are 3 different ones (ignore {} for now):

1a. ONS ≫ {PARSE ≫ FILL} ~ PARSE ≫ ONS ≫ FILL   Preferred: ^V.
1b. ONS ≫ {FILL ≫ PARSE} ~ FILL ≫ ONS ≫ PARSE   Preferred: ⟦V⟧
2. {PARSE ≫ FILL} ≫ ONS ~ {FILL ≫ PARSE} ≫ ONS   Preferred: V.

(Here, ~ denotes the equivalence of two orderings, that is, that they determine the same maximally harmonious forms.)

Using the first instance of each case, we can view the three cases hierarchically by letting

1-1: Prefer one-to-one pairing of segments and syllable positions
= either {PARSE $\gg$ FILL} or {FILL $\gg$ PARSE}

we get:

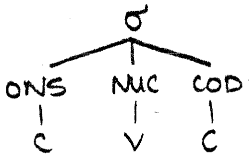
1. ONS $\gg$ 1-1
    - a. 1-1 = {PARSE $\gg$ FILL} Preferred: ^V.
    - b. 1-1 = {FILL $\gg$ PARSE} Preferred: ⟦V⟧
2. 1-1 $\gg$ ONS
    - 1-1 = {PARSE $\gg$ FILL} ~ 1-1 = {FILL $\gg$ PARSE} Preferred: V.

That is:

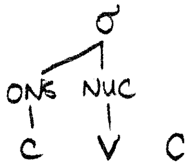
1. Onsets are mandatory, since ONS is the predominant preference
    - a. an onset is epenthesized;
    - b. an onsetless nucleus is deleted.
2. Onsets are optional, since it is more important not to epenthesize or delete (1-1) than it is to have an onset.

Codas:

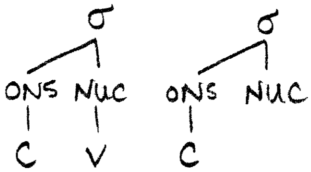
Same story, *mutatis mutandis*. Look at /CVC/ (preferred onset, coda needing repair).
Possible structures:



CVC.
{ * -COD}



CV.⟦C⟧
{ * PARSE}



CV.C^.
{ * FILL}

Cases:

1. -COD $\gg$ 1-1
    - a. 1-1 = {PARSE $\gg$ FILL} Preferred: CV.C^.
    - b. 1-1 = {FILL $\gg$ PARSE} Preferred: CV.⟦C⟧
2. 1-1 $\gg$ -COD
    - 1-1 = {PARSE $\gg$ FILL} ~ 1-1 = {FILL $\gg$ PARSE} Preferred: CVC.

That is:

1. Codas are prohibited, since -COD is the predominant preference
    - a. an extra nucleus is epenthesized;
    - b. a coda is deleted.
2. Codas are optional, since it is more important not to epenthesize or delete (1-1) than it is to lack a coda.

Note: For the onset discussion, the relevant case of FILL is 'Prefer filled onsets' (FILL_{ONS}), while for the coda discussion, it is 'Prefer filled nuclei' (FILL_{NUC}). These need not be assumed to be subsumed under a single general preference, FILL. Similarly, the relevant cases of PARSE are PARSE_{V} and PARSE_{C} for the onset and coda discussions, respectively. If there is but one general preference FILL and another such PARSE, then these general preferences must be relatively ranked once and for all in a given grammar, predicting that, $\text{PARSE}_{\text{V}} \gg \text{FILL}_{\text{ONS}}$ iff $\text{PARSE}_{\text{C}} \gg \text{FILL}_{\text{NUC}}$, which in turn implies that in any Σ_{CV} language, either onsets and nuclei are both epenthesized (/V/ → ^V. and /CVC/ → CV.C^.) or nuclei and codas are both deleted (/V/ → ⟦V⟧ and /CVC/ → CV.⟦C⟧). Presumably this is empirically false, so we must allow the two kinds of PARSE and two kinds of FILL preferences to each have their own place in the dominance hierarchy.

LARDIL — Hale, 1973; Wilkinson, 1988

	uninflected	nonfuture	future	gloss	
	(A)	/-in/	/-uṛ/		
	kentapal	kenṭapal-in	kentapal-uṛ	dugong	
	keṭar	keṭar-in	keṭar-uṛ	river	
	miyaṛ	miyaṛ-in	miyaṛ-uṛ	spear	
	yaṛput	yaṛput'-in	yaṛpuṭ-uṛ	snake, bird	
	yaraman	yaraman-in	yaraman-kuṛ	horse	∅ → k / N — uṛ
	pirŋen	pirŋen-in	pirŋen-kuṛ	woman	
	(B)		V → ∅/V—		
	mela	mela-n	mela-ṛ	sea	lowers
—a	wanka	wanka-n	wanka-ṛ	arm	≈ [illegible]
	kuŋka	kuŋka-n	kuŋka-ṛ	groin	
	ŋuka ŋuku	ŋuku-n	ŋuku-ṛ	water	— ?
-u/→a	kaṭa kaṭu	kaṭu-n	kaṭu-ṛ	child	
	ŋawa ŋawu	ŋawu-n	ŋawu-ṛ	wife	
	keṇṭe kenti	keṇṭi-n	keṇṭi-wuṛ	wife	ŋawi [illegible] — [illegible]
— i → e	ŋiṇe ŋini	ŋiṇi-n	ŋiṇi-wuṛ	skin	
	pape papi	papi-n	papi-wuṛ	father's mother	
	t'empe	t'empe-n	t'empe-ṛ	mother's father	
—e	wiṭe	wiṭe-n	wiṭe-ṛ	interior	
	(C) "NOT [illegible] CONSONANT" in 2nd C				
	yalulu	yalulun	yaluluṛ	flame	
	mayara	mayaran	mayaraṛ	rainbow	
	wiwala	wiwalan	wiwalaṛ	bush mango	
	karikari	karikarin	karikariwuṛ	butter-fish	
	yiliyili	yiliyilin	yiliyiliwuṛ	oyster (sp.)	
	(D)		NB		
—i	ṭuraraŋ[illegible]	ṭuraraŋin	ṭurarŋkuṛ	shark	
	ŋaluk[illegible]	ŋalukin	ŋalukuṛ	story	
—a	putuka	putukan	putukaṛ	short	
	murkunima	murkuniman	murkunimaṛ	nullah	
—u	ŋawuŋawu	ŋawuŋawun	ŋawuŋawuṛ	termite	
	tipitipi	tipitipin	tipitipiwuṛ	rock-cod (sp.)	
—i	ṭaput'i[illegible]	ṭaput'in	ṭaput'iwuṛ	older brother	
-u	muŋkumuŋku	muŋkumuŋkun	muŋkumuŋkuṛ	wooden axe	
	t'umput'umpu	t'umput'umpun	t'umput'umpuṛ	dragon fly	
	(E)				
	yukarpa	yukarpan	yukarpaṛ	husband	
	wulunka	wulunkan	wulunkaṛ	fruit (sp.)	
	wuṭalt'i	wuṭalt'in	wuṭalt'iwuṛ	meat	
	kantukantu	kantukantun	kantukantuṛ	red	
	karwakarwa	karwakarwan	karwakarwaṛ	wattle (sp.)	

CVC-

uninflected	*nonfuture*	*future*	
ṭera	ṭer-in	ṭer-uṛ	'thigh'
yura	yir-in	yur-uṛ	'body'
ṛilta	ṛil-in	ṛil-uṛ	'neck'
tjulta	tjil-in	tjul-uṛ	'hair'
maṛṭa	maṛ-in	maṛ-uṛ	'hand'
ṭuṛṭa	ṭuṛ-in	ṭuṛ-uṛ	'excrement'
wunta	wun-in	wun-kuṛ	'rain'
kaṇṭa	kaṇ-in	kaṇ-kuṛ	'grass'
njuta	njitj-in	njith-uṛ	'fire'

$\Sigma_{cv(c)}$

INVIOLABLE: PrWd, Avoid *σ non-finally, Coda Lic., ONS

	PrWd	Avoid *σ non-finally	Coda Lic.	ONS		V-ExM √(V)#	FILL *σ	PARSE *(X)	Prefer]=]stem σ	-COD
(A) **kentopal**										
ken.ta.pal. ✓										–
ken.ta.pa(l)								–	–	
cen.ta.p(al)			–					–	–	–
ken.ta.(pal)								–	–	
ken.ta.pals.							–		–	
(B) **wite**										
wi.te. ✓	○					⊖				
wit.e.				–		–				
wit.(e)	⊖					○		–	–	
wit.Δe.						–	–			
wi.te.ΔΔ.						–	–		–	
wi.teΔ.						–				–
(C) **yalulu**										
ya.lul.(u) ✓								⊖	–	–
ya.lu.lu.						⊖				
ya.lu.lua						–	–		–	–
ya.lu.(lu)								– –		
(D) **ŋaluk**										
ŋa.lu(k) ✓								(⊖)	–	
ŋa.luk.			⊖							–
ŋa.lu.kΔ.							(⊖)		–	

(A) ←]

(B) No final V deletion: block by minimal word con:

(C) V-ExM >> PARSE

Delete final V (but not preceding C)

(D) Delete final non-coda C (do not epenthesize)*

	PrWd	Avoid σ Non-Finally	Coda Lic.	ONS		V-ExM √(V)#	FILL *σ	PARSE *(X)	Prefer]=]stem σ	-COD
(E) **yukarpa**										
yukar(pa) ✓								⊖	–	–
yu.kar.pa.						⊖				–
yu.karp(a)			⊖					–		
yu.kar.paΔ.						(–)	–			–
'MW) **wik**										
wi.kΔ. ✓						○	⊖		–	
wi.k.	–		–							–
wi.k	⊖							(–)	–	
wi.ΔΔ.(k)		–					–			
wi.(k)ΔΔ.							–	–	–	
mar										
mar.ΔΔ. ✓						○	–		○	○
marΔ.							–		⊖	
mar.	⊖									
ma.(r)	–							–	–	
mar.ΔΔΔ.							–			–
ΔΔ.mar.		–							–	

(E) Delete final V & penult non-coda C (do not epenthesize)

Do NOT delete non-codal C. epenthesize*

Insert final CV (do not 'reanalyze' across +

Connectionism for Linguistics

Paul Smolensky

LSA Linguistic Institute
July 17, 1991

Collaborators:

- Géraldine Legendre (Linguistics, Colorado)
- Yoshiro Miyata (Cognitive Science, Chukyo)
- Alan Prince (Linguistics, Brandeis)

Connectionist Computation

- What is it? Low level characterization:
    - Data: activation values
    - Processing: spreading activation
    - Learning: associationist strength modification with experience
- Why is it? Cognitive motivations:
    - Biological neural networks: abstract out a *computational* description
    - Psychological models of:
        - Perception
        - Memory
        - Associationist processing
        - Empiricist learning

Computation & linguistics

- Psycholinguistics
    - Neurolinguistics
    - On-line processing
    - Acquisition
    - The form of grammar
        - **If cognition is *connectionist* computation, what are the implications for grammar?**
- Relation between connectionist and symbolic computation
    - Eliminativism: Connectionist *replaces* symbolic
    - Implementationalism: Connectionist *implements* symbolic
    - Revisionism: Connectionist *revises/enriches* symbolic
        - [A major research program is required to pull this off; multiple levels.]
- Grammar
    - Eliminativism: Grammar is a fiction.
    - Implementationalism: Grammar is unchanged.
    - Revisionism: **Grammar inherits new *kinds* of rules & representations from connectionist computation**
        - Requires *high-level* characterization of connectionist computation [which in turn requires a major research program to develop]

Computing done by linguist vs. computing done by speaker/hearer.

Examples of new ways of computing/thinking:

1. Continuous representations
2. Similarity-based processes
3. Parallel *soft* constraint satisfaction
4. Quantitative measures of well-formedness
5. Optimality as main organizing principle

Are these *good* ways of computing/thinking? You tell me.

Plan for lectures:

For $p \in \{1,2,3,4,5\}$

A. Develop high-level principle p of connectionist computation.
B. Propose a way of importing it into linguistics.
C. Develop an example application.

Logical progression of principles: 1 → 5 (lowest, 'most connectionist' → highest, 'most symbolic')

Conservative presentation of linguistic applications: 5 → 1 — wins.

Today: 5; Next week: 4 - 1

5. Optimality as an organizing principle of grammar

A. Connectionist principle:

Harmony maximization. In a central and interesting class of connectionist networks, activation flow creates from an input a pattern of activity that [locally] maximizes a numerical well-formedness measure, the *Harmony*:

$$\Sigma_i\ \Sigma_j\ a_i\, w_{ij}\, a_j$$

Linguistic interpretation:

connectionist input → linguistic 'input' to grammar (string of words, phonemes)
connectionist well-formedness → linguistic well-formedness
pattern of activity arising from an input → structural description of the input
pattern of connections → grammar

Here, all grammatical representations will be symbolic: traditional, unrevised. The rules too will be traditional in form, but their interpretation and interaction will be new. (Conservatism.)

B. Proposed linguistic principle:

Optimality. A grammar is a *preference relation* $\succ$ among structural descriptions; it assigns to any input that structure which is *optimal* (prefered by $\succ$ to all other structures).

The preference relation $\succ$ among structures is built up compositionally from preference relations among the substructures or subdimensions from which structures are built.

The basic preference relations and their means of combination are drawn from a universal repertoire.

Linguistic-internal motivation:

Grammars are more explanatory when they explicity state well-formedness constraints rather than describing procedures to satisfy those constraints.
(Prefer principles to rewrite rules/transformations.)

Phonology has not made the progress towards this goal that syntax has — because the constraints are so often of the form, 'prefer ... if possible' rather than 'obey ... or be ill-formed'.
(But re:syntax, cf. Sadock's constraints on inter-component discrepancies.)

C. Applications in phonology [for many more, cf. Prince & Smolensky, LINS 247]

a. Berber syllabification

[insert brief statement of problem]

Idea: Dell—El-Medlaoui algorithm is trying to build the best syllables possible, as defined primarily by the goodness of their peaks. Replace the algorithm with a formal statement of this well-formedness metric.

More sonorous segments make better syllable peaks:

align sonority prominence hierarchy $8 \sqsupset 7 \sqsupset \ldots \sqsupset 1$ with syllable structure prominence hierarchy: peak$\sqsupset$non-peak

$$(8,\text{peak}) > (7,\text{peak}) > \ldots > (1,\text{peak}).$$

Given a syllabification s, let M_s = the (multi-)set of peaks in s.
Let $first(M_s)$ = best peak of M_s, and let $rest(M_s) = M_s \setminus first(M_s)$.
Then recursively define:

$M_s \succ M_{s'}$ iff

$first(M_s) > first(M_{s'})$

or

$first(M_s) = first(M_{s'})$ and $rest(M_s) \succ rest(M_{s'})$

Note: This means of composing the ordering $>$ on peaks to an ordering $\succ$ on syllabifications is a generalization of alphabetical ordering of strings: reinterpret *first* as 'left-most element'. It turns out to be useful quite generally in the theory, so we dub it Generalized Lexicographic Ordering, **GLO**. To define a GLO $\succ$, we must specify the sets M_s, the function *first*, and an ordering $>$ of individual elements of M_s.

b. Lardil

A **Strict Domination Grammar** is a preference relation $\succ$ defined from a sequence of Boolean predicates over possible structural descriptions

$$p_1 \gg p_2 \gg \ldots \gg p_n$$

by the following case of GLO:

> For a given structure s, let $M_s = \{\pm p_1, \pm p_2, \ldots.\}$ where the sign of p_k indicates whether s satisfies predicate p_k (+) or not (−). Let $first(M_s)$ = first element of M_s. On each predicate separately, $+p_k$ (preference satisfaction) $>$ $-p_k$ (preference violation).

Note: This is a form of *markedness theory*. Each '$-p_k$' is a kind of 'mark' on the structure, and the ordering $\succ$ gives a formal means of deciding which structure as a whole is least marked, by examining individual marks.

For clarity, we write 'Prefer p_k' for p_k and 'Avoid q_k' when $p_k = \sim q_k$.

The Lardil preferences:

p_1: Avoid subminimal words [less than one foot].
p_2: Avoid non-final defective syllables.
p_3: Avoid complex consonants in coda position [allowed: one non-back coronal].
p_4: Prefer filled onsets.
p_5: Prefer free final V.
p_6: Avoid defective syllables.
p_7: Avoid free material.
p_8: Prefer coincidence of stem & syllable boundaries.