Strict Domination in Connectionist Networks

Adam Wayment
wayment@cogsci.jhu.edu
Department of Cognitive Science
Johns Hopkins University
HUMDRUM
April 23, 2005

Overview

- Review strict dominance and additive numerical ranking in Brbrnet.
- Why additive numerical ranking doesn’t work
- A new dynamic/architecture for strict domination in connectionist networks
- Case study: Brbrnet revisited

Berber facts

- Berber allows any segment to occur in syllable nucleus position (Dell & Elmedlaoui 1989, 2003).
- A sonority hierarchy specifies that some segments make better nuclei than others (Prince & Smolensky 1993/2004):

  [a] > [i] > [r] > [n] > [z] > [s] > [d] > [t]

  *[a]/M >> *[i]/M … *[d]/M >> *[t]/M
- A high ranked onset constraint requires that there be an onset between every nucleus:

  ONSET>>*[a]/M >> *[i]/M … *[d]/M >> *[t]/M

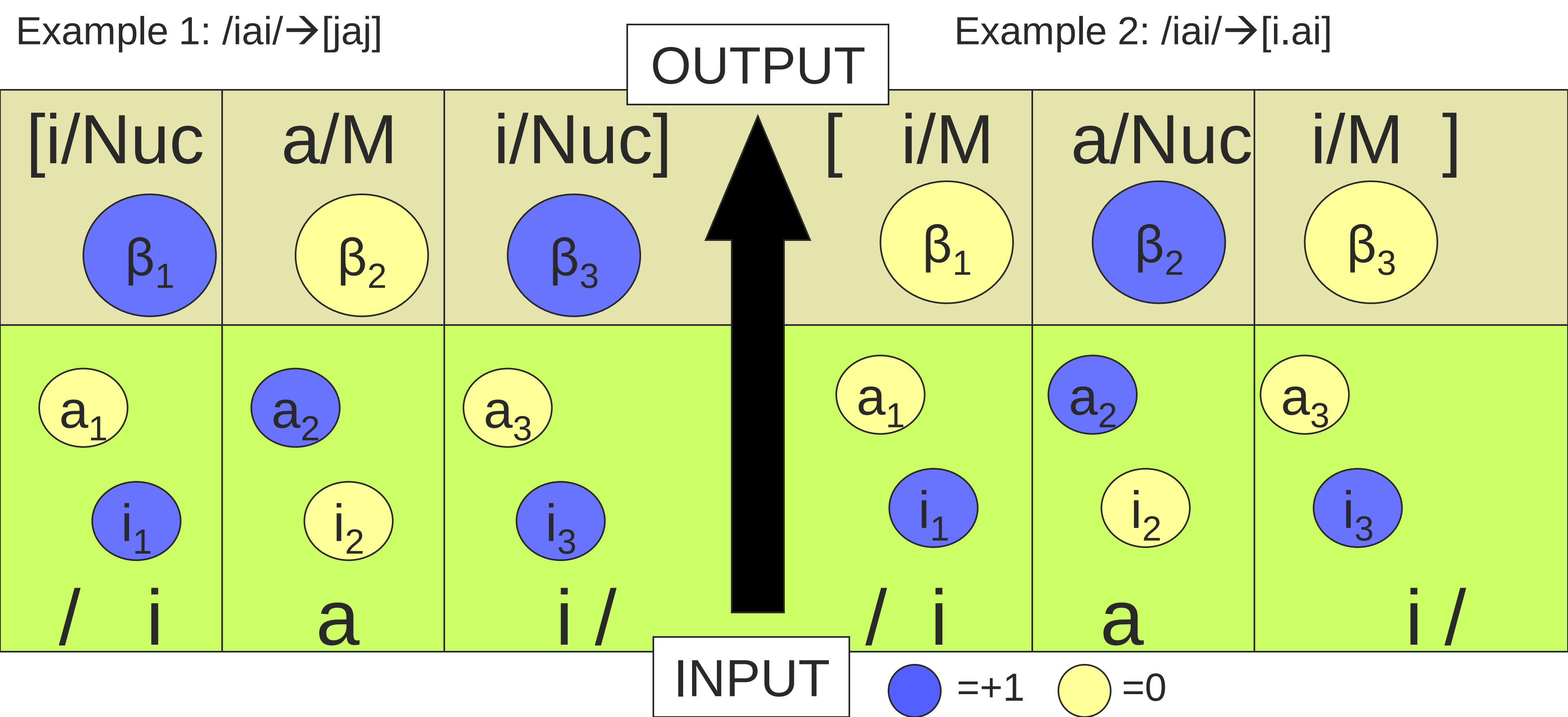
Strict dominance in Berber

- Grammar in Optimality Theory (Prince & Smolensky,1993/2004)
  - A set of constraints: $C_1 \dots C_n$
  - A strict dominance relation among $C_1 \dots C_n$
- No amount of violations of lower ranked constraints can "overpower" a higher ranked constraint.

/iai/		ONS	*a/M	*i/M
☹**a.**	[jaj]			2
b.	[i.ai]		1	

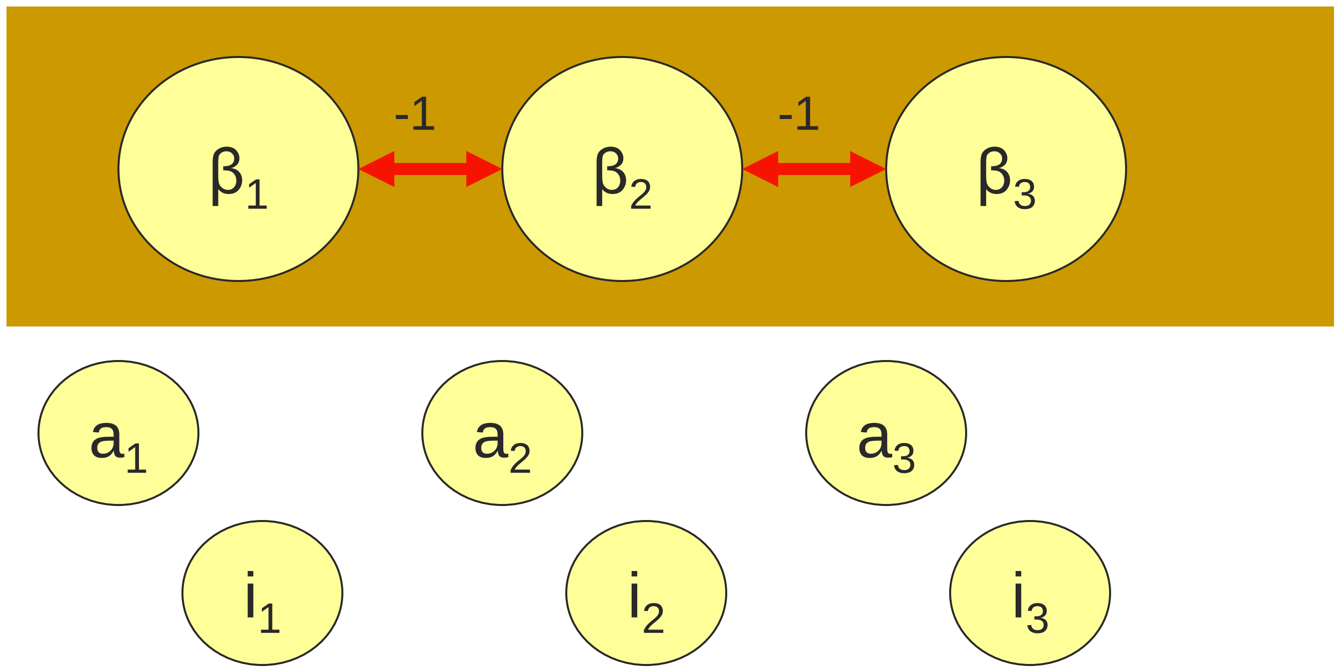
Brbrnet architecture
(Legendre, Sorace, Smolensky 2003)

- Input: Turn on the sonority unit corresponding to the sonority of each input segment α_i
- Output $\beta_i = 0 \rightarrow \alpha_i$/M; $\beta_i = 1 \rightarrow \alpha_i$/Nuc



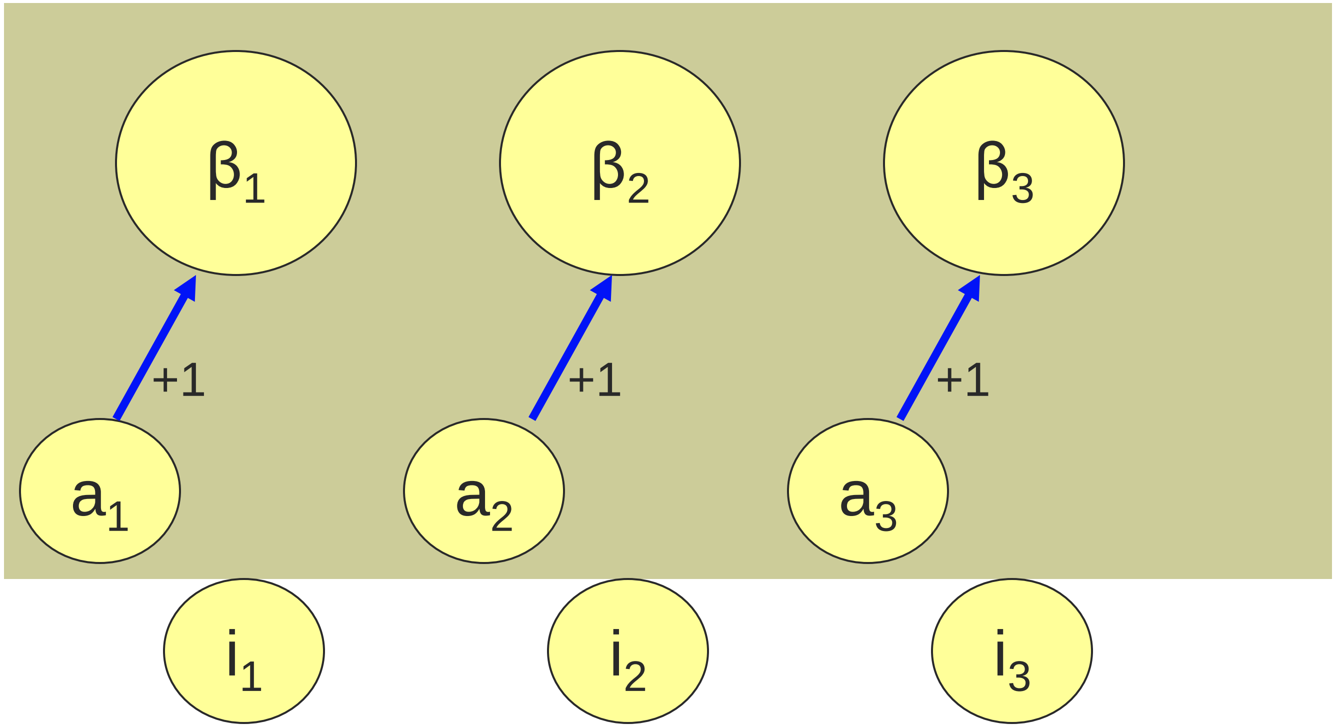
Brbrnet constraints: ONSET

- Each constraint is implemented by a weight matrix.
- ONSET: *(β_i=1,β_{i+1}=1)



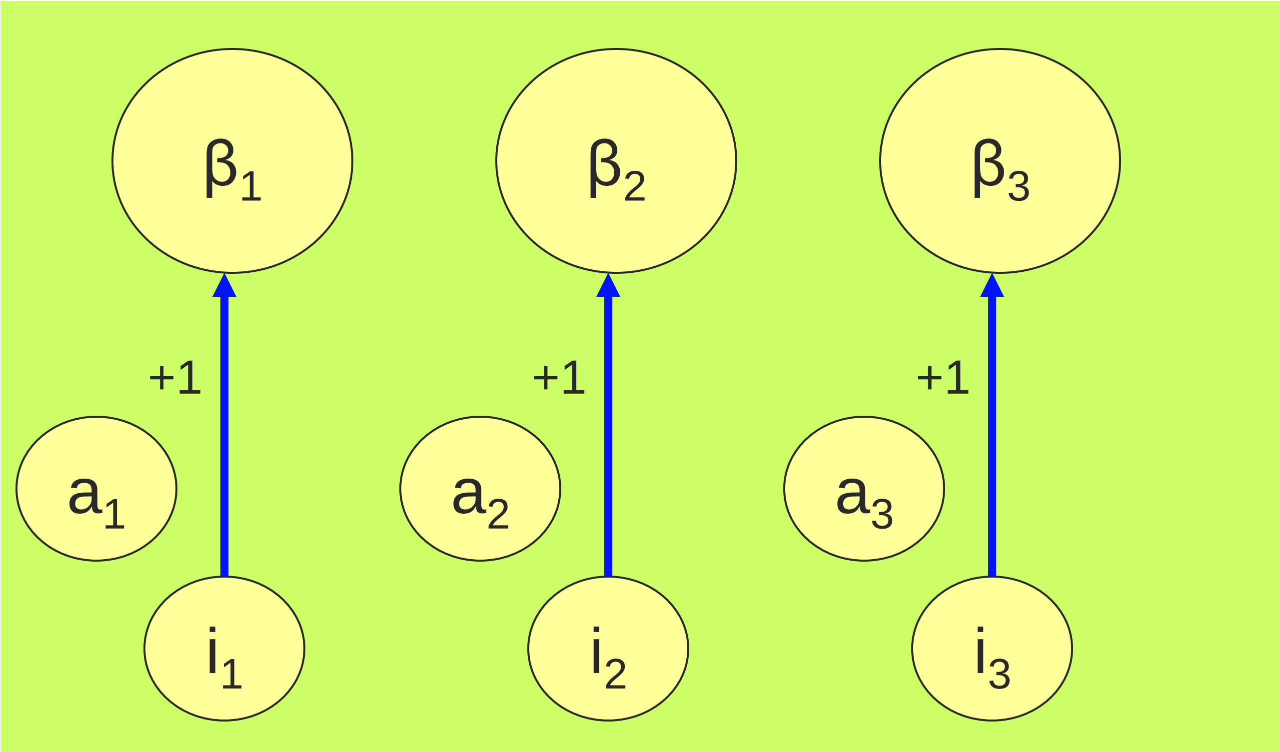
Brbrnet constraints: *a/M

- *a/M: $*(a_i = 1, \beta_i = 0)$



Brbrnet constraints: *i/M

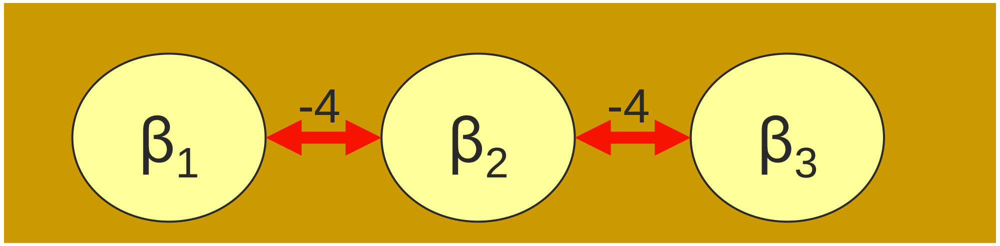
- *i/M: *($i_i = 1, \beta_i = 0$)



Harmonic grammar

(Legendre et al. 1990)

- A markedness constraint is represented by a weight matrix on a group of surface units.
- The constraint's strength is a multiplier on that weight matrix.
- Example. Let $S_{ONSET} = 4$, then the weight matrix for the Onset is



- The final weight matrix for the network is given by

(1) $$\mathrm{W}_{final} = \sum_i s_i \mathrm{W}_{\mathrm{C}i}$$

- For a specific pattern of activation α,

(2) $$H(\alpha) = \sum_{i,j} \alpha_i \mathrm{W}_{ij} \alpha_j$$

Additive numerical ranking

- The candidate with the lowest **score**, the weighted sum of violations, is the optimum.
- Prince (2002) showed that, in general, linear additive numerical ranking and strict domination are not equivalent
  - i.e. $s_{C1} > s_{C2}$ does not usually imply $C_1 >> C_2$.

CC-Counter-example

- Consider the following OT grammar:
  - GEN = {PEAK, IDENT-IO}
  - Ranking:

    DEP[1] >> MAX >> {NoCODA, *COMPLEX}
  - Deletion is the repair mechanism to satisfy GEN constraints.
  - No repairs are made to satisfy the low ranked MARKEDNESS constraints NoCODA, *COMPLEX

---

1. DEP and MAX come from McCarthy & Prince's (1994) Correspondence Theory.

Why doesn't linear ranking work?

/CC/	DEP	MAX	No CODA	*Complex	S_{DEP} 1.00	S_{MAX} 0.90	S_{NoCODA} 0.10	$S_{*COMPLEX}$ 0.10	Score
☹❄**a.** ■■		2			0.00	1.80	0.00	0.00	❄ 1.80
b. CvC	1		1		1.00	0.00	0.10	0.00	💣 1.10
c. CCv	1			1	1.00	0.00	0.00	0.10	💣 1.10

Exponential ranking facts

- Prince and Smolensky (1993/2004) state that exponentially weighting the strengths can provide strict dominance **if** there is an upper bound on the number of violations (p. 236).

(3) For $C_1 \gg C_2 \gg \ldots \gg C_k$.

$C_1 = 1.0$

$s_i < (1/n)s_{i-1} \quad s_i \approx (1/n)^{(i-1)}$

- Brbrnet example:
  - $S_{ONSET} = 2^8 > S_{*a/M} = 2^8 - 1 > S_{*i/M} = 2^7 - 1 > S_{*r/M} = 2^6 - 1 \ldots$
- What should be the base of exponentiation?
- How do you know you have picked the right one?

Exponential ranking in action

/CC/	DEP	MAX	NoCODA	*COMPLEX	S_{DEP}	S_{MAX}	S_{NoCODA}	$S_{*Complex}$	Score
					1.00	0.49	0.10	0.10	
☹❄**a.** ■■		2			0.00	0.98	0.00	0.00	☹❄0.98
b. CvC	1		1		1.00	0.00	0.10	0.00	1.10
c. CCv	1			1	1.00	0.00	0.00	0.10	1.10

Uh-oh

/CCC/	DEP	MAX	MARKEDNESS	S_{DEP}	S_{MAX}	$S_{MARKEDNESS}$	Score
				1.00	0.49	0.10	
☹❄**a.** ■■■		3		0.00	1.47	0.00	❄1.47
b. CCvC	1		2	1.00	0.00	0.20	💣1.20
c. CvCC	1		2	1.00	0.00	0.20	💣1.20
d. CCCv	1		2	1.00	0.00	0.20	💣1.20

An additive numerical ranking must depend on input length for strict domination

- If the base of exponentiation does not depend on the size of the input, then

    given a numerical ranking

    there exists (by Richness of the Base) an input of C's

    that will break strict domination.

- Fixes for a Harmonic Grammar:
    - Invoke a high-ranked self-conjunction of *COMPLEX.
    - Posit an additional Null-parse candidate that does not violate correspondence constraints (McCarthy & Wolf, 2005).
    - Place an upper bound on input length or on the number of constraint violations. (see Finite State OT. Frank & Satta, 1998).

Beyond additive numerical ranking.

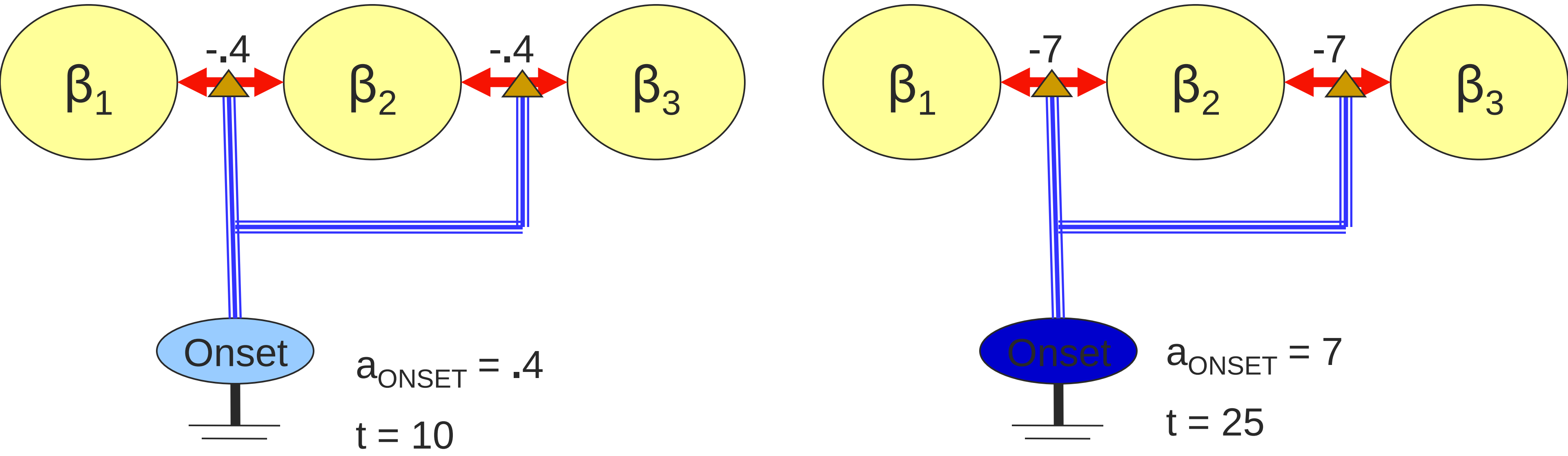
- Critically the result does depend on unbounded input size.
- Requiring a dependence on input is tantamount to requiring an adjustment of ranking for each input.
- BUT by the Nativist hypothesis implicit in OT typologies, ranking is the only means of grammar definition.
- So changing the strengths means *changing the grammar* for each input.
- Therefore, I reject additive numerical ranking as a viable connectionist computational model of OT.

A new strict dominance dynamic (SDD) for networks

- Motivations:
  - Add Constraint units to the network whose activation represents the strength of constraints.
  - The weights of a constraint are multiplied by the activation (strength) of the Constraint unit.
  - Because the strength is an activation, it can change in time.
  - Allow constraint units activation to grow exponentially in hope that the stronger constraints will always get strong enough to force strict domination.

Recipe for strict domination 1

- The strength of a constraint will be encoded as a *multiplier* connection from a *constraint unit* to all connections in W_{Ci}.
- Now, the strength of a constraint can change over time, during ***the course of performing an optimization*** (an input-output mapping).



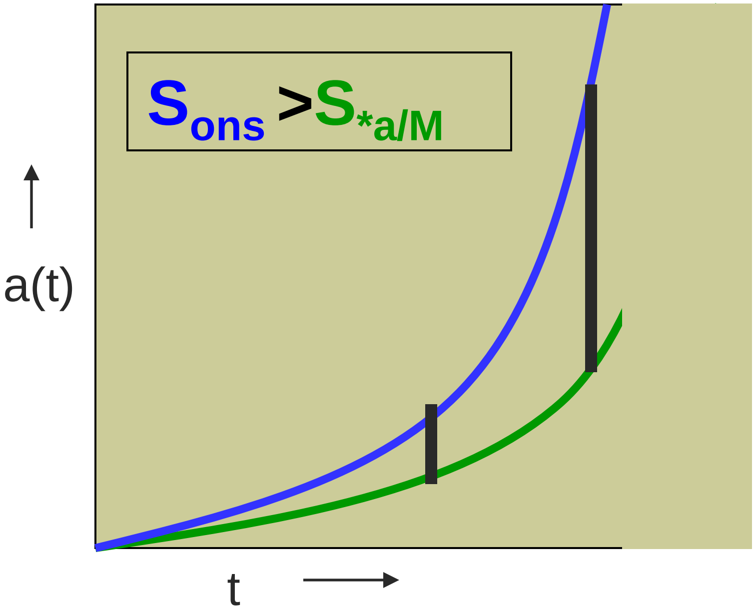
Recipe for strict domination 2

- Swap exponential weighting with exponential activation.
- Now all constraint units grow exponentially, but the stronger constraint will grow faster, so the relative strengths grows exponentially.
- Multiplier connections apply to all potential violations at the same time, so the strengths do not depend on the input.

$$a_{ONSET}(t) \approx e^{Sons \cdot t}$$

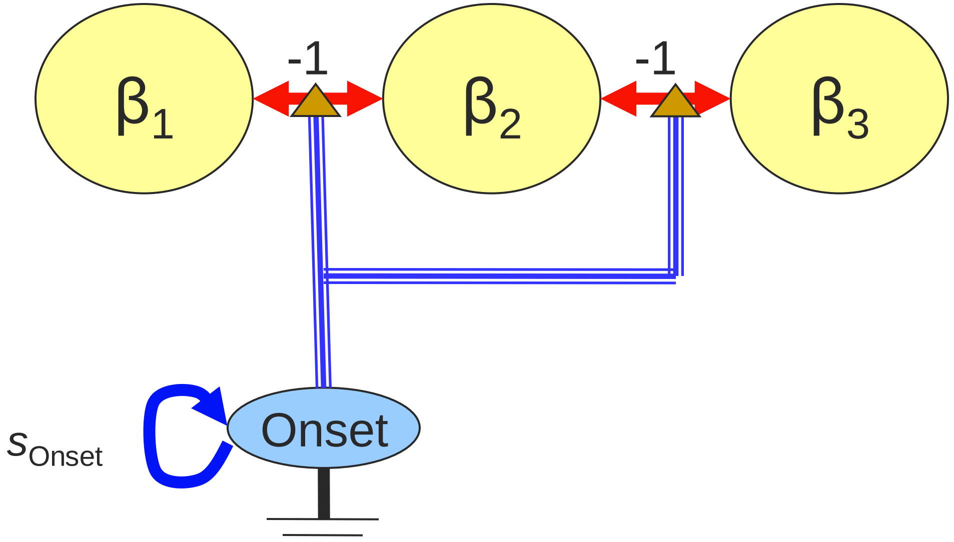
$$a_{*a/M}(t) \approx e^{S*a/M \cdot t}$$

$$a_{ONSET} - a_{*a/M} = e^{(Sons - S*a/M) \cdot t}$$



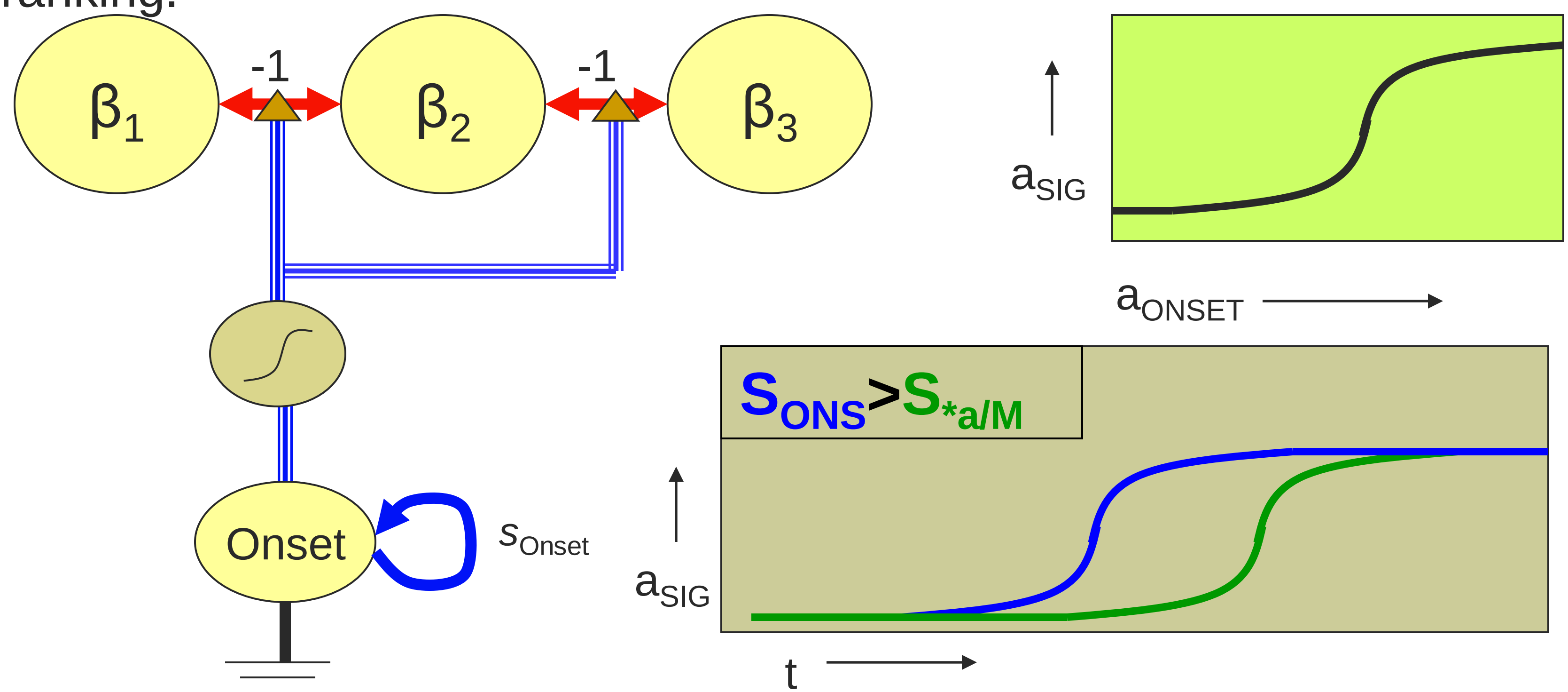
Exponential activation through self-connections.

- Add excitatory recurrent connection on the Constraint units.
- The strength of this self-connection is the ranking of the constraint.



Recipe for strict domination 3

- Initially, constraint unit's activations are all very small; so the system is nearly linear, and ganging up effects are readily observed
- Adding sigmoid units spreads the constraints' strengths apart in time.
- The constraints turn "on" in the order[1] specified by the numerical ranking.



Recipe for strict domination 4

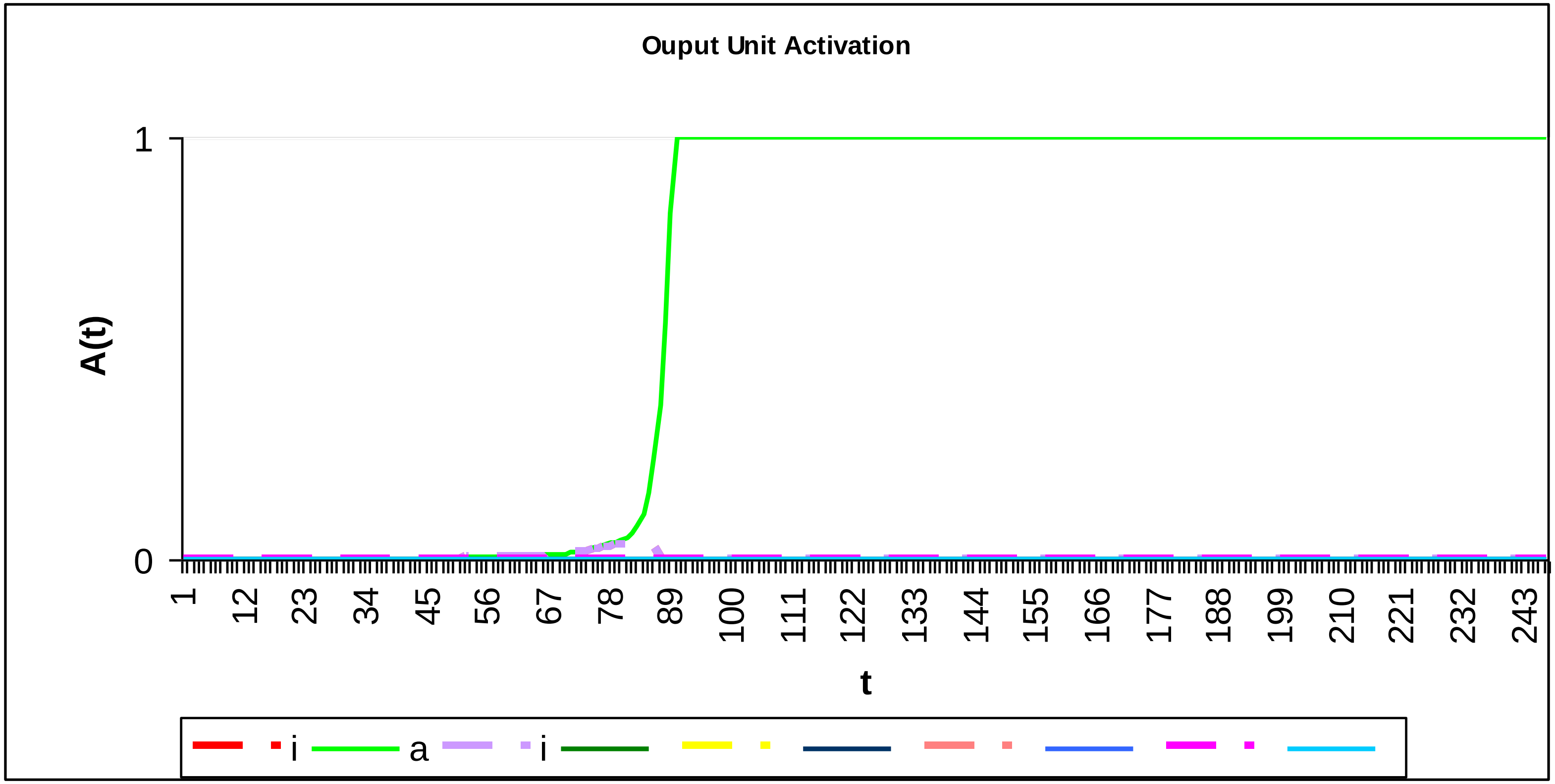
- Start optimization by
  - fixing the input pattern of activation
  - Starting the constraint units with equally small initial activations.
- End optimization when
  - The output pattern of activation stabilizes
    - Note constraint units will not stop changing.

Testing Brbrnet with SSD.

/iai/		ONS	*a/M	*i/M	S_{ONS}	$S_{*a/M}$	$S_{*i/M}$	Score	
					1	0.9	0.88		
❄☹**a.**	[jaj] = [010]			2	0	0	1.76	❄	1.76
b.	[i.ai]=[101]		1		0	0.9	0	☹	0.9

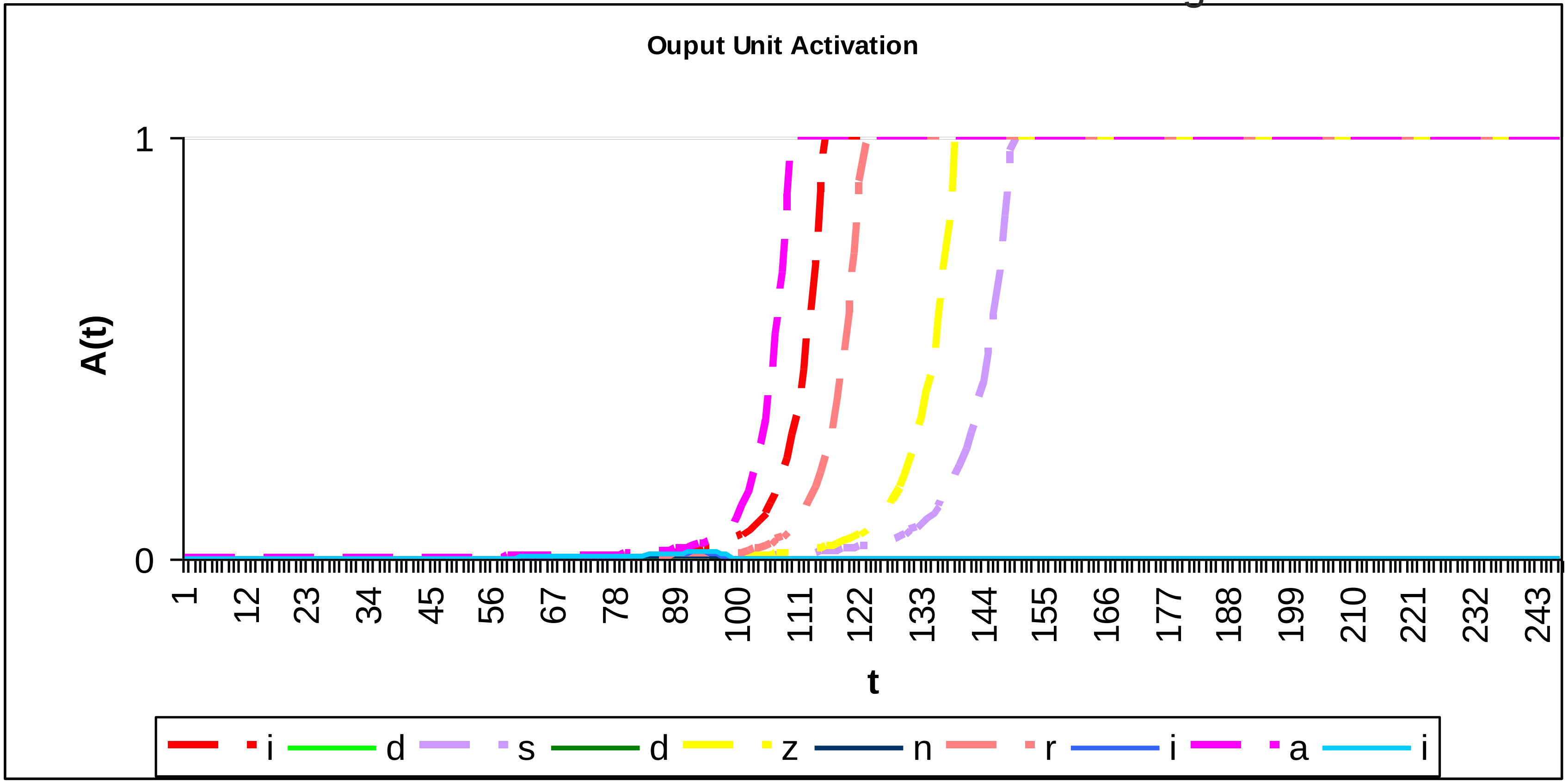
BrbrNet with SDD

On input /iai/ produces output [jaj] = [010].



/idsdznriai/→[i.ds.dz.nr.jaj]

- Turning constraints on in the order of their ranking gives a connectionist instantiation of the Dell-Elmedlaoui Algorithm:



Future Work

- Truly test SDD by implementing a syllabification network that can test strict dominance in the CC-Counter-Example Grammar.
- Developing a learning algorithm that adjusts constraint strengths based on input-output training pairs.
- Adding noise to the constraint units biases to perform Stochastic OT (Boersma, 1997).
- More Math.
  - Analytical descriptions of the network properties.

Conclusion

- Additive numerical ranking is a flawed implementation of strict dominance.
- My SDD avoids these complications by making constraints active in the order of their numerical ranking.
  - In general, SDD networks are similar to the optimization by successive intersections described by (Frank & Satta, 1998).

Acknowledgements

- Paul Smolensky, my advisor, for among many things, the Bnchmrk.
- Don Mathis for network engineering help.
- JHU Linguistics Lab and JHU faculty for thoughtful comments and suggestions.
- Paul Boersma for initially suggesting the need for a new dynamic.

References

Boersma, P. 1997. How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the Univ. of Amsterdam* 21:43-58.

Frank, R. and Satta G. 1998. Optimality Theory and Generative Complexity of Constraint Violability. *Computational Linguistics* 24:307-315.

Legendre, G. Mityata, Y. and Smolensky, P. 1990. Harmonic grammar – a formal multi- level connectionist theory of linguistic well-formedness: Theretical foundations. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society,* 288-295, Cambridge MA. Erlbaum.

Legendre, G., Sorace, A. & Smolensky, P. 2003. The Optimality Theory -- Harmonic Grammar connection**.** Chapter 20 in Smolensky, P. and Legendre, G. 2005. *The Harmonic Mind: From Neural Computation to Optimality: Theoretic Grammar*. MIT Press.

McCarthy, J. & Prince, A. 1994 Faithfulness and Reduplicative Identity, in *University of Massachusetts Occasional Papers in Linguistics* 18: Papers in Optimality Theory. Ed. by Jill Beckman, Suzanne Urbanczyk and Laura Walsh Dickey. Pp. 249–384.

McCarthy, J. & Wolf, M. 2005. Less than zero: Correspondence and the null output. ROA-722-0305.

Prince, A. Anything Goes. 2002. ROA-536.

Prince, A. and Smolensky, P. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Cambridge, MA: Blackwell. Technical Report, Linguistics Department, Rutgers University, New Brunswick, NJ and Department of Computer Science, University of Colorado at Boulder (1993). Rutgers Optimality Archive 537 (2002).

Smolensky, P. and Legendre, G. 2005. *The Harmonic Mind: From Neural Computation to Optimality: Theoretic Grammar*. MIT Press.

Soderstrom, M., Mathis, D., and Smolensky, P. 2001. Toward computational empirical testing of linguistic innateness: Abstract genomic encoding of an Optimality-Theoretic grammar. *Proceeding of the Third International Conference on Cognitive Science*, Beijing, China, 14-25. University of Science and Technology of China Press.

Appendix A

/CC/	DEP	MAX	No CODA	*Complex	S_{DEP} 1.00	S_{MAX} 0.90	S_{NoCODA} 0.10	$S_{*COMPLEX}$ 0.10	Score
☹❄**a.** ■■		2			0.00	1.80	0.00	0.00	❄ 1.80
b. CvC	1		1		1.00	0.00	0.10	0.00	☹ 1.10
c. CCv	1			1	1.00	0.00	0.00	0.10	☹ 1.10
d. Cv.Cv	2				2.00	0.00	0.00	0.00	2.00
e. Cv ■	1	1			1.00	0.90	0.00	0.00	1.90
f. vCC	1		1	1	1.00	0.00	0.10	0.10	1.20
g. vC ■	1	1	1		1.00	0.90	0.10	0.00	2.00

Appendix B

/CCC/	DEP	MAX	MARKEDNESS	S_{DEP}	S_{MAX}	$S_{MARKEDNESS}$	Score
				1.00	0.49	0.10	
☹❄**a.**■■■		3		0.00	1.47	0.00	❄1.47
b.CCvC	1		2	1.00	0.00	0.20	☹1.20
c.CvCC	1		2	1.00	0.00	0.20	☹1.20
d.CCCv	1		2	1.00	0.00	0.20	☹1.20
e.Cv**.**Cv**.**Cv	3			3.00	0.00	0.00	3.00

Appendix C

/CCC/	DEP	MAX	**MARKEDNESS**	S_{DEP} 1.00	S_{MAX} 0.49	$S_{MARKEDNESS}$ 0.10	Score
☹❄**a.**■■■		3		0.00	1.47	0.00	❄1.47
b.CCvC	1		2	1.00	0.00	0.20	☹1.20
c.CvCC	1		2	1.00	0.00	0.20	☹1.20
d.CCCv	1		2	1.00	0.00	0.20	☹1.20
e.Cv**.**Cv**.**Cv	3			3.00	0.00	0.00	3.00

Appendix D: Adding SDD to a Harmonic Grammar.

For each constraint matrix W_{Ci}

- Add multiplier connections (+1) to sigmoid unit SIG_i.
- Connect (+1) SIG_i to constraint unit C_i.
- Add recurrent connection (s_i) on the constraint unit C_i.
- Add a bias connection w_b to C_i.

Appendix E: Timing details

- The constraint units and the surface units were placed on different time scales in order to give the surface units time to respond to changes in the activity along the multiplier connections.
- The time scale for surface units was much faster than the time scale of the constraint units.
- It is not clear at this point how these relative timing scales interact to determine the outcome because of the complex interaction of initial activation value, numerical ranking strength, and the slope and mean of the sigmoid function.