Partitions & Parameters

Roeland Hancock & Thomas G. Bever
Department of Psychology
University of Arizona
Tucson, AZ US

Overview

- Acquisition theories
    - Nativism: Parameter setting
    - Empiricism: Connectionism, statistical learning
- The middle road: a dynamical systems model
    - Dynamical systems primer
    - Learning mechanisms
    - Examples

Theoretical Issues

- Acquisition
    - Poverty of the stimulus (POS)
    - Creating new candidate grammars (NCG)
- Possible and impossible languages
- Universals
- Language change

Solution 1: Parameters (Chomsky & Lasnik)

- Innate UG
- Set parameters with critical examples
- Addresses
  - POS
  - Universals

Possible Parameters

- Word Order
    - VO (*e.g.* Modern English)
    - OV (*e.g.* German, Dutch)
- Branching
    - right-branching (*e.g.* English)
    - left-branching (*e.g.* Japanese)

Parameters

- Need some structure before setting parameters
- How would such a system actually work in the brain?
- Language change
  - “language change sometimes takes place through an abrupt change in grammars, reflecting a new parameter setting. In that case, one cannot view language acquisition as a function of children matching their input...” (Lightfoot, 1997)
  - or may be gradual, occurring over multiple generations

Mixed Evidence

- Degree-0 learnability (Lightfoot, 1991)

- German and Dutch are OV, but also have V2 movement

(5a) Ich$_{Subj}$ denke$_{TensedVerb}$, das ich [*den Fuchs*]$_{Obj}$ *sehen*$_{Non\text{-}TensedVerb}$ *kann*$_{TensedVerb}$
I think that I the fox see can
'I think that I can see the fox'.

(5b) Ich$_{Subj}$ kann$_{TensedVerb}$ t_{Subj} [den Fuchs]$_{Obj}$ sehen$_{Non\text{-}TensedVerb}$ $t_{TensedVerb}$
I can the fox see
'I can see the fox'.

Example from Pearl & Weinberg (2007)

Solution 2: Connectionism

- Goal: form connections that are equivalent to those of a native speaker
- biologically relevant (sometimes)
- minimal built-in assumptions
- well-understood mechanics

Connectionism

- POS
- poor at generalizing
- minimal linkage between different structural descriptions
- Not realistically constrained (cf Love et al, 2006; Hancock, 2008)

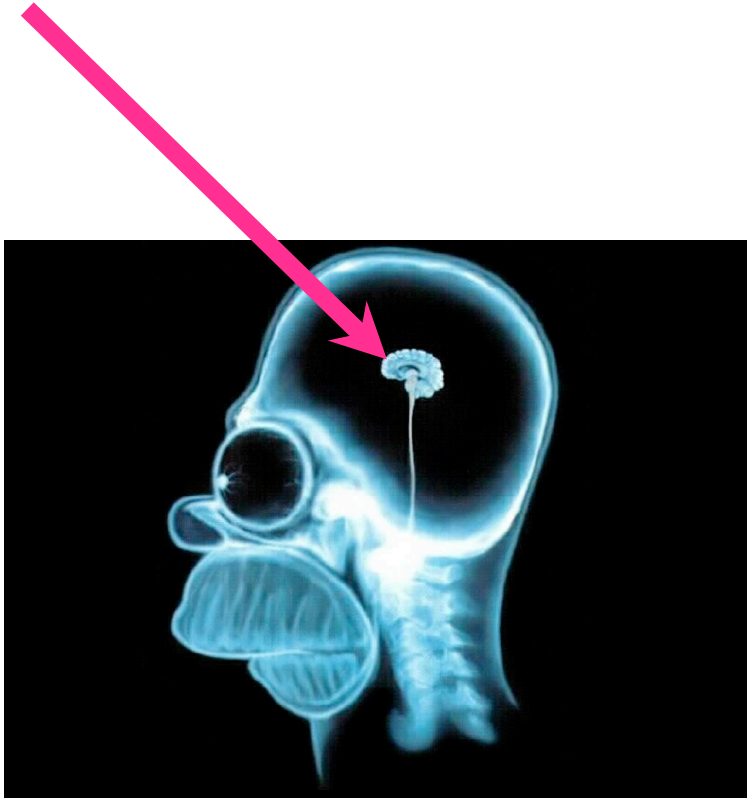
Statistical Learning

- Behavioral evidence
- Notable that languages have statistical cues (e.g. canonical forms; Bever, 2007)
- Not powerful enough to produce grammars

The Middle Road

- Provide the explanatory power of P&P in a learning framework
    - account for abrupt, parametric changes; creolization
    - without structural analysis
    - using minimal data
    - constraints on possible languages
- Using dynamical systems

A (small) dynamical system



Dynamical Systems

Iterated Function Systems (IFS)

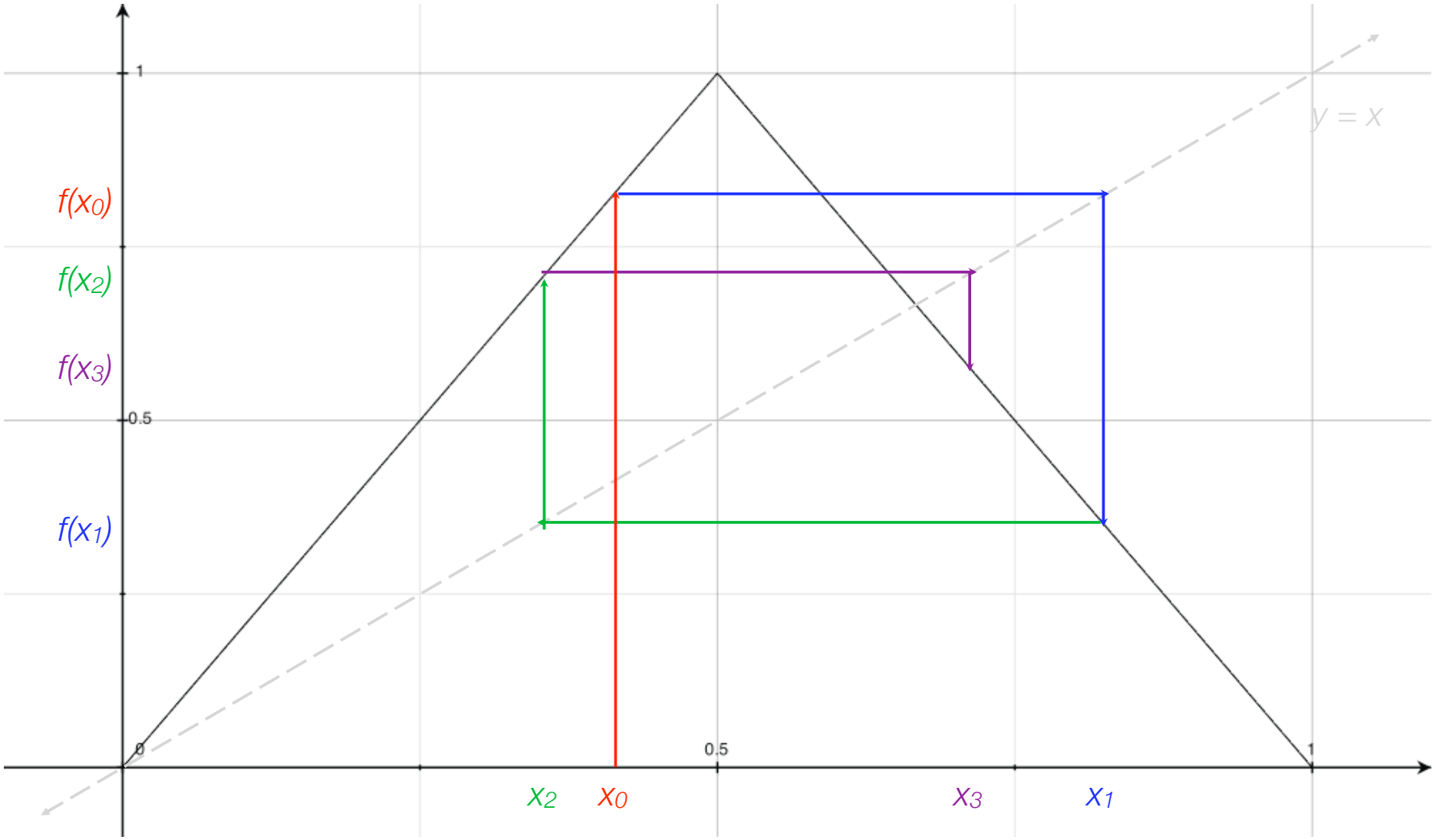
- A function, *f*, over a real interval, *I*, with *f(I)* ⊆ *I*

$$x_1 = f(x_0), x_2 = f(x_1), x_3 = f(x_2), ..., x_n = f(x_{n-1})$$

- Orbits:

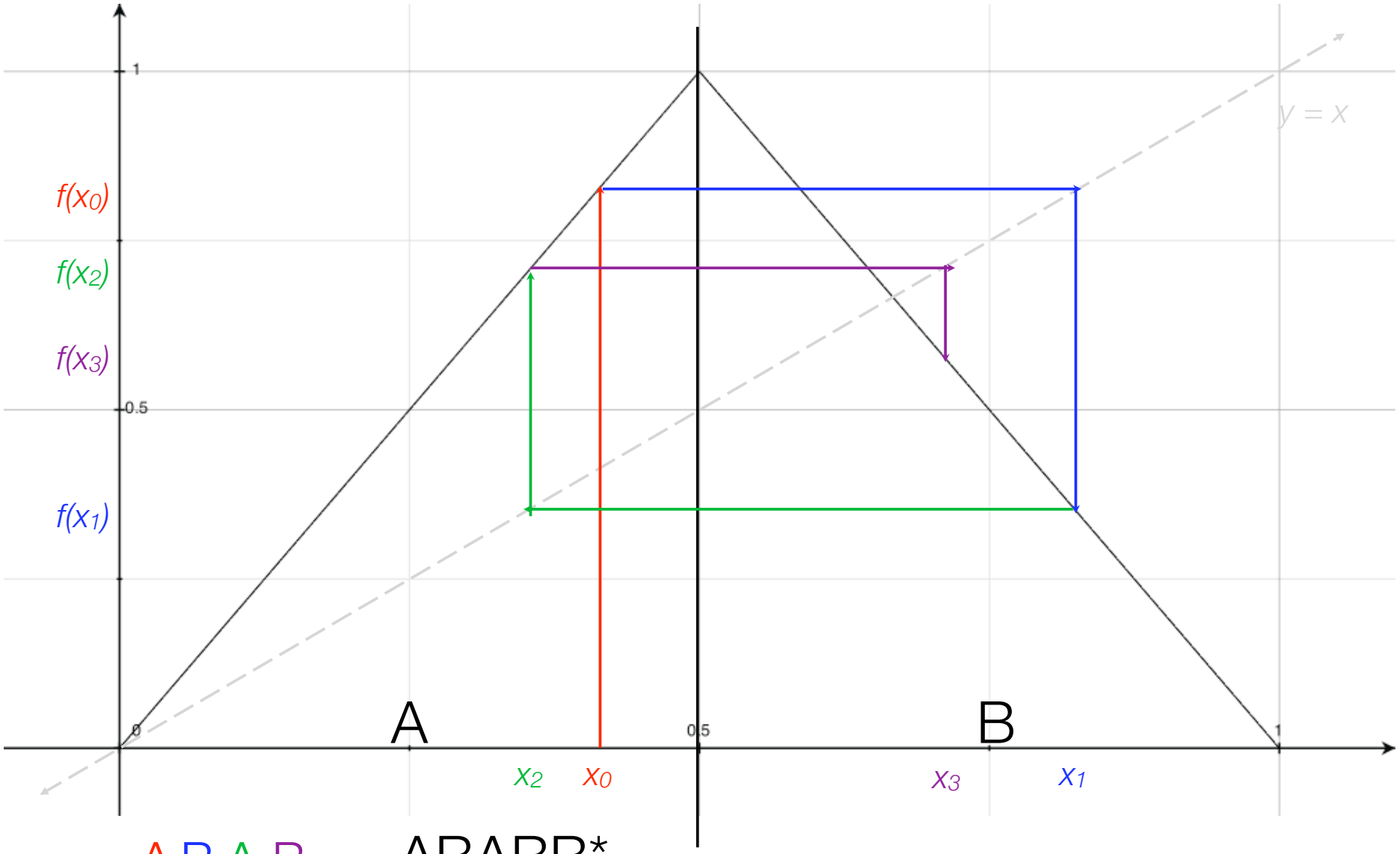
$$\{x_0, x_1, ..., x_n, ...\}$$

- Connectionist models (e.g. SRNs) can also be described as dynamical systems (Andrews, 2003)



Iterated Map

Tent Map on [0,1]



A B A B... ABABB*

Symbolic Dynamics

Tent Map on [0,1]

Model

Present Model

- Assume a *fixed* state space (something like UG)
- Find a partition that is consistent with the input data
- POS is only relevant insofar as the poverty limits the identification of a partition
    - also true for any parameter theory
- with a suitable partition
    - unobserved regions of state-space can be accessed
    - and previously accessible regions can be blocked

Tent Map as an Example

- A very simple example filling in for a fixed function (something like UG)
- Orbits form derivations
- Real interval is an abstract lexicon
- Placement of partition is critical-functions like parameters
- In the brain-thalmocortical loops (Ullman, 2006)

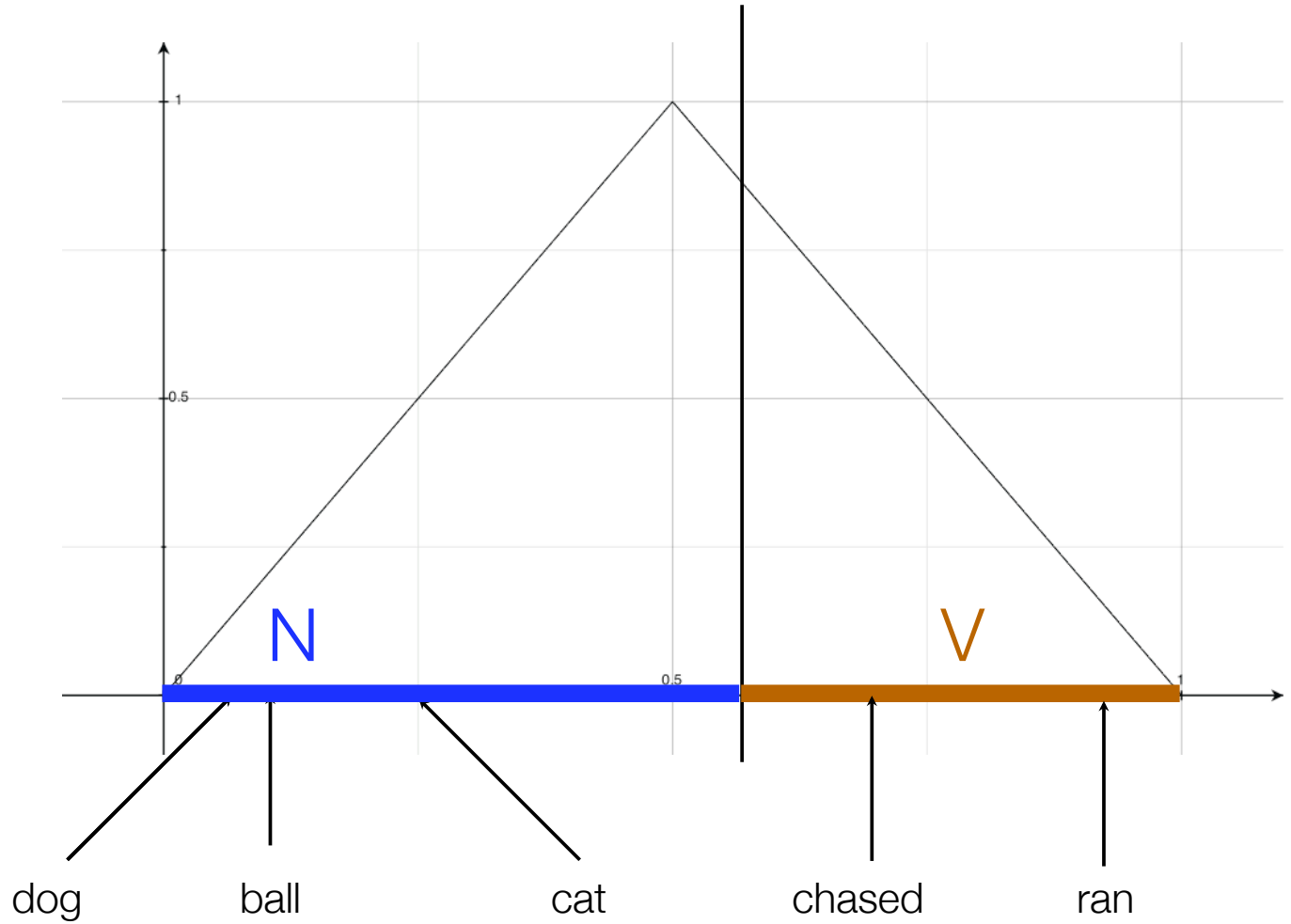
Dynamical Grammar

- Statistical Processes
  - Chunker
    - break a sentence into labeled chunks in the set C
  - Partition $\mathcal{P}$
  - A map Φ: $C \rightarrow \mathcal{P}$
- Innate Process
  - A function *G:* $\mathcal{P}$ →Structural Descriptions (*SD)*

Partition

↑ ϕ

Chunks

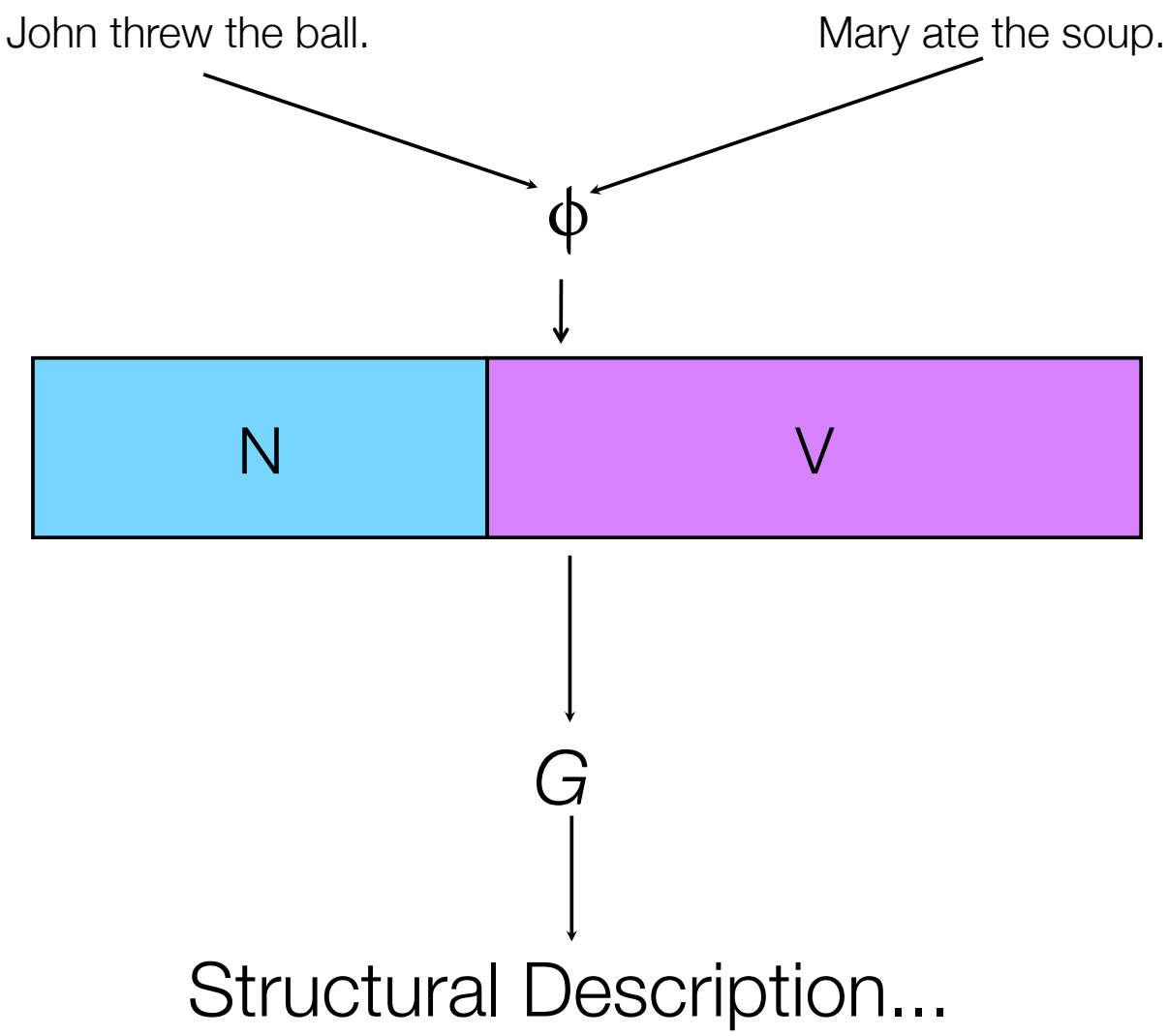


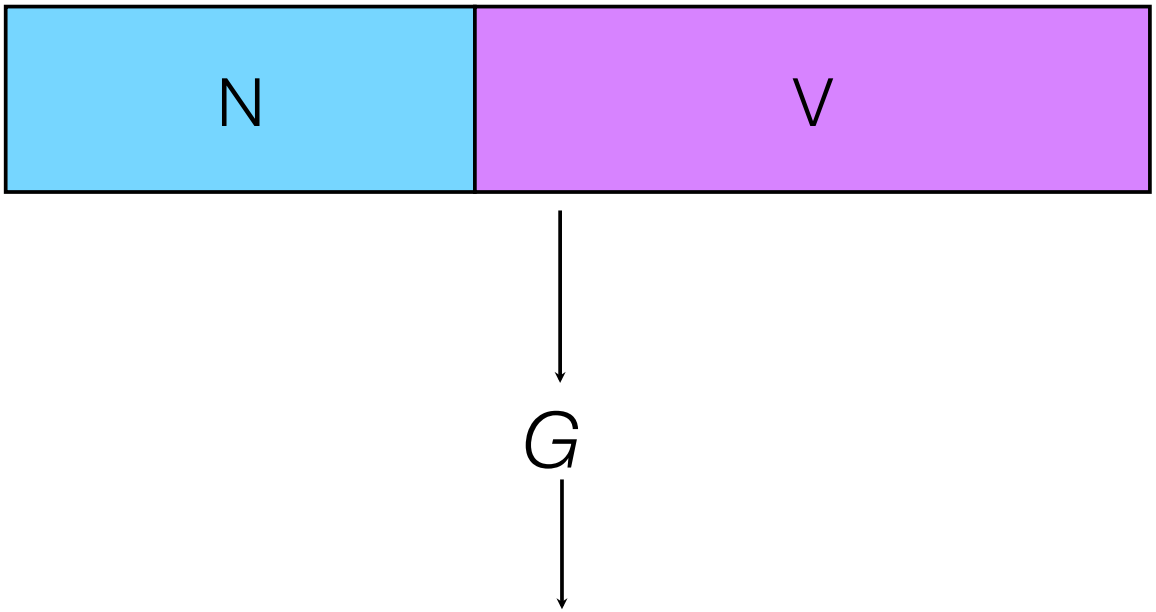
Statistical Process

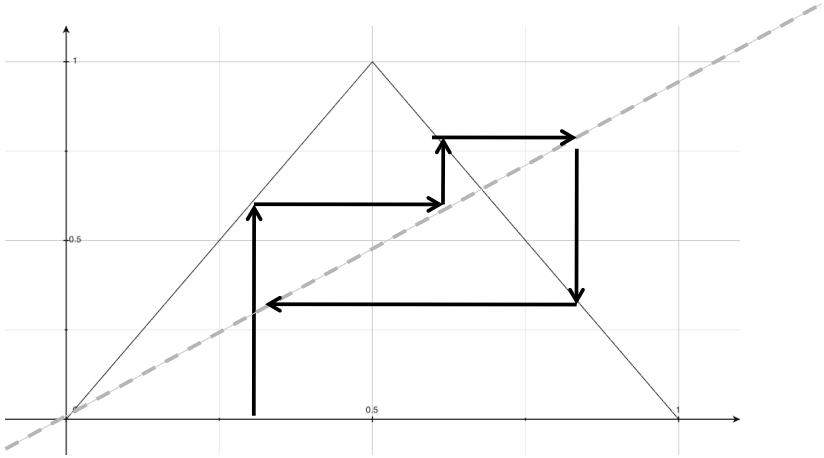
Learning a Partition

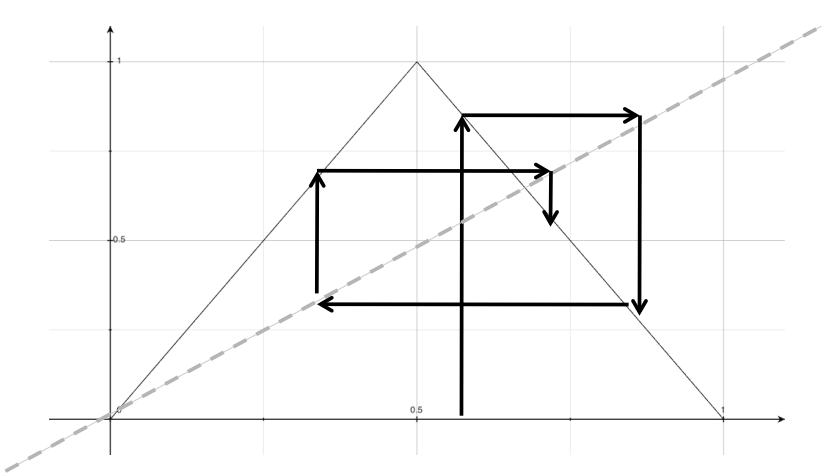
- A learner needs to discover *a* partition consistent with the partitions of other native speakers
- Most of the time, superficially similar sentences will have similar derivations
    - e.g. NVN templates
- Find partitions which minimize the difference in analyses of the most common sentence forms
- Minimize the size of each partition
- This can be done with fairly minimal computations (Kennel & Buhl, 2003)

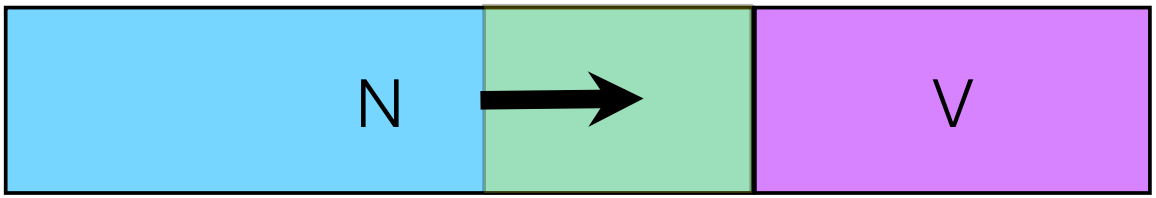
Matching Derivations



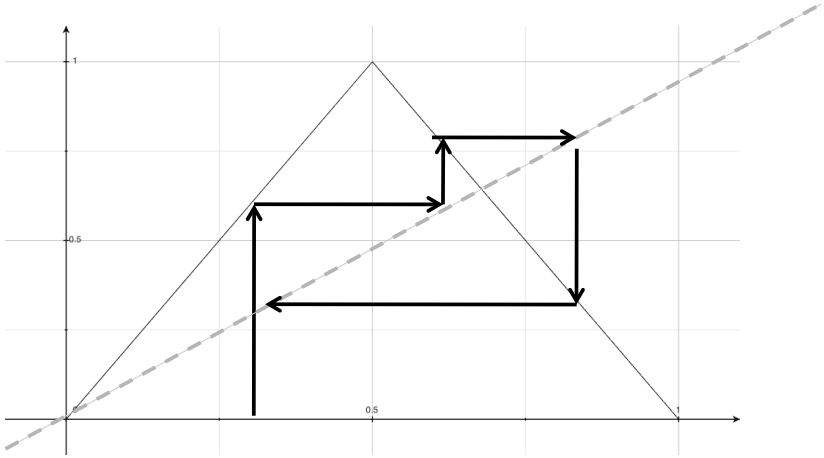

N
V
G

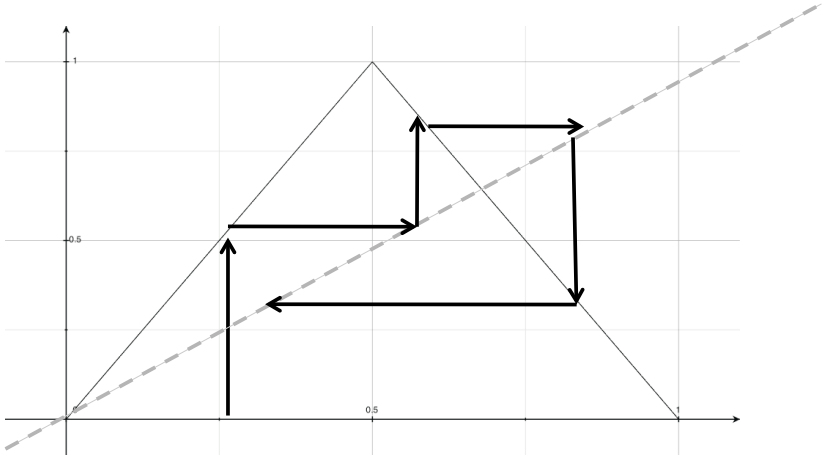

1
0.5
0
0.5
1

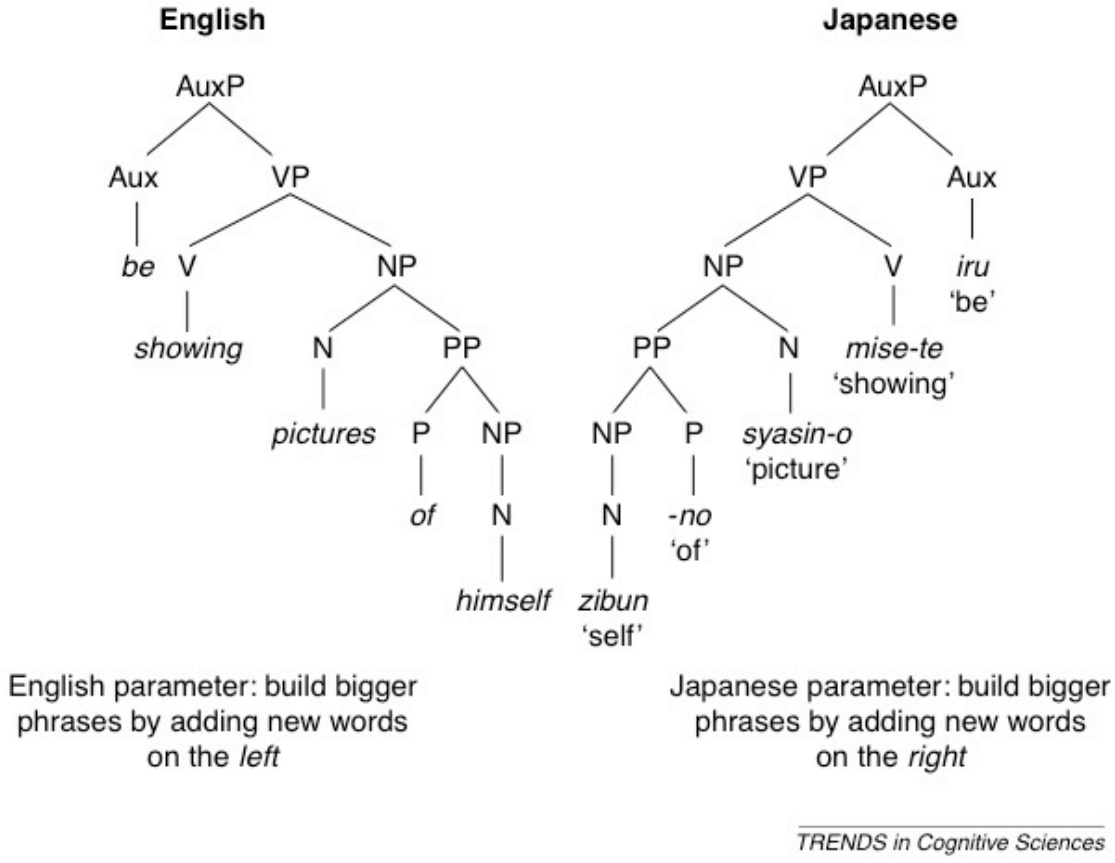

1
0.5
0
0.5
1



G







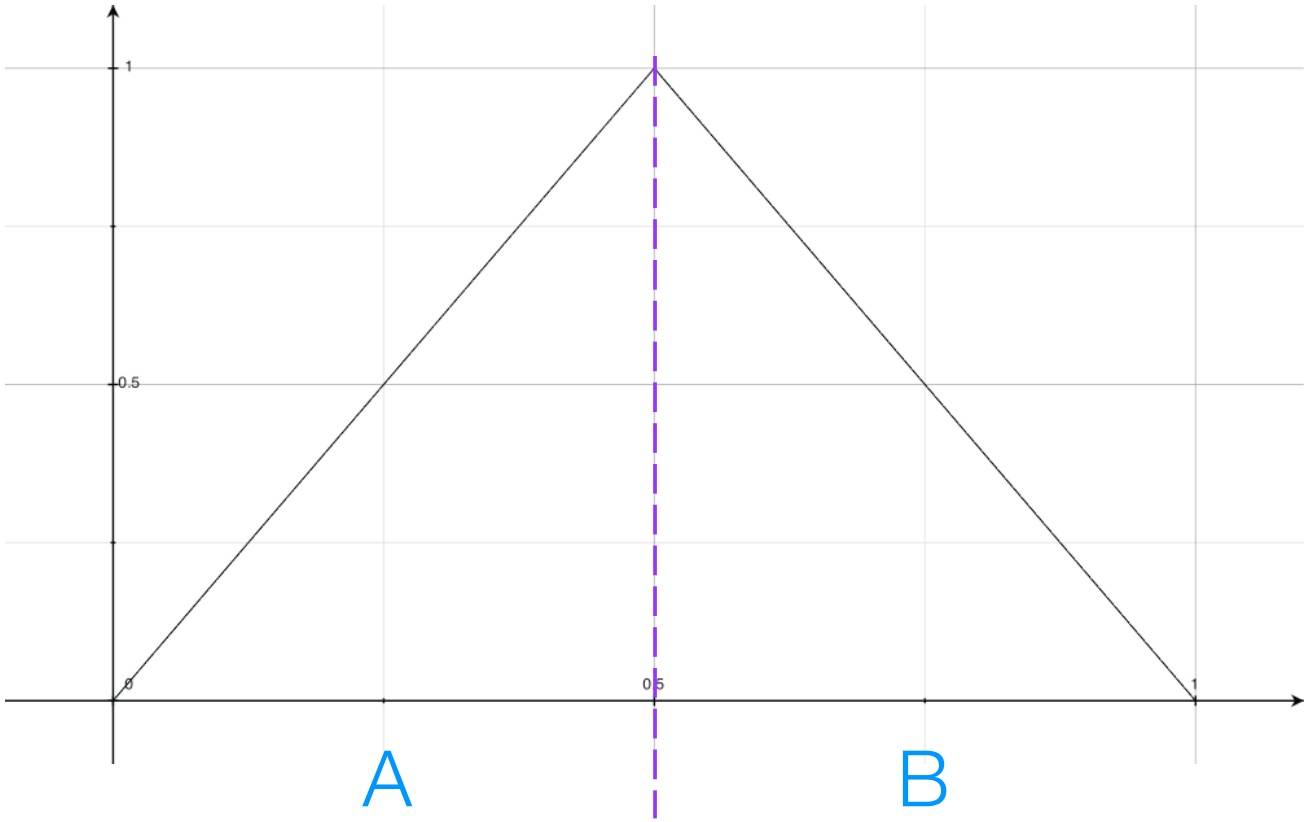
English parameter: build bigger phrases by adding new words on the *left*

Japanese parameter: build bigger phrases by adding new words on the *right*

TRENDS in Cognitive Sciences

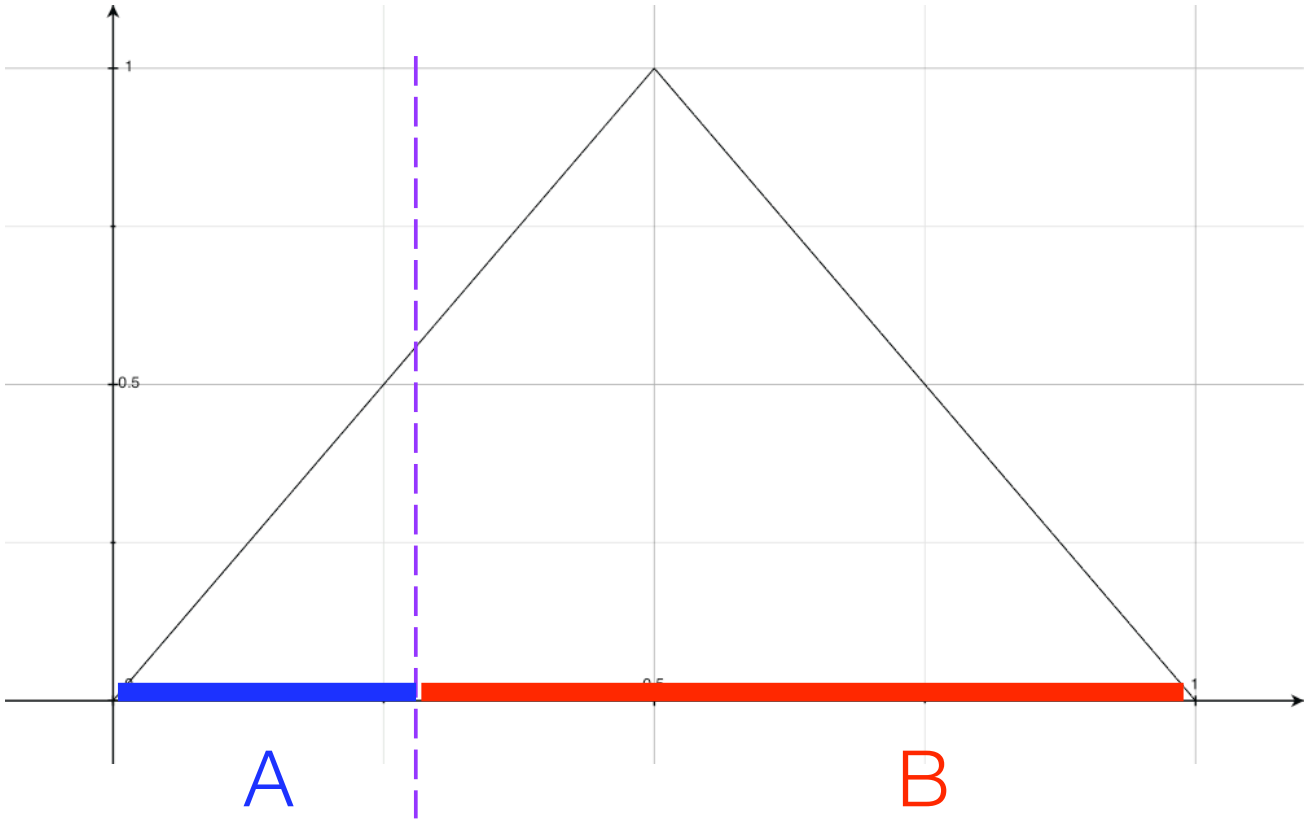
Branching Parameter

Baker (2003)



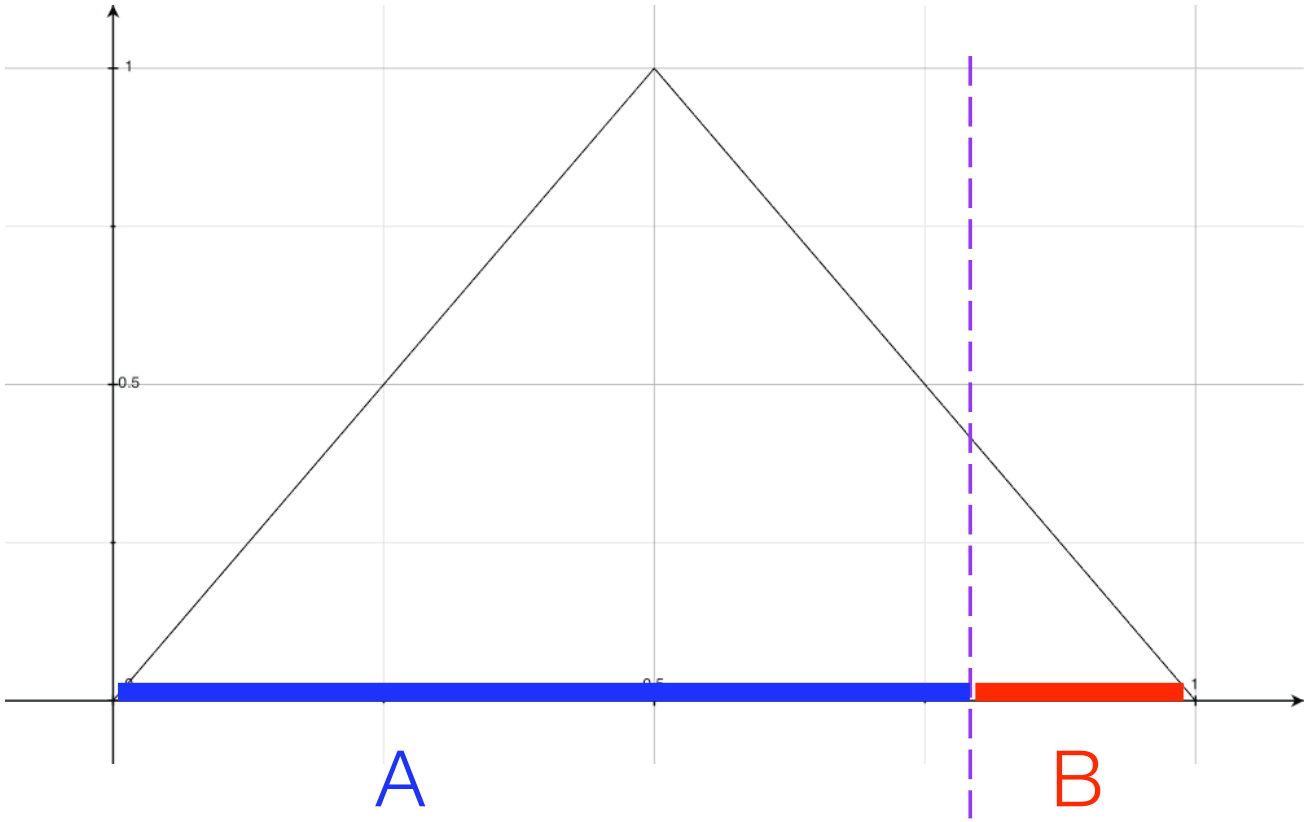
ABBA vs BAAB

Symbolic Dynamics | Tent Map



BAAB

Symbolic Dynamics | Tent Map

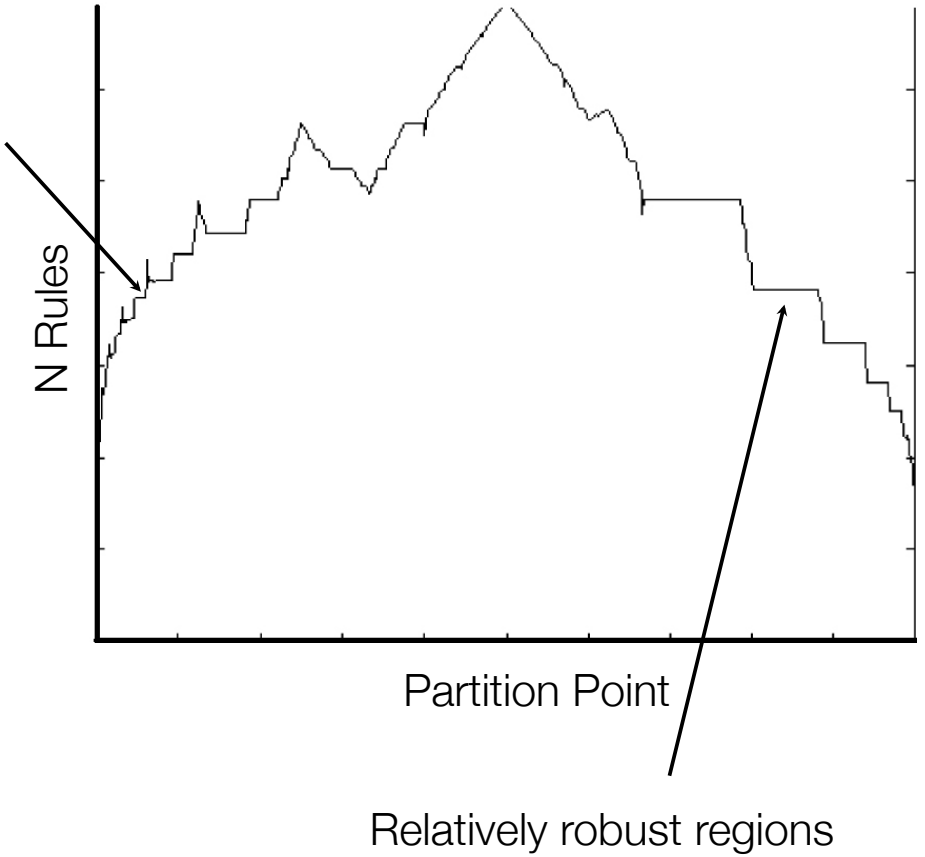


ABBA

Symbolic Dynamics | Tent Map

Non-monotonicity

- Not all partition adjustments are equal
  - Some changes can be induced by a few critical examples
  - Others require a much larger body of evidence



- These are ‘fuzzy parameters’, not strict binary settings
    - Some variance can be tolerated
    - Language change (OV -> VO)
    - Unidirectionality from topology
- The ‘parameters’ (i.e. partition points) themselves are learned, not just their values

Comparison

Present Model	Connectionism
use a fixed state space	reconstruct a state space
lightweight statistical computations; variable rates of convergence	many computations; constant, slow convergence
structures appear and disappear in a linked fashion	little relation between different SDs

Present Model	P&P
use a fixed state space	UG
statistical computations	structural analysis
no *a priori* parameters	a set of possible parameters
not all boundaries are sharp	coarse

Summary

- Parametric variation without innate parameters
- Statistically learned lexicon
- Computationally efficient
- Underlying universal structure