Language Acquisition
Fall 2010/Winter 2011

Morphology & Syntax

Afra Alishahi, Heiner Drenhaus

Computational Linguistics and Phonetics
Saarland University

Rules that Govern Form

- Moving from fixed forms (e.g. '*apple*') to derivational forms

```
play → plays, played, playing

I, you, admire → "I admire you"
```

- Morphology and syntax
  - In all languages, the formation of words and sentences follows highly regular patterns
  - How are the regulations and exceptions represented?
- The study and analysis of language production in children reveals common and persistent patterns

U-shaped Learning Curves

- Observed U-shaped learning curves in children
  - Imitation: an early phase of conservative language use
  - Generalization: general regularities are applied to new forms
  - Overgeneralization: occasional misapplication of general patterns
  - Recovery: over time, overgeneralization errors cease to happen

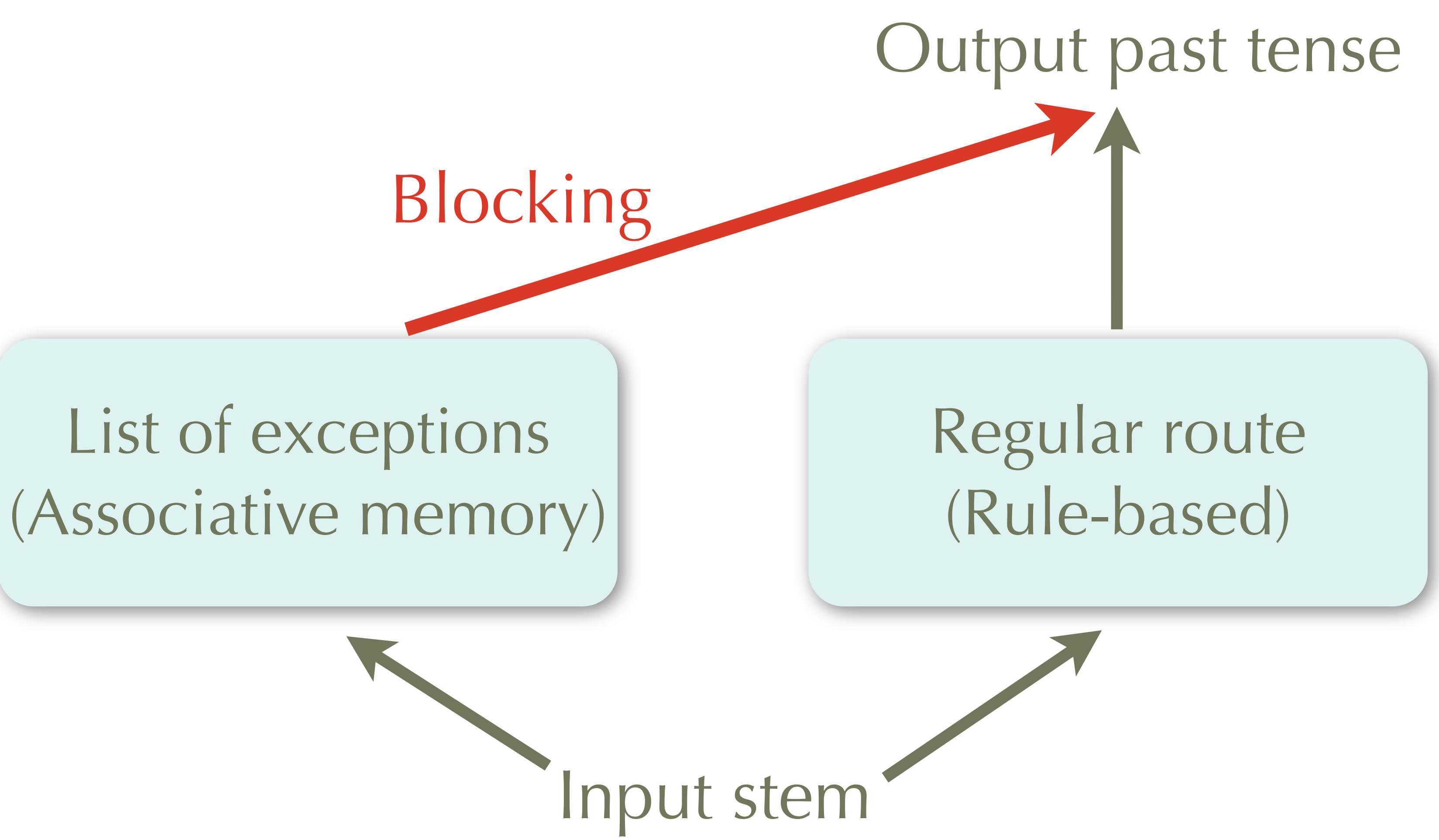
- Lack of Negative Evidence
  - Children do not receive reliable corrective feedback from parents to help them overcome their mistakes (Marcus, 1993)

Case Study: Learning English Past Tense

- The problem of English past tense formation:
  - Regular formation: `stem + 'ed'`
  - Irregulars do show some patterns
    - No-change: **`hit`** → **`hit`**
    - Vowel-change: **`ring`** → **`rang,`** **`sing`** → **`sang`**
- Over-regularizations are common: **`goed`**
  - These errors often occur after the child has already produced the correct irregular form: **`went`**
- What causes the U-shaped learning curve?

A Symbolic Account of English Past Tense

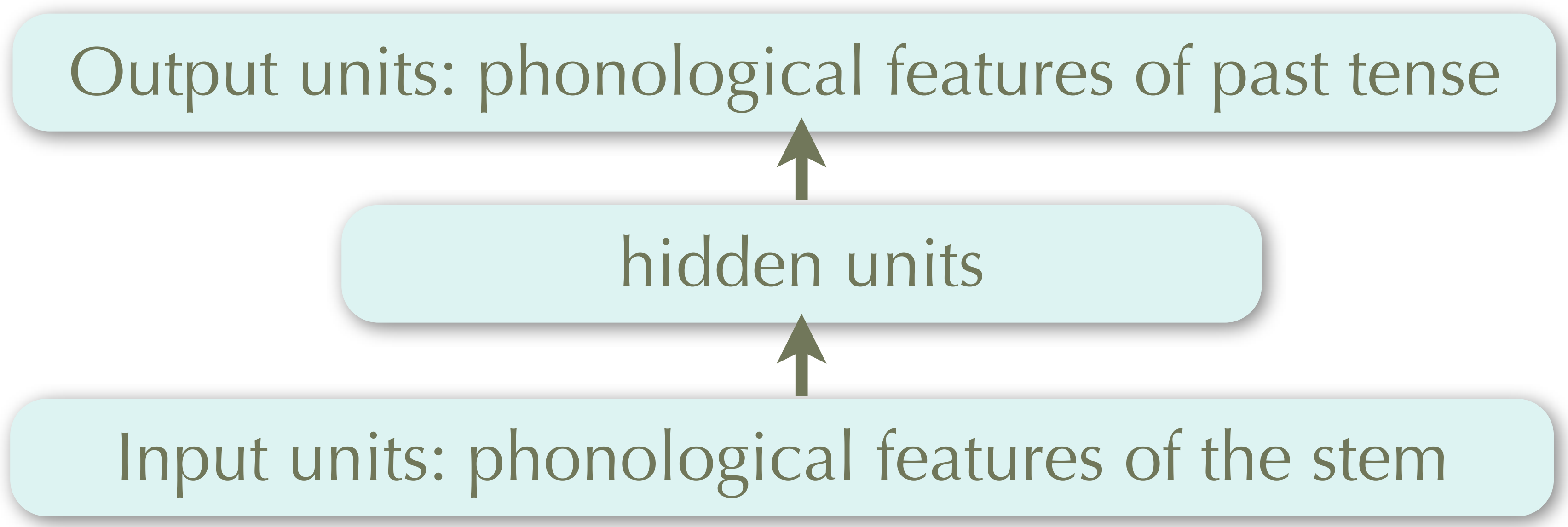
- Dual-Route Account: two qualitatively different mechanisms



- Prediction:
  - Errors result from transition from rote learning to rule-governed
  - Recovery occurs after sufficient exposure to irregulars

A Connectionist Account of Learning English Past Tense

- A connectionist model (Plunkett & Marchman, 1993)



- Properties:
  - Early in training, the model shows tendency to overgeneralize; by the end of training, it exhibits near perfect performance
  - U-shaped performance is achieved using a single learning mechanism, but depends on sudden change in the training size

A Hybrid, Analogy-based Account

- Taatgen & Anderson (2002): an rational model of learning past tense based on the ACT-R architecture
  - Declarative memory chunks represent past tenses, both as a goal and as examples

```
PAST-TENSE-GOAL23
        ISA PAST
        OF WALK
        STEM NIL
        SUFFIX NIL
```

goal to determine past tense of walk

```
PAST-TENSE-GOAL23
        ISA PAST
        OF WALK
        STEM WALK
        SUFFIX ED
```

accomplished goal, stored in the memory

A Hybrid, Analogy-based Account

- The analogy strategy is implemented by two production rules, based on simple pattern matching:

RULE ANALOGY-FILL-SLOT
IF the goal has an empty <u>suffix</u> slot
AND there is an example in which <u>suffix</u> has a value
THEN set the <u>suffix</u> of the goal to the <u>suffix</u> value of the example

RULE ANALOGY-COPY-A-SLOT
IF the goal has an empty <u>stem</u> slot and the <u>of</u> slot has a certain value
AND in the example the values of the <u>of</u> and <u>stem</u> slots are equal
THEN set the <u>stem</u> to the value of the <u>of</u> slot

ACT-R Equations

Equation	Description
Activation	
$A = B + \text{context} + \text{noise}$	The activation of a chunk has three parts: base-level activation, spreading activation from the current context and noise. Since spreading activation is a constant factor in the models discussed, we treat activation as if it were just base-level activation.
Base-level activation	
$B(t) = \log \sum_{j=1}^{n} (t - t_j)^{-d}$	n is the number of times a chunk has been retrieved from memory, and t_j represents the time at which each of these retrievals took place. So, the longer ago a retrieval was, the less it contributes to the activation. d is a fixed ACT-R parameter that represents the decay of base-level activation in declarative memory.
Retrieval time	
$\text{Time} = F\text{e}^{-fA}$	Activation determines the time required to retrieve a chunk. A is the activation of the chunk that has to be retrieved, and F and f are fixed ACT-R parameters. Retrieval will only succeed as long as the activation is larger than retrieval threshold τ, which is also a fixed parameter.
Expected outcome	
$\text{Expected outcome} = P_{\text{p}}G - C_{\text{p}} + \text{noise}$	Expected outcome is based on three quantities, the estimated probability of success of a production rule (P), the estimated cost of the production rule (C), and the value of the goal (G).

A Hybrid, Analogy-based Account

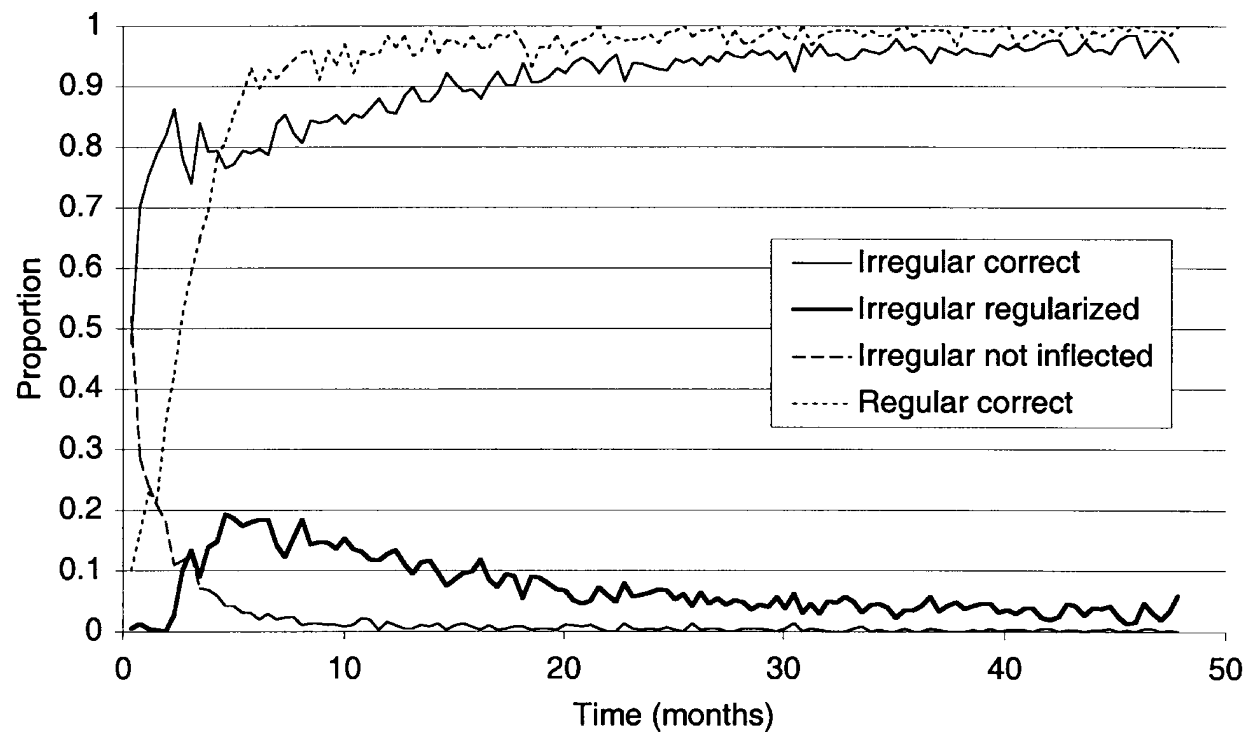
- ACT-R's production rule mechanism learns new rules by combining two rules that have fired consecutively into one:

RULE LEARNED-REGULAR-RULE

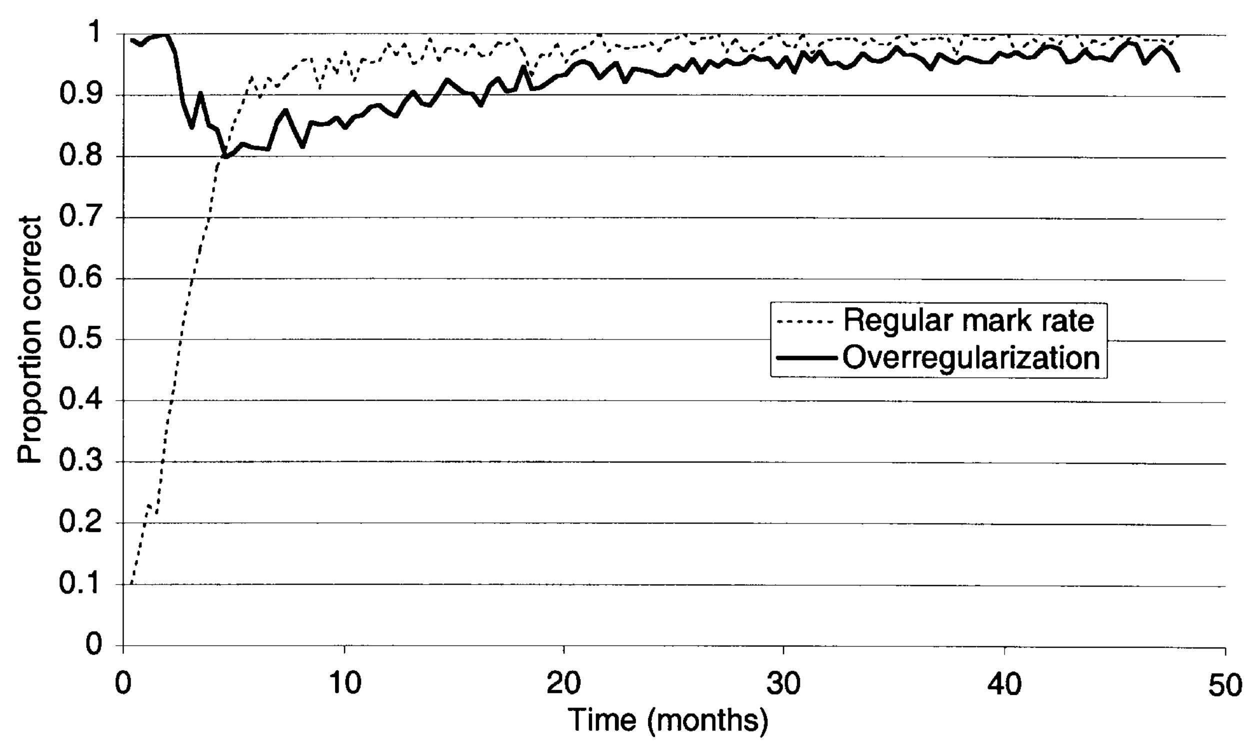
IF the goal is to find the past tense of a word and slots <u>stem</u> and <u>suffix</u> are empty

THEN set the <u>suffix</u> slot to ED and set the <u>stem</u> slot to the word of which you want the past tense

A Hybrid, Analogy-based Account



(a)



(b)

Innateness of Language

- Central claim: humans have innate knowledge of language
  - Assumption: all languages have a common structural basis
- Argument from the Poverty of the Stimulus (Chomsky 1965)
  - Linguistic experience of children is not sufficiently rich for learning the grammar of the language, hence they must have some innate specification of grammar
  - Assumption: knowing a language involves knowing a grammar
- Universal Grammar (UG)
  - A set of rules which organize language in the human brain

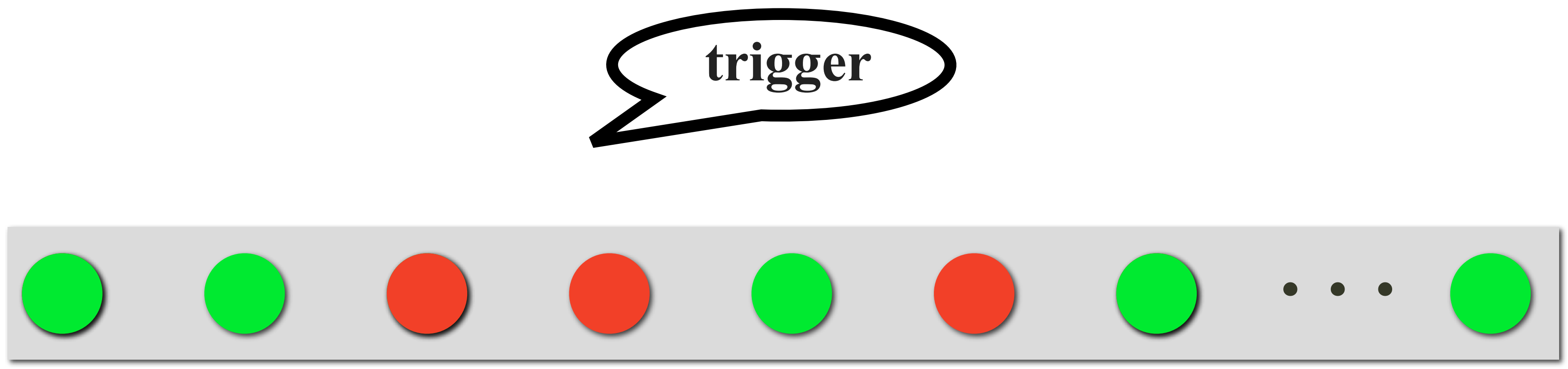
Principles & Parameters

- A framework for representing UG
  - A finite set of fundamental principles that are common to all languages
    - E.g., "a sentence must have a subject"
  - A finite set of parameters that determine syntactic variability amongst languages
    - E.g., a binary parameter that determines whether the subject of a sentence must be overtly pronounced
  - Learning involves identifying the correct grammar
    - I.e., setting UG parameters to proper values for the current language

Computational Implementation of P&P

- Formal parameter setting models for a small set of grammars
  - Clark 1992, Gibson & Wexler 1994, Niyogi & Berwick 1996, Briscoe 2000
- General approach:
  - Analyze current input string and set the parameters accordingly
  - Set a parameter when receiving evidence from an example which exhibits that parameter (trigger)
- Representative models:
  - Triggering Learning Algorithm or TLA [Gibson & Wexler, 1994]
  - Structural Triggers Learner or STL [Fodor, 1998]
  - Variational Learner or VL [Yang, 2002]

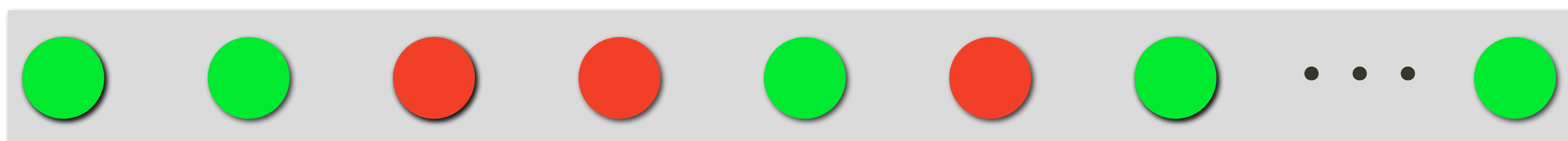
Computational Implementation of P&P



- TLA: randomly modifies a parameter value if it cannot parse the input
- STL: learns sub-trees (treelets) as parameter values
- VL: assigns a weight to each parameter, and rewards or penalizes these weights depending on parsing success

Computational Implementation of P&P

What if it is ambiguous?



- TLA: chooses one of the possible interpretations of the ambiguous trigger
- STL: ignores ambiguous triggers and waits for unambiguous ones
- VL: each interpretation is parsed and the parameter weights are changed accordingly

Computational Challenges of P&P

- Practical limitations:
  - Formalizing a UG that covers existing languages is a challenge
  - Learning relies on well-formed sentences as input
- P&P framework predicts a huge space of possible grammars
  - 20 binary parameters lead to > 1 million grammars
- Search spaces for a grammar contain local maxima
  - I.e. learner may converge to an incorrect grammar
- Most of the P&P models are psychologically implausible
  - They predict that a child may repeatedly revisit the same hypothesis or jump randomly around the hypothesis space

Usage-based Accounts of Language Acquisition

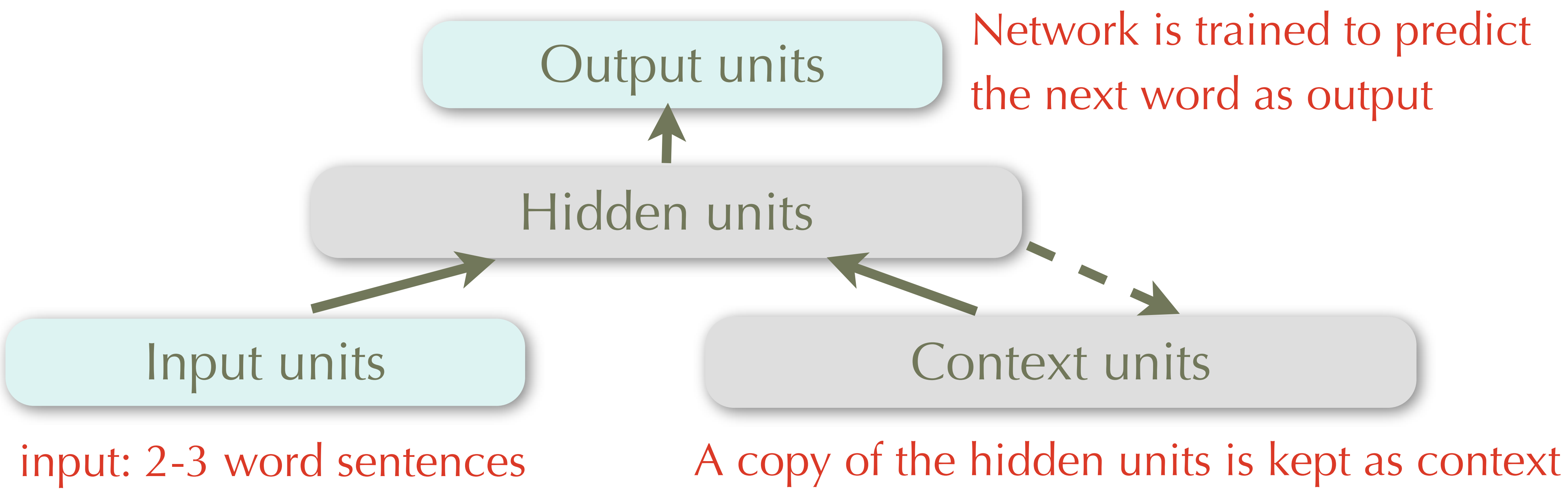
- Main claims:
  - Children learn language regularities from input alone, without guidance from innate principles
  - Mechanisms of language learning are not domain-specific
- Verb Island Hypothesis (Tomasello, 1992)
  - Children build their linguistic knowledge around individual items rather than adjusting general grammar rules they already possess
  - Children use cognitive processes to gradually categorize the syntactic structure of their item-based constructions
  - General-purpose cognitive tools are used for this purpose: imitation, analogy, structure mapping

Distributional Representation as an Alternative to Grammar

- Knowing a language is <u>not</u> equated with knowing a grammar
  - Knowledge of language is developed to perform communicative tasks of comprehension and production
- Neural networks for language representation and acquisition
  - Different levels of linguistic representation are emergent structures that a network develops in the course of learning
  - E.g., Elman (1990, 1991), Allen (1997), Allen & Seidenberg (1999)

Case Study: Elman (1990)

- A model of leaning lexical classes and word order



Word Categories

Category	Examples
NOUN-HUM	man, woman
NOUN-ANIM	cat, mouse
NOUN-INANIM	book, rock
NOUN-AGRESS	dragon, monster
NOUN-FRAG	glass, plate
NOUN-FOOD	cookie, break
VERB-INTRAN	think, sleep
VERB-TRAN	see, chase
VERB-AGPAT	move, break
VERB-PERCEPT	smell, see
VERB-DESTROY	break, smash
VERB-EAT	eat

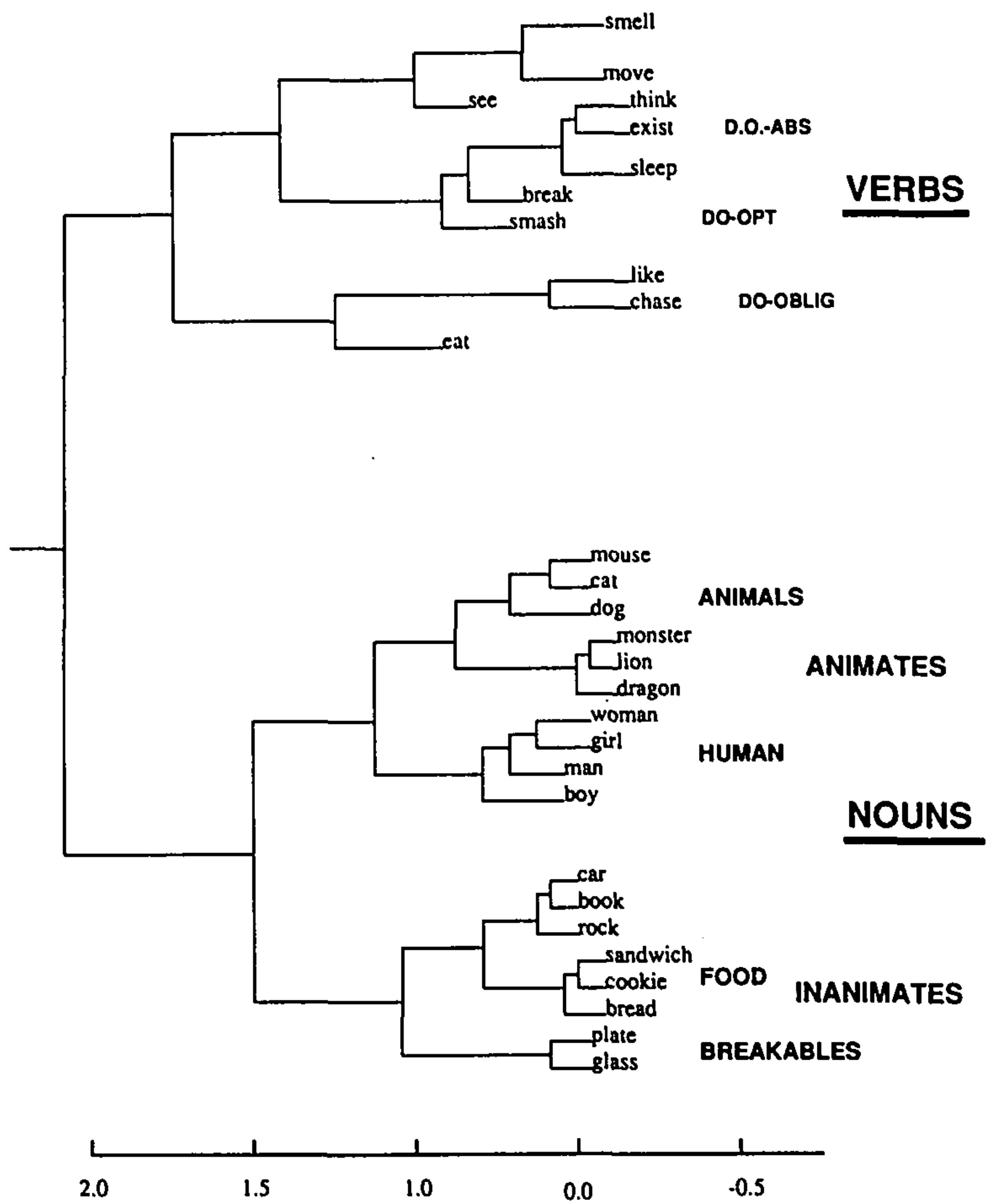
Templates for Sentence Generation

WORD 1	WORD 2	WORD 3
NOUN-HUM	VERB-EAT	NOUN-FOOD
NOUN-HUM	VERB-PERCEPT	NOUN-INANIM
NOUN-HUM	VERB-DESTROY	NOUN-FRAG
NOUN-HUM	VERB-INTRAN	
NOUN-HUM	VERB-TRAN	NOUN-HUM
NOUN-HUM	VERB-AGPAT	NOUN-INANIM
NOUN-HUM	VERB-AGPAT	
NOUN-ANIM	VERB-EAT	NOUN-FOOD
NOUN-ANIM	VERB-TRAN	NOUN-ANIM
NOUN-ANIM	VERB-AGPAT	NOUN-INANIM
NOUN-ANIM	VERB-AGPAT	
NOUN-INANIM	VERB-AGPAT	
NOUN-AGRESS	VERB-DESTROY	NOUN-FRAG
NOUN-AGRESS	VERB-EAT	NOUN-HUM
NOUN-AGRESS	VERB-EAT	NOUN-ANIM
NOUN-AGRESS	VERB-EAT	NOUN-FOOD

Sample Training Sequence

Input	Output
0000000000000000000000000000010 (*woman*)	0000000000000000000000000010000 (*smash*)
0000000000000000000000000010000 (*smash*)	0000000000000000000001000000000 (*plate*)
0000000000000000000001000000000 (*plate*)	0000010000000000000000000000000 (*cat*)
0000010000000000000000000000000 (*cat*)	0000000000000000000100000000000 (*move*)
0000000000000000000100000000000 (*move*)	0000000000000000100000000000000 (*man*)
0000000000000000100000000000000 (*man*)	0001000000000000000000000000000 (*break*)
0001000000000000000000000000000 (*break*)	0000100000000000000000000000000 (*car*)
0000100000000000000000000000000 (*car*)	0100000000000000000000000000000 (*boy*)
0100000000000000000000000000000 (*boy*)	0000000000000000000100000000000 (*move*)
0000000000000000000100000000000 (*move*)	0000000000001000000000000000000 (*girl*)
0000000000001000000000000000000 (*girl*)	0000000000100000000000000000000 (*eat*)
0000000000100000000000000000000 (*eat*)	0010000000000000000000000000000 (*bread*)
0010000000000000000000000000000 (*bread*)	0000000010000000000000000000000 (*dog*)
0000000010000000000000000000000 (*dog*)	0000000000000000000100000000000 (*move*)
0000000000000000000100000000000 (*move*)	0000000000000000001000000000000 (*mouse*)
0000000000000000001000000000000 (*mouse*)	0000000000000000001000000000000 (*mouse*)
0000000000000000001000000000000 (*mouse*)	0000000000000000000100000000000 (*move*)
0000000000000000000100000000000 (*move*)	1000000000000000000000000000000 (*book*)
1000000000000000000000000000000 (*book*)	0000000000000001000000000000000 (*lion*)

Analysis of Hidden Unit Activation Patterns

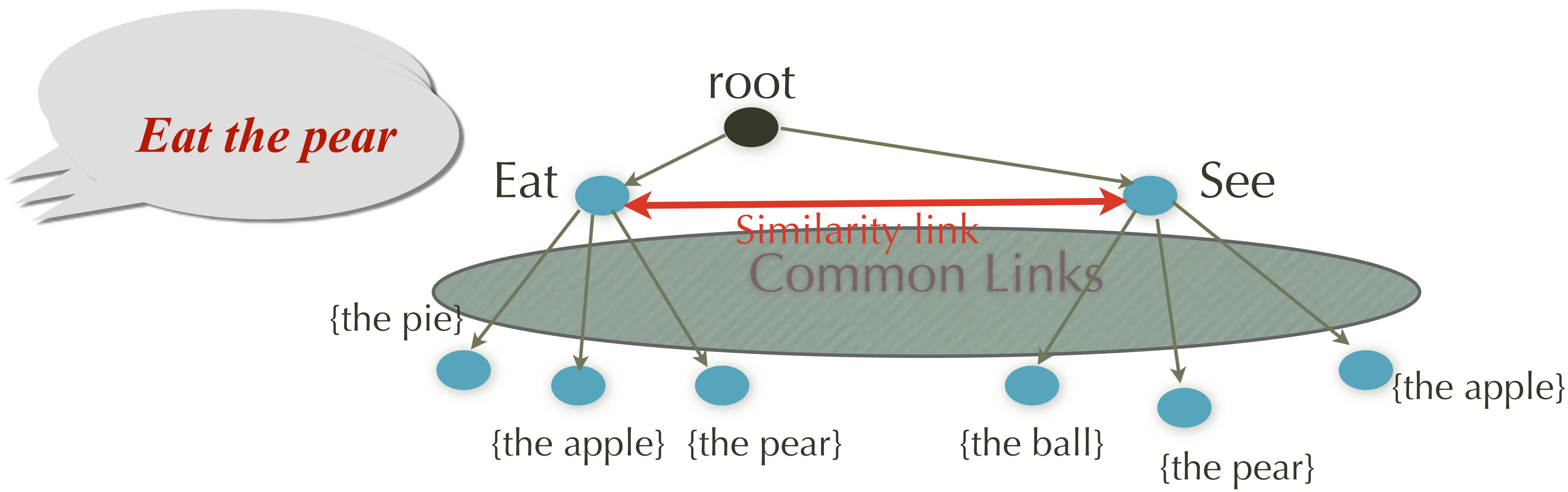


Learning Grammar from Corpora

- Many computational models show the possibility of learning a grammar from corpus data
  - Machine learning techniques induce a grammar that fits data
  - Jones, Gobet, & Pine (2000), Clark (2001), Gobet, Freudenthal, & Pine (2004), Solan, Horn, Ruppin & Edelman (2004)
- Common properties:
  - Most of these models are not incremental
  - They mostly focus on the acquisition of syntax (usually a CFG), but not semantics

Case Study: MOSAIC (Jones et al., 2000)

- MOSAIC (Model Of Syntax Acquisition In Children; Jones et al 2000)
    - Learns from raw text, and produces utterances similar to what children produce using a discrimination network



Case Study: MOSAIC (Jones et al., 2000)

- Underlying mechanisms
  - Learning: expand the network based on input data
  - production: traverse the network and output contents of the nodes
- Generalization
  - Generative links allow limited generalization abilities
  - Lack of semantic knowledge prevents meaningful generalization
  - Generalized sentences are limited to high-frequency terms
- Evaluation
  - The model was trained on a subset of CHILDES
  - It was used to simulate verb island phenomenon, optional infinitive in English, subject omission, ...