

Language Evolution and the Appearance of Design Argument

Willem Zuidema¹²

¹ Language Evolution & Computation, University of Edinburgh

² Institute for Cell, Animal and Population Biology, University of Edinburgh





I-language

production

I-language

interpretation
acquisition

E-language



What is special about language?

(Pinker & Jackendoff, forthcoming in *Cognition*)

Conceptual structure (theory of mind, language dependent concepts)

Speech perception (double dissociations, subjective similarity space different from primates)

Speech production (permanently lowered larynx, anatomical changes in tongue & jaw, neural control architecture)

Phonology (distinct level of organization, vocal imitation, no combinatorial phonology in primates, no rhythm in bird song)

Words (≥ 50.000 , Saussurean, relation to other words)

Morphosyntax (phrases, ordering, agreement, case-marking)

not in the primary data \neq innate

special \neq adaptation

adaptation \neq explanation

Learning algorithm

Context-free grammars

- Rules of the forms: $A \mapsto t$, $A \mapsto BC$, $A \mapsto Bt$
- Start symbol S , terminal symbols (lexicon), non-terminal symbols

Lexical

$S \mapsto$ mary shouts
 $S \mapsto$ angry mary shouts
 $S \mapsto$...

Combinatorial

$S \mapsto$ N V
 $S \mapsto$ N walks
 $N \mapsto$ mary
 $N \mapsto$ angry mary
 $N \mapsto$...
 $V \mapsto$ shouts

Recursive

$S \mapsto$ N V
 $N \mapsto$ A N
 $N \mapsto$ mary
 $A \mapsto$ beautiful
 $A \mapsto$ angry
 $V \mapsto$ walks

Learning algorithm

Incorporation: extend the language, such that it includes the encountered string

Compression: substitute frequent and long substrings with a nonterminal (the grammar becomes smaller and the language remains unchanged)

Generalization: equate two nonterminals if they occur frequently in the same context

(Wolff, 1983; Stolcke, 1994; Kirby, 2000; Zuidema 2002)

Example

Training sentences: abcabcabcd, abcabcabcabcd, abcd, abcabcd

(a) Incorporation

S \mapsto abcabcabcd
S \mapsto abcabcabcabcd
S \mapsto abcd
S \mapsto abcabcd

(b) Compression

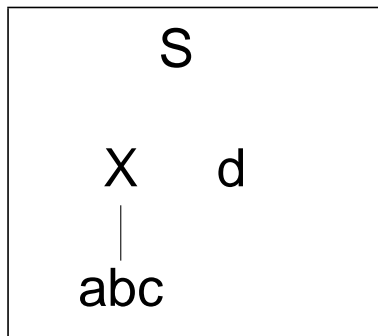
S \mapsto Xabcd
S \mapsto Xabcabcd
S \mapsto Yd
S \mapsto Xd
X \mapsto YY
Y \mapsto abc

(c) Generalization

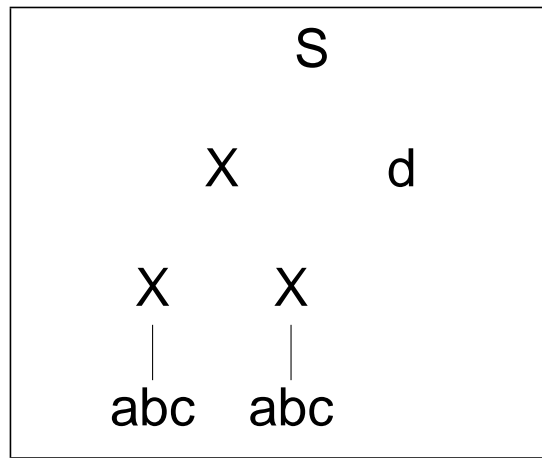
S \mapsto Xabcd
S \mapsto Xabcabcd
S \mapsto Xd
X \mapsto XX
X \mapsto abc

Nurture nor nature

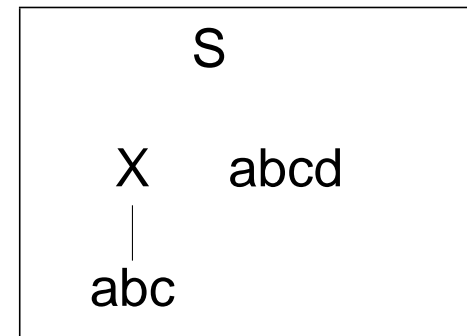
$S \mapsto Xabcd$, $S \mapsto Xabcabcd$, $S \mapsto Xd$, $X \mapsto XX$, $X \mapsto abc$



abcd



abcabcd

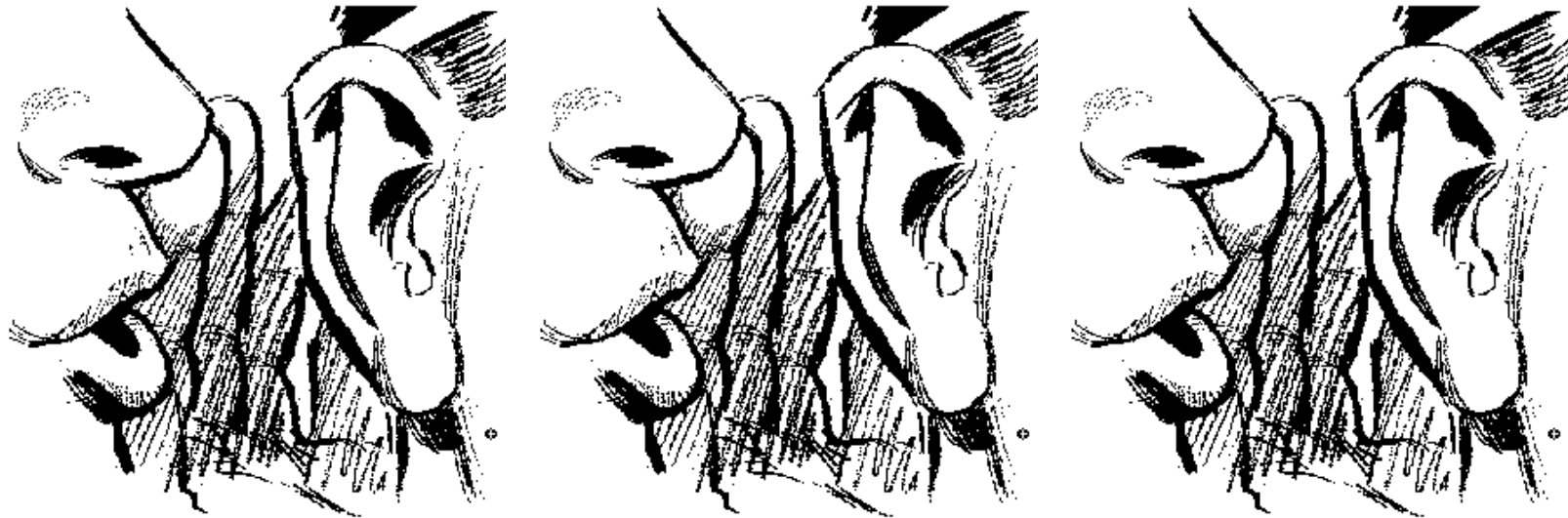


abcabcd

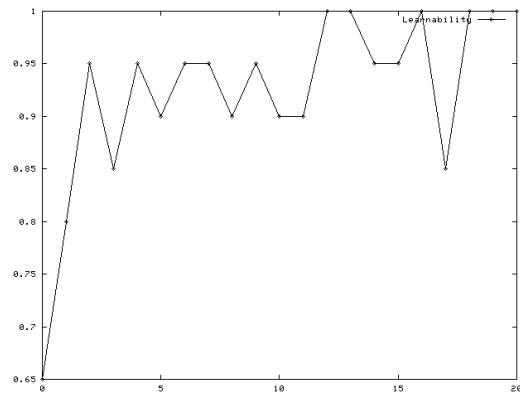
Structural ambiguity! But this property is neither in the data, nor in the algorithm; it results from the interaction between the two.

Iterated Learning

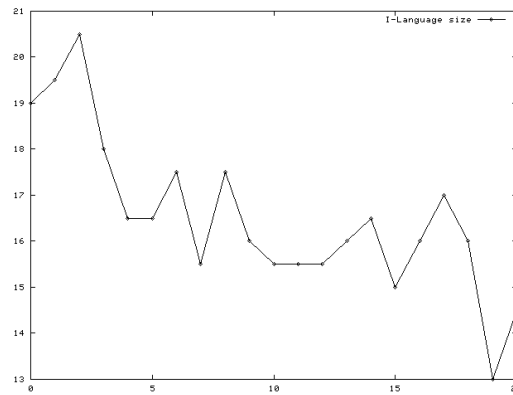
Individuals in a *chain* learn from the previous individual and teach to the next



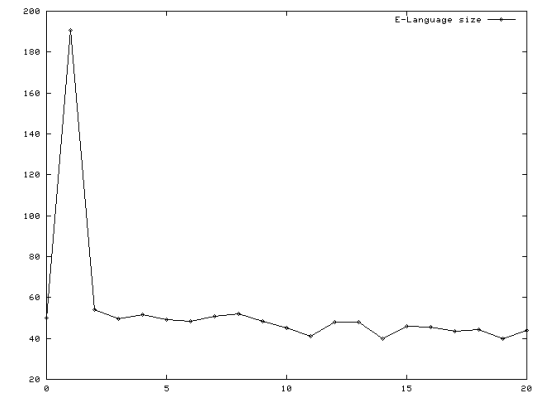
Iterated Learning - results



learnability



number of rules



number of sentences

Parameters: $V_t = \{a, b, c, d\}$, $V_{nt} = \{S, X, Y, Z, A, B, C\}$, $T=30$, $E=20$, $l_0=3$.
Shown are the average values of 2 simulations.

Example

1. “ada”, “ddac”, “adba”, “bcbd”, “cdca”
2. “dcac”, “bcac”, “caac”, “daac”
3. $S \mapsto dcX$, $S \mapsto bcX$, $S \mapsto caX$, $S \mapsto daX$, and $X \mapsto ac$
4. $Y \mapsto b$
5. “dcac”, “bcac”, “caac”, “bcb”, “cab”, “dab”,

Language Adaptation

1. Languages are transmitted culturally and are subject to change
2. Languages will change more if they are difficult to learn
3. Over time, languages that are easy to learn are more likely to occur

(Christiansen, 1995; Deacon, 1996; Kirby 2000)

Conceptual structure (theory of mind, language dependent concepts)

Speech perception (double dissociations, subjective similarity space different from primates)

Speech production (permanently lowered larynx, anatomical changes in tongue & jaw, neural control architecture)

Phonology (distinct level of organization, vocal imitation, no combinatorial phonology in primates, no rhythm in bird song)

Words (≥ 50.000 , Saussurean, relation to other words)

Morphosyntax (phrases, ordering, agreement, case-marking)

Speech Perception

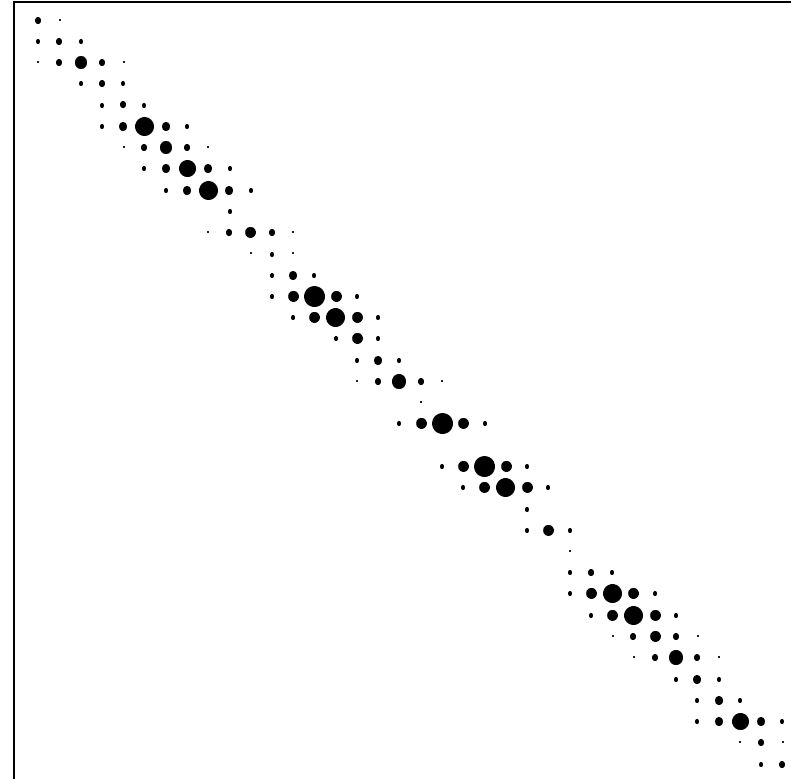
Confusion probabilities

U	signal 1	signal 2	...	signal n
signal 1	0.8	0.1	...	0.0
signal 2	0.05	0.9	...	0.0
...
signal n	0.0	0.0	...	0.7

Production probabilities:

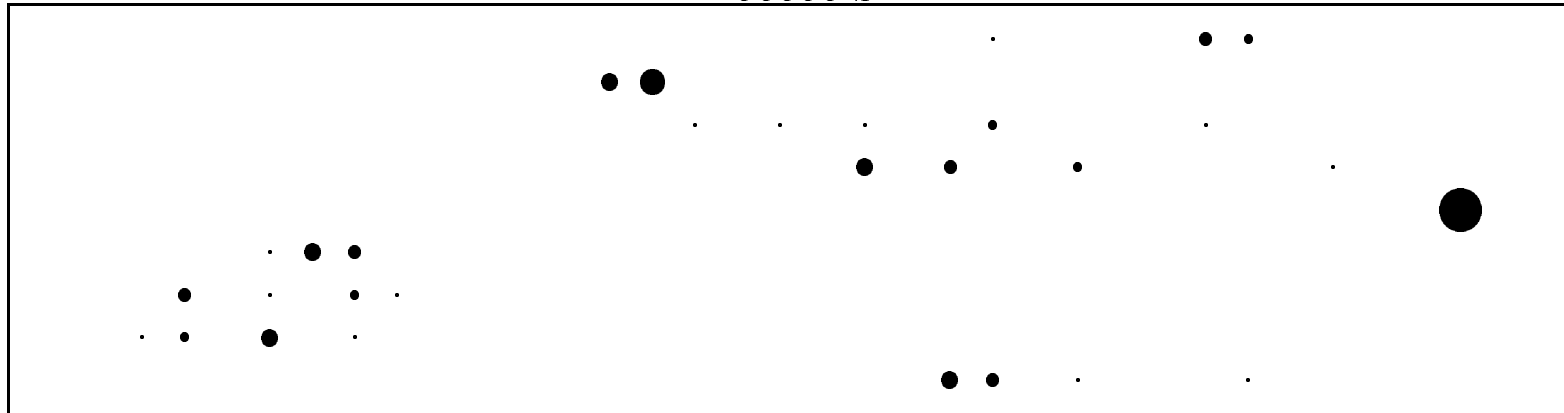
S	signal 1	signal 2	...	signal n
meaning 1	0.1	0.8	...	0.0
meaning 2	0.0	0.0	...	1.0
...
meaning n	0.9	0.0	...	0.0

U =

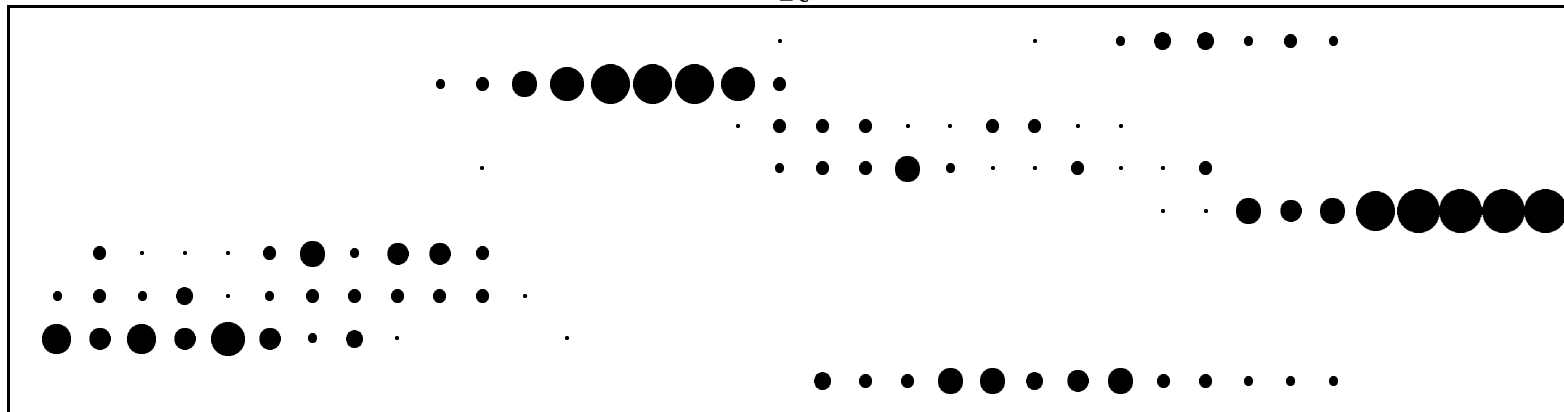


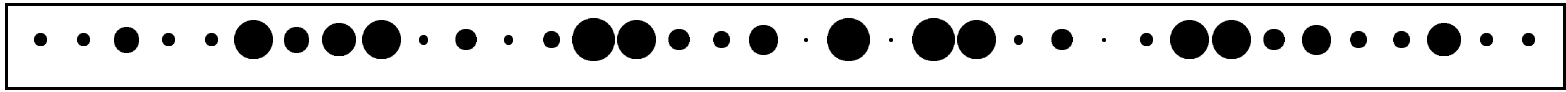


100000 S

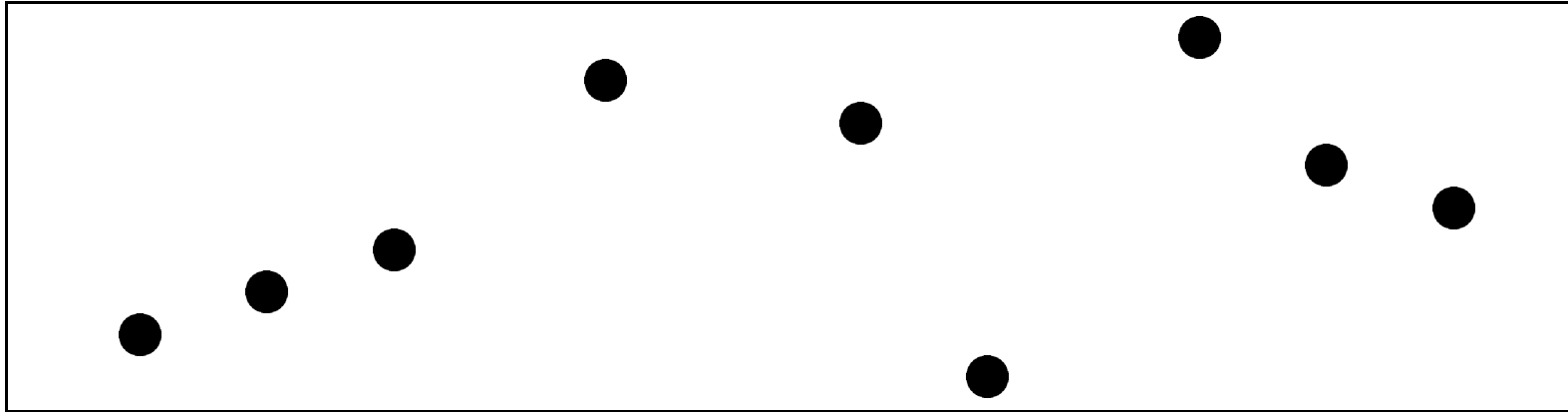


R

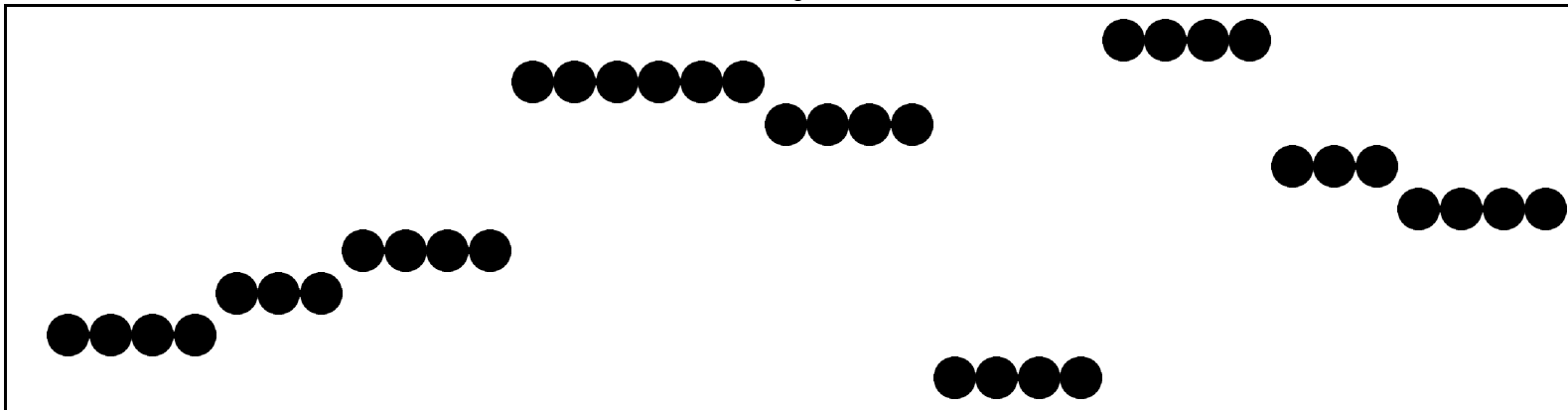


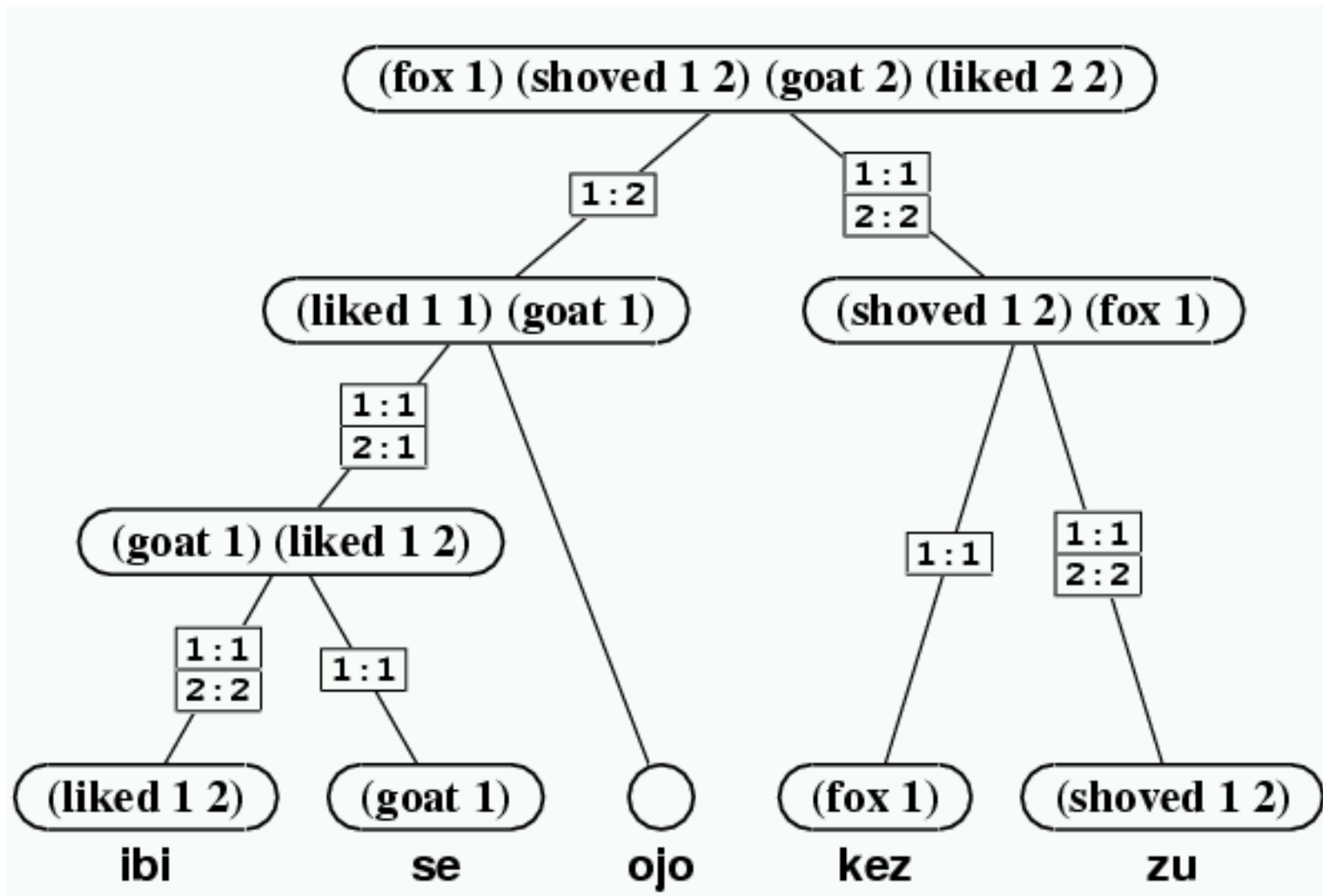


100000 S:



R:





(Batali, 2002)