Intelligent Control and Cognitive Systems

brings you...

Culture & Language in Cognitive Systems

Joanna J. Bryson

University of Bath, United Kingdom

Outline

- What is culture for? (computationally)
  - Why are we social?
  - Why do we communicate?
- Language as a special case:
  - Phonetics/phonology/morphology, Syntax, Semantics, & Pragmatics.
- Natural Language Processing (NLP)

Outline

- What is culture for? (computationally)
  - Why are we social?
  - Why do we communicate?
- Language as a special case:
  - Phonetics/phonology/morphology, Syntax, Semantics, & Pragmatics.
- Natural Language Processing (NLP)

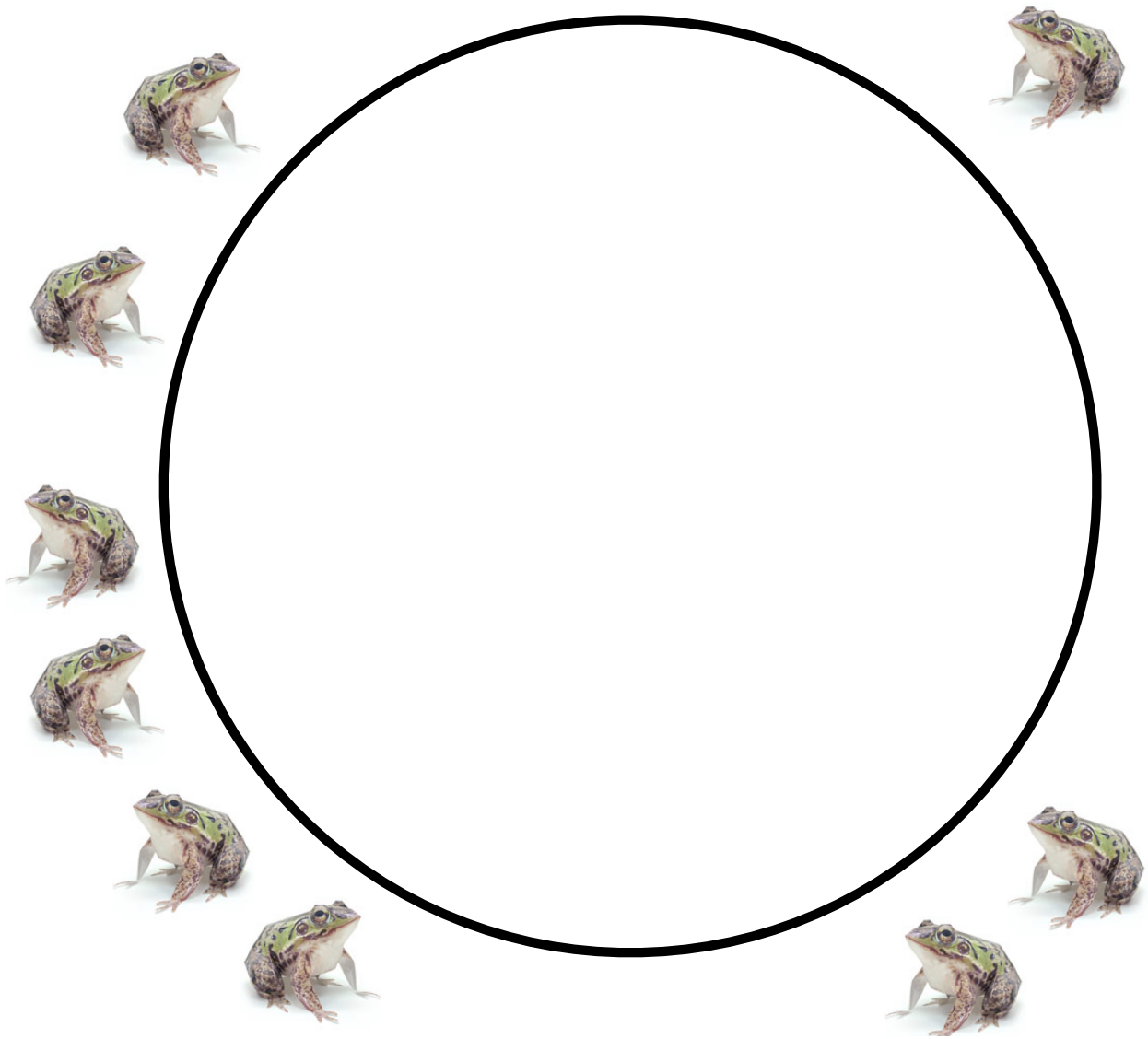

Sociality

Why not be social?

- Disease & parasites.
- Competition for food, shelter, mates.
    - Time spent maintaining social structure.

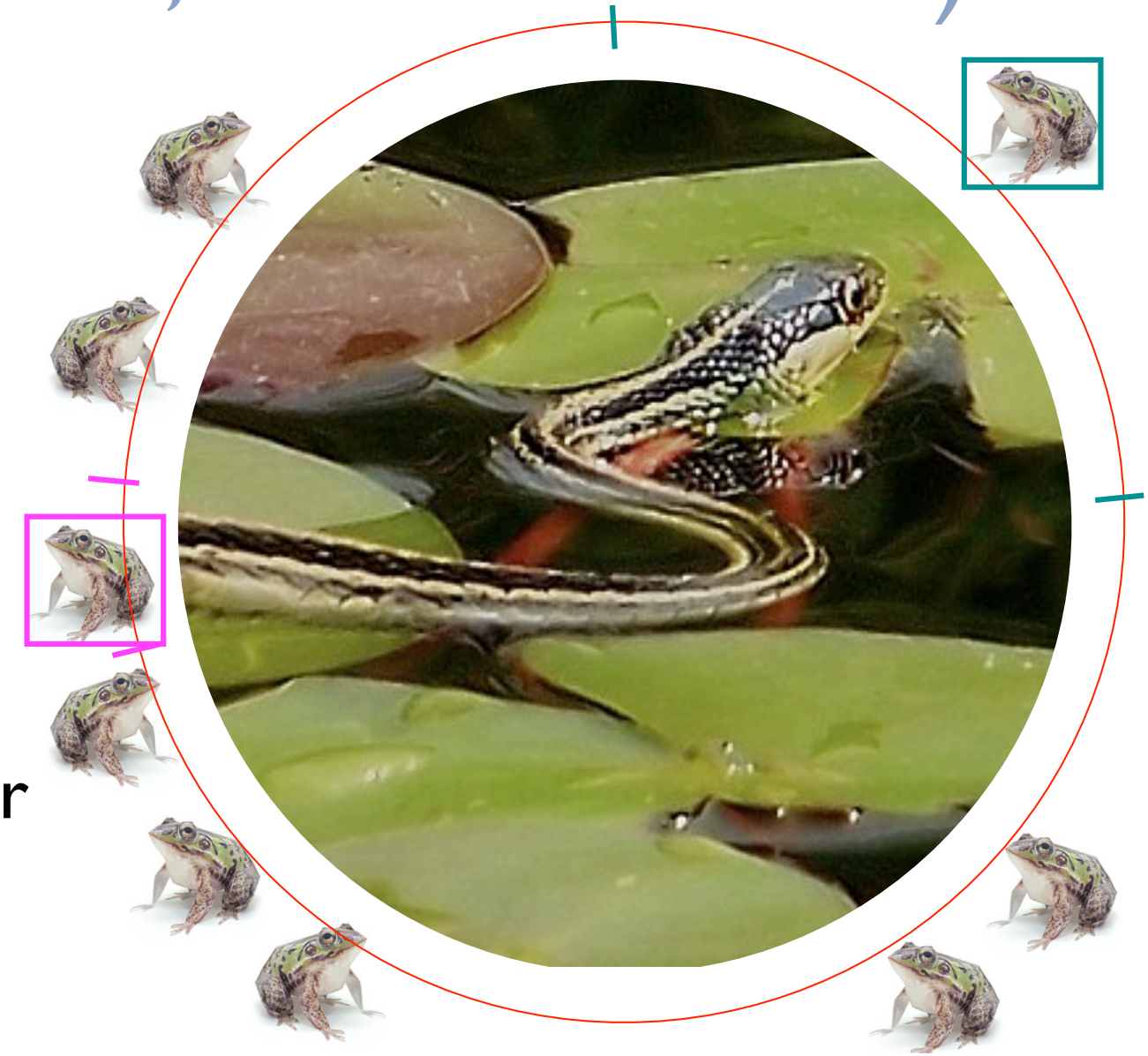
Traditional Explanation (Galton 1871, Hamilton 1973)

- Aggregation as a form of cover seeking.



Traditional Explanation (Galton 1871, Hamilton 1973)

- Aggregation as a form of cover seeking.
- Isolation increases probability of being near a predator.

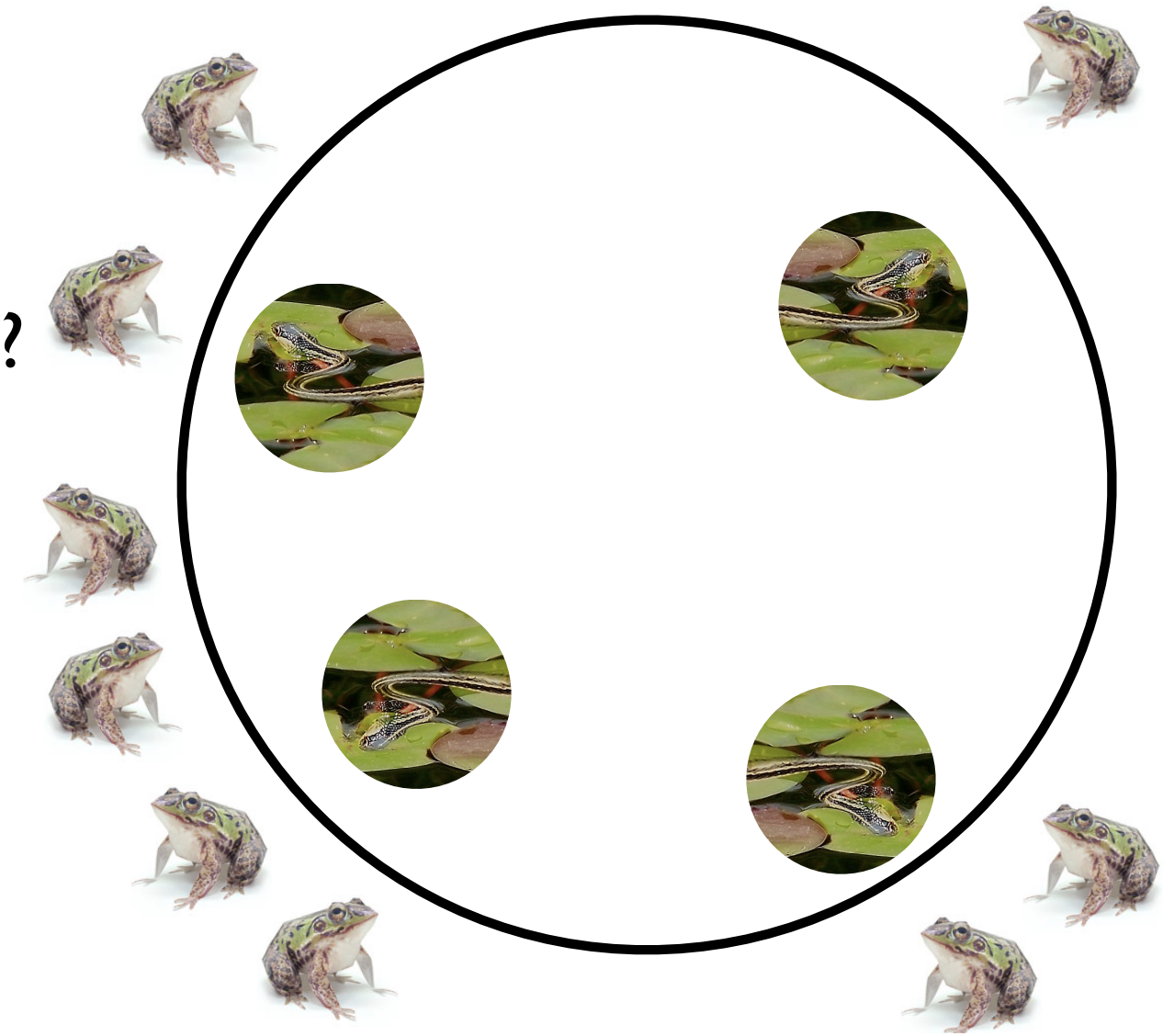


Why not be social?

- Disease & parasites.
- Competition for food, shelter, mates.
    - Time spent maintaining social structure.

Traditional Explanation (Galton 1871, Hamilton 1973)

- Aggregation as a form of cover seeking?
- Aren't predators a form of parasite?

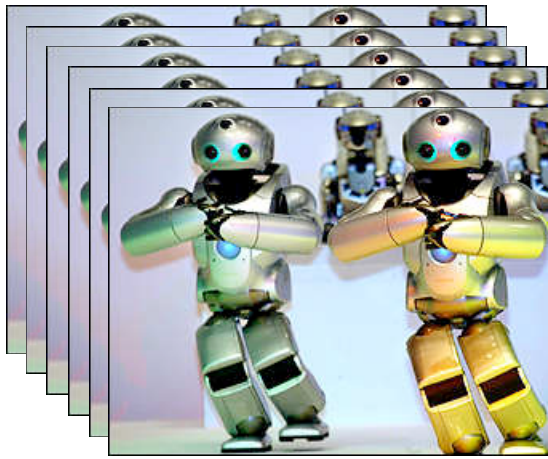


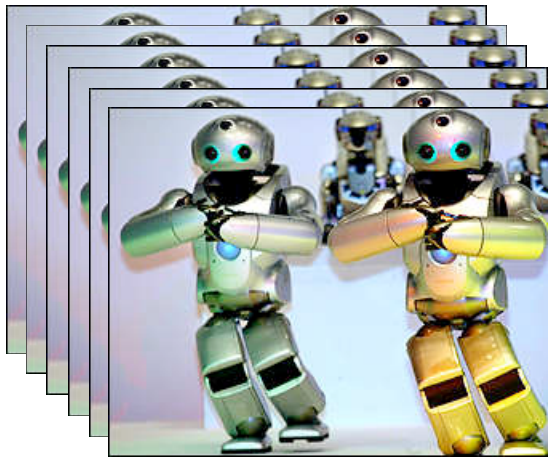
Culture – Biological Perspective

- Culture: Behaviour acquired from conspecifics by non-genetic means (Richerson & Boyd 2005).
- Neo-diffusionist hypothesis: cultural diffusion of adaptive behaviours more likely than neutral or negative traits (Kashima 2008).

Culture as Concurrency

- If each agent has a 1% chance of discovering a skill (e.g. making yogurt) in its lifetime and there are 2000 agents, at any instant probably some agents will know the skill.
- If it is easier to learn the skill from a knowledgeable agent than by discovery, then selective pressure for **culture**.
- Inclusive fitness $c < b \times r$ (Hamilton 1964; West *et al* 2007).





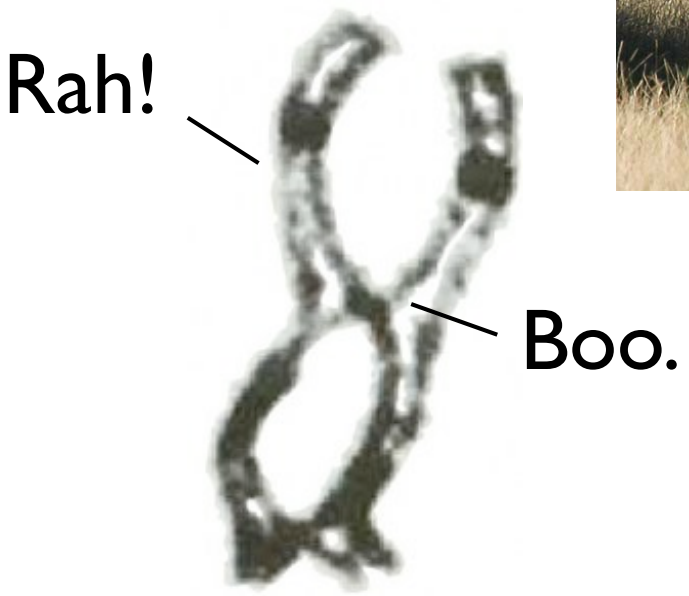
What About Selfish Genes?

- How can evolution select traits that help the community but hurt the individuals?
- Inclusive fitness & kin / group selection:
  - What is transmitted is the replicator.
  - The unit of selection is the vehicle (or interactor.)
  - Most current vehicles are composed of many, many replicators.

(Dawkins e.g. The Extended Phenotype)

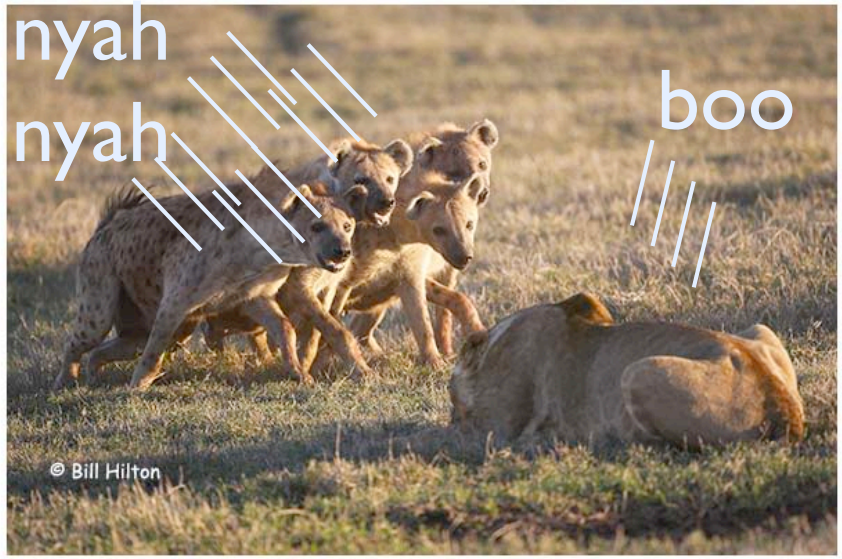
Multiple Levels of Interaction ⇒Cooperation

Replicator (Gene)



Organism

Group



Strategies for Speeding Search

- Concurrency
  - multiple searches at the same time,
  - only effective if solutions can be communicated.
- Pruning
  - limit search to likely space of solutions

Culture Lets Humans Search Faster

Language

Built Culture

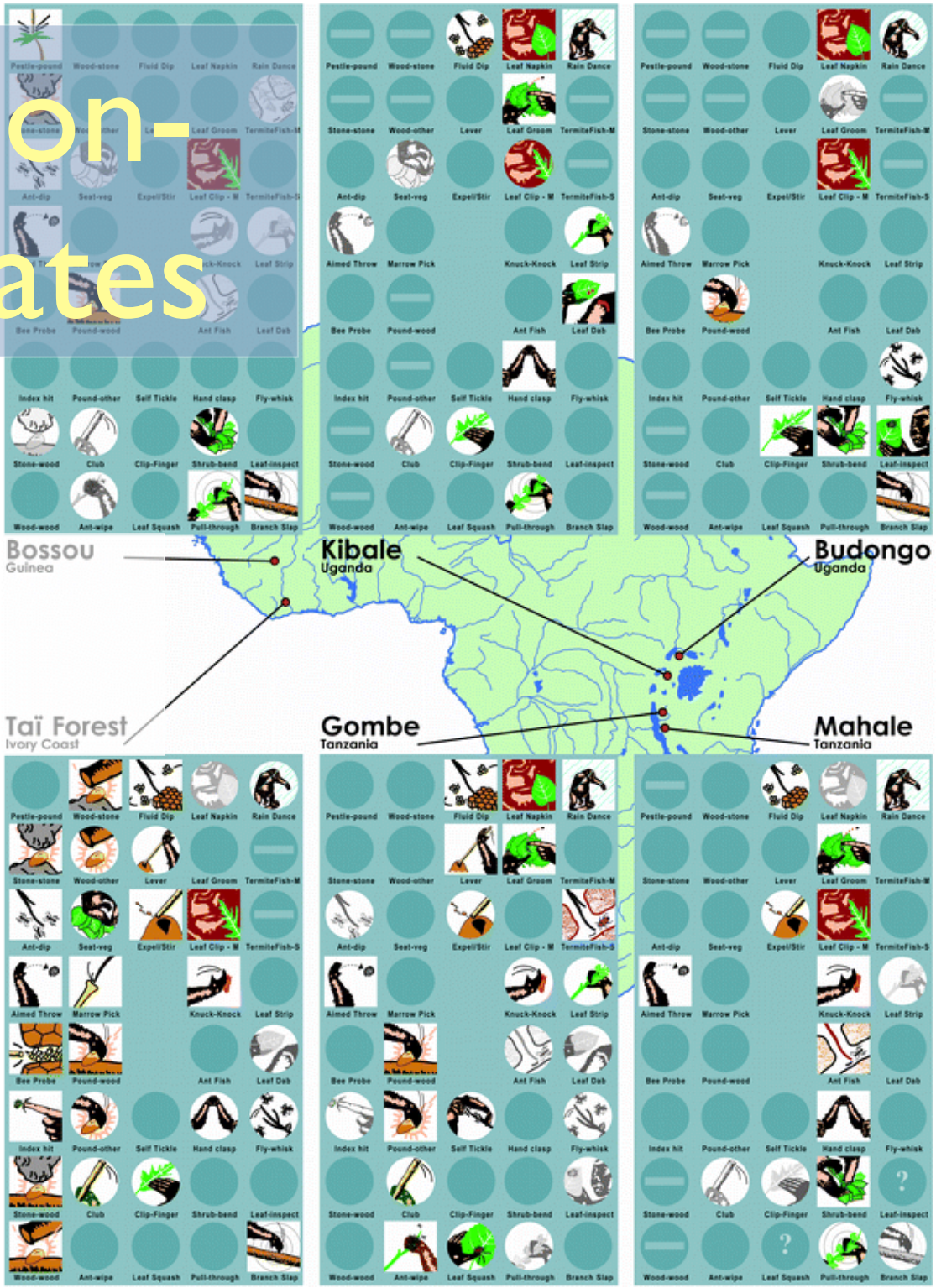
Why Don't Other Species Use It?

They Do

Culture in non-human primates

Chimpanzees (Whiten, Goodall, McGew, Nishida, Reynolds, Sugiyama, Tutin, Wrangham, & Boesch 1999, p . 684).

Macaques (de Waal & Johanowicz 1993); Capuchins (Perry *et al* 2003); Orangutans (van Schaik *et al* 2003).

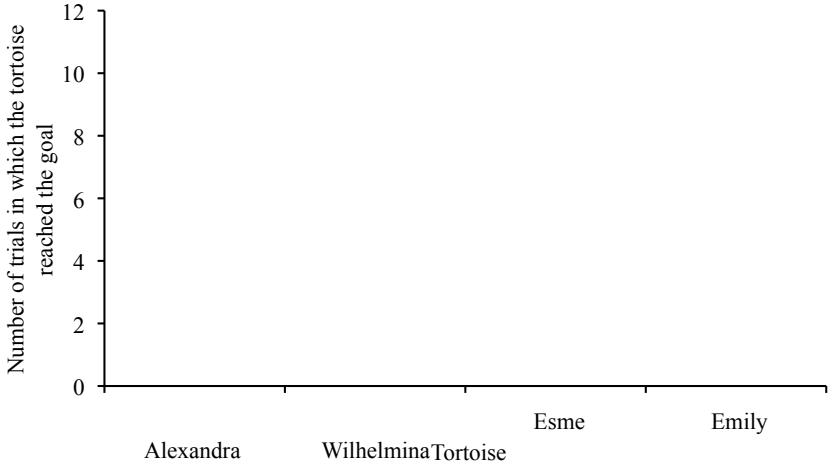


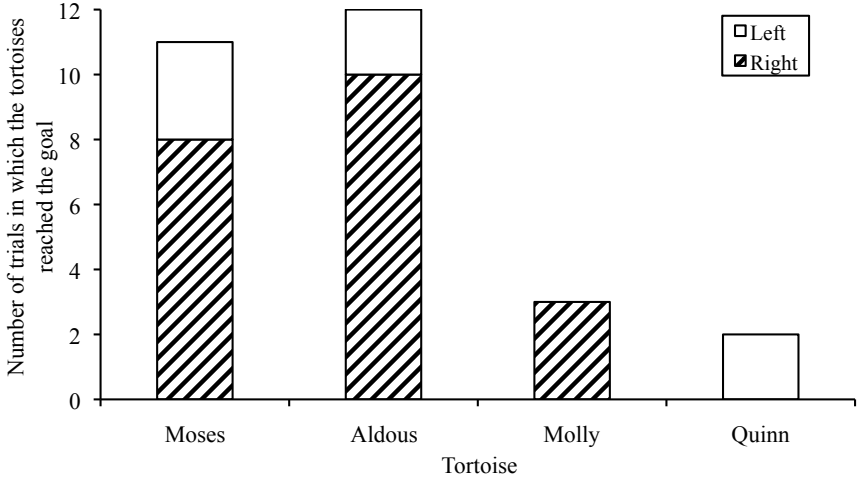
Culture in non-human primates

Chimpanzees (video from Whiten)

‘Solitary’ Tortoises Use Culture if It’s Available

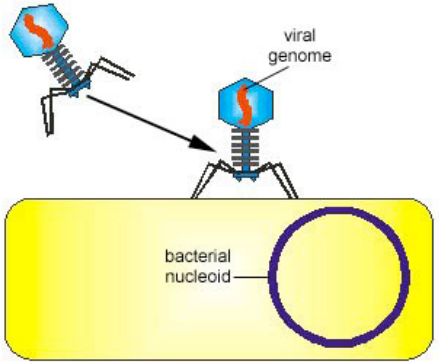
Social Learning in a Non-Social Tortoise
Anna Wilkinson, Karin Künstner
Julia Müller& Ludwig Huber 2010.

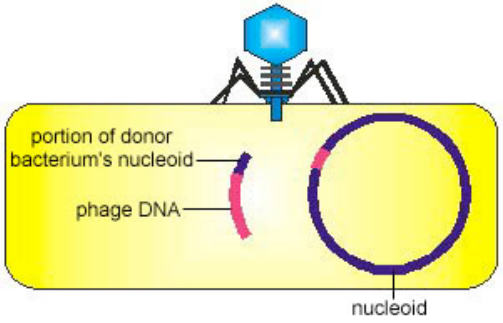


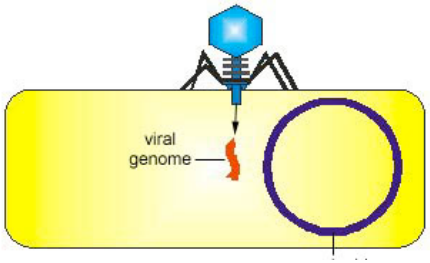


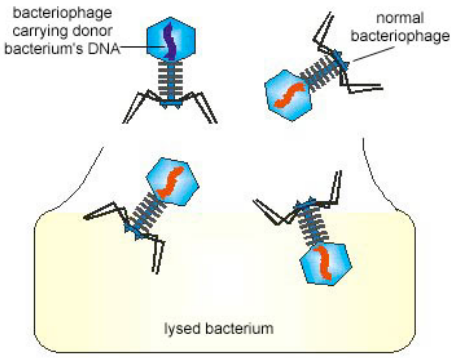
Even Bacteria Share Info

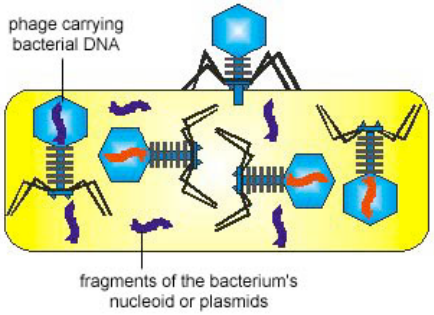
MGEs: e.g. Phages & Plasmids



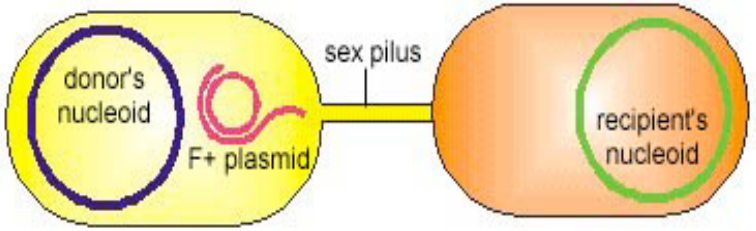


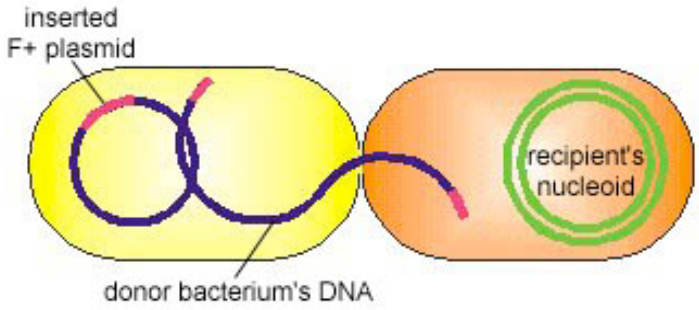






One on One 'speech'?



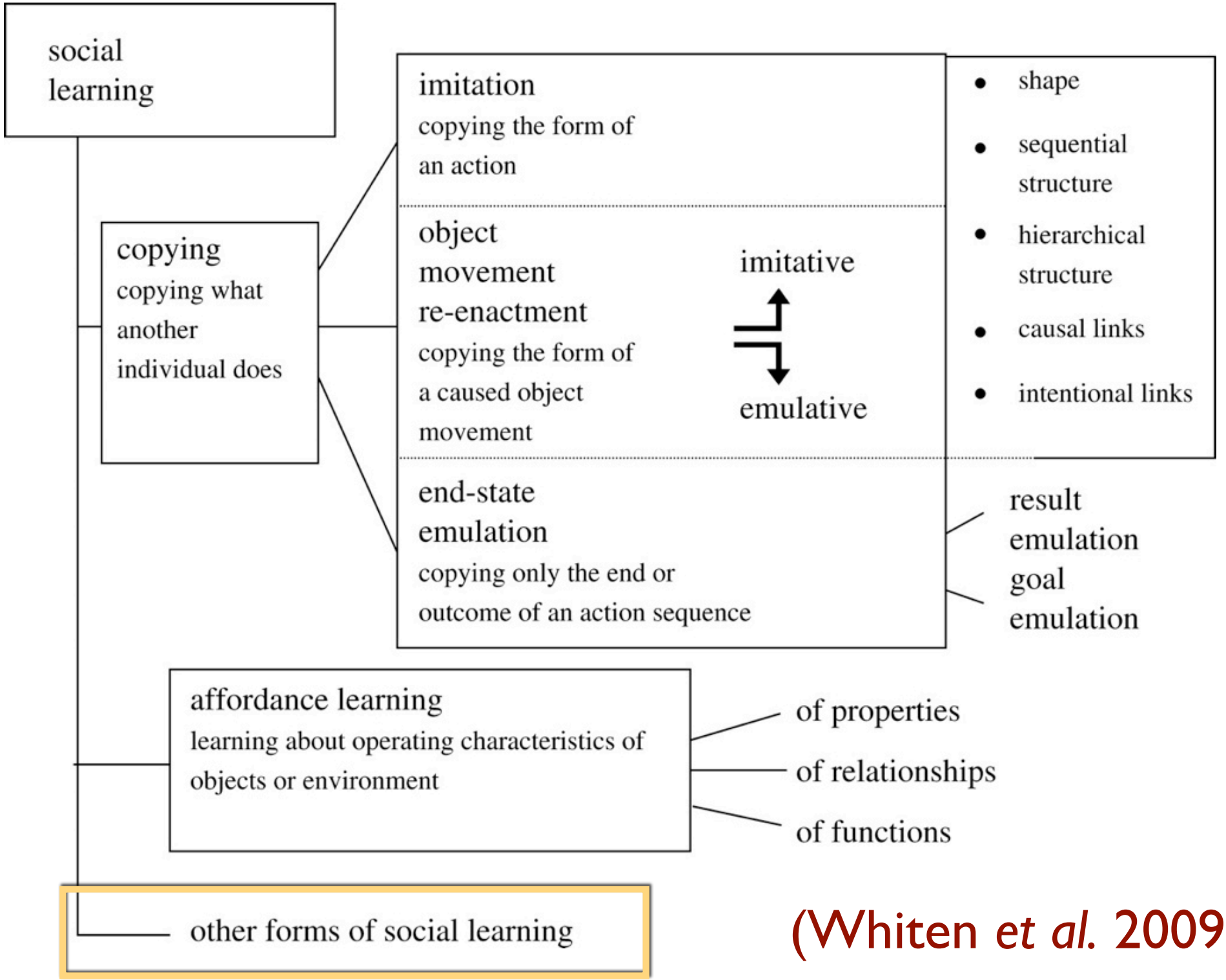


'Books'?

Images from Bharat Kumar Chimanlal Patel

How Culture is Transmitted

- Intentionally versus unintentionally
- By instruction or by demonstration



(Whiten *et al.* 2009)

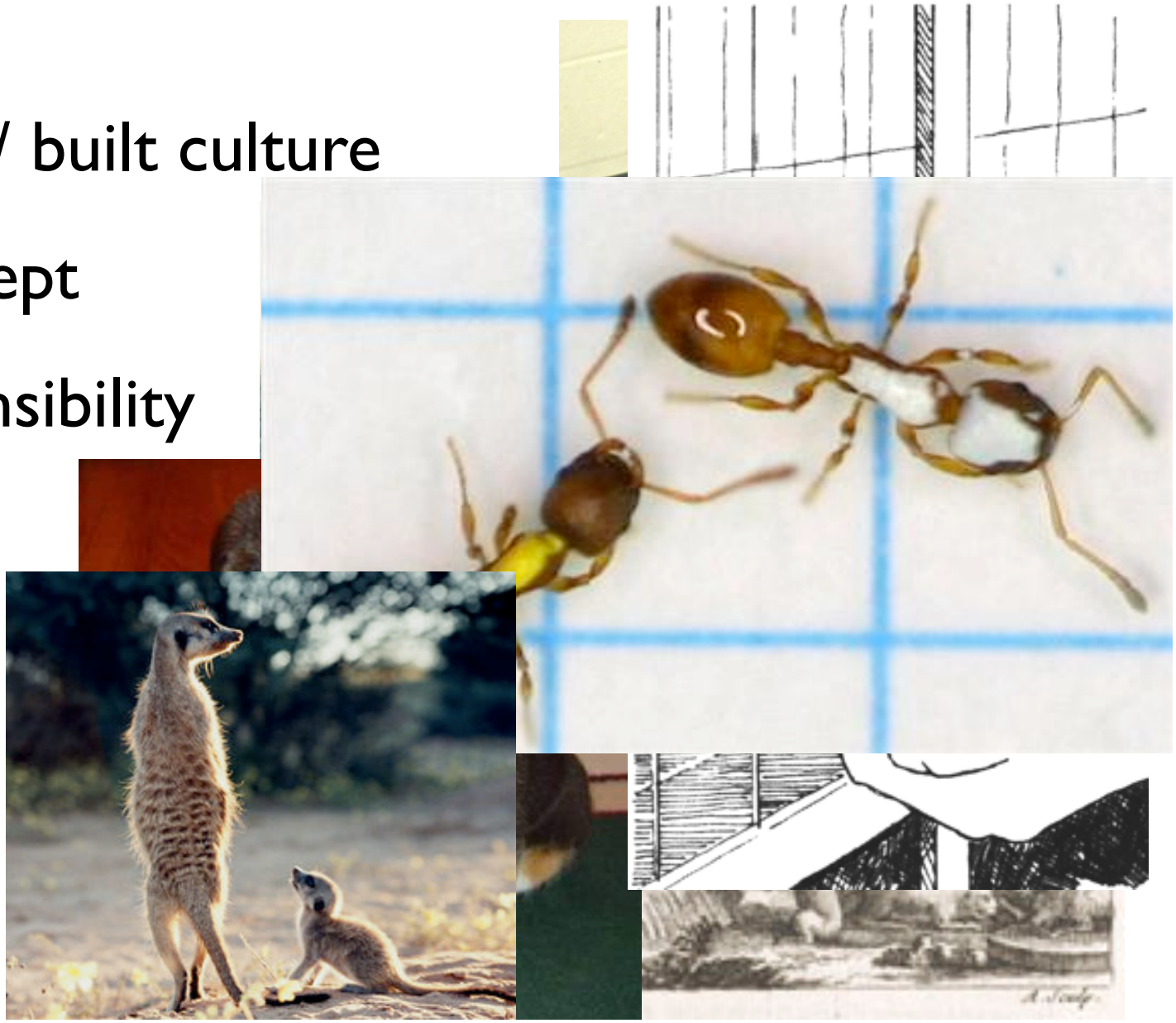
Ways to Transmit Culture

- Intentionally versus unintentionally
- By instruction or by demonstration
- Language and teaching

Uniquely human?

Human Uniqueness

- Tool use / built culture
- Self concept
- Moral sensibility
- Culture
- Teaching
- Language



Human Uniqueness

- Tool use / built culture
- Self concept
- Moral sensibility
- Culture
- Teaching
- Language

Outline

- What is culture for? (computationally)
  - Why are we social?
  - Why do we communicate?
- Language as a special case:
  - Phonetics/phonology/morphology, Syntax, Semantics, & Pragmatics.
- Natural Language Processing (NLP)

Language: Many Types of Information

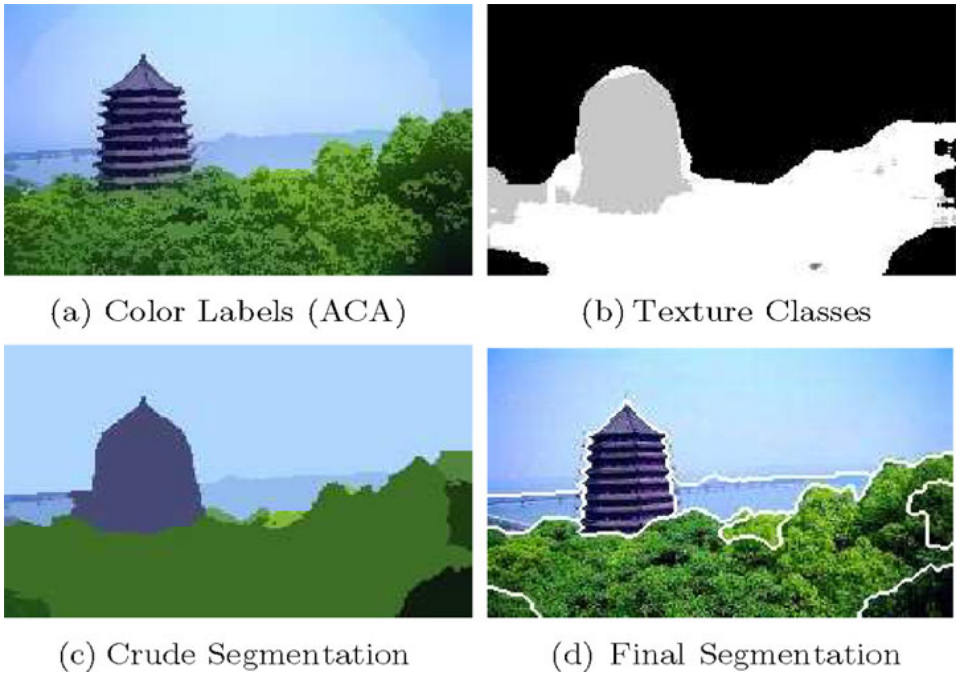
- Phonetics/phonology/morphology: what words (or subwords) are we dealing with?
- Syntax: What phrases are we dealing with? Which words modify one another?
- Semantics: What's the literal meaning?
- Pragmatics: What should you conclude from what was said? How should you act?

Phonetics / phonology / morphology

- Understanding a speech (or character) stream requires decomposing it into the units that have meaning: segmentation.
- Phonemes are relatively discrete (though they can be merged in transitions.)
  - Infants babble all(?) initially then settle on the ones they hear / in their language.

Segmentation

- Objects in a scene.
- Gestures in a video.
- Words in speech.
- Actions in sequence.

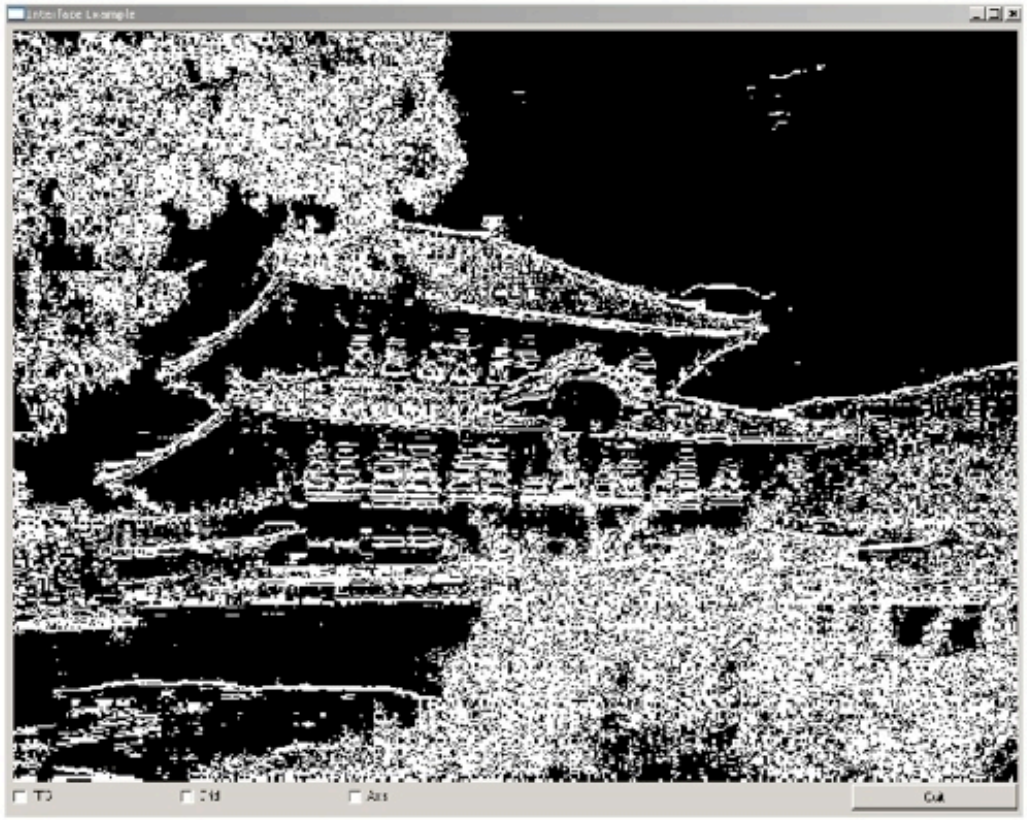


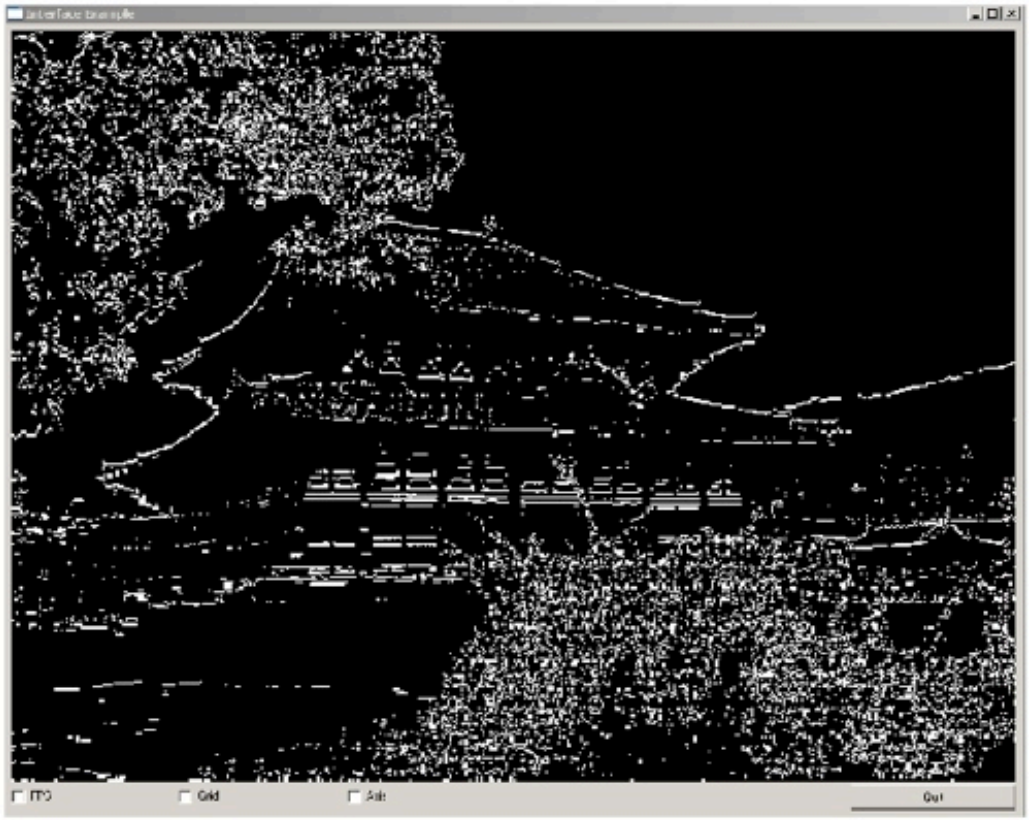
(a) Color Labels (ACA) (b) Texture Classes

(c) Crude Segmentation (d) Final Segmentation

Junqing Chen and Thrasyvoulos Pappas

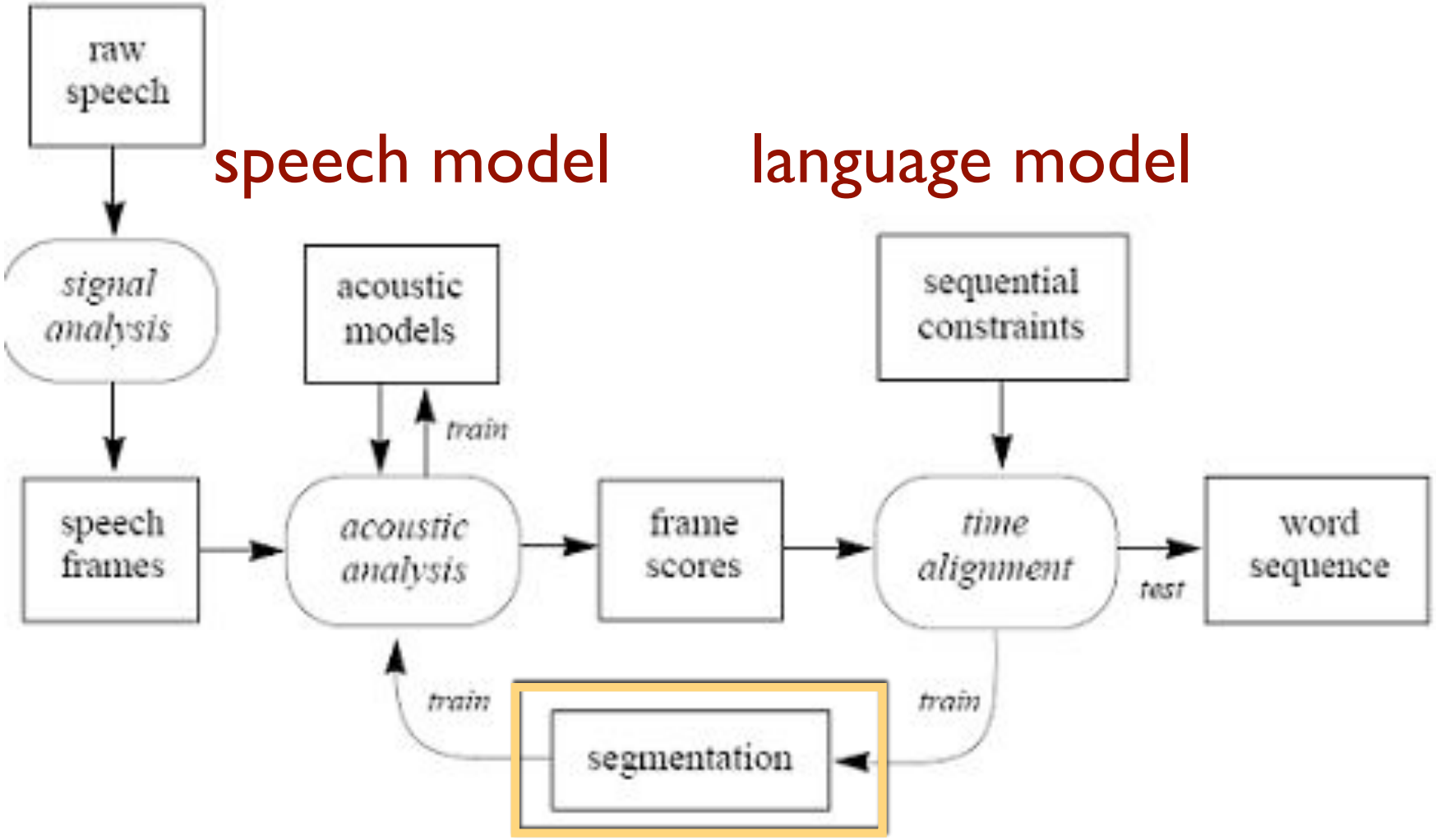
Very, very hard in all domains; better with multiple information sources.



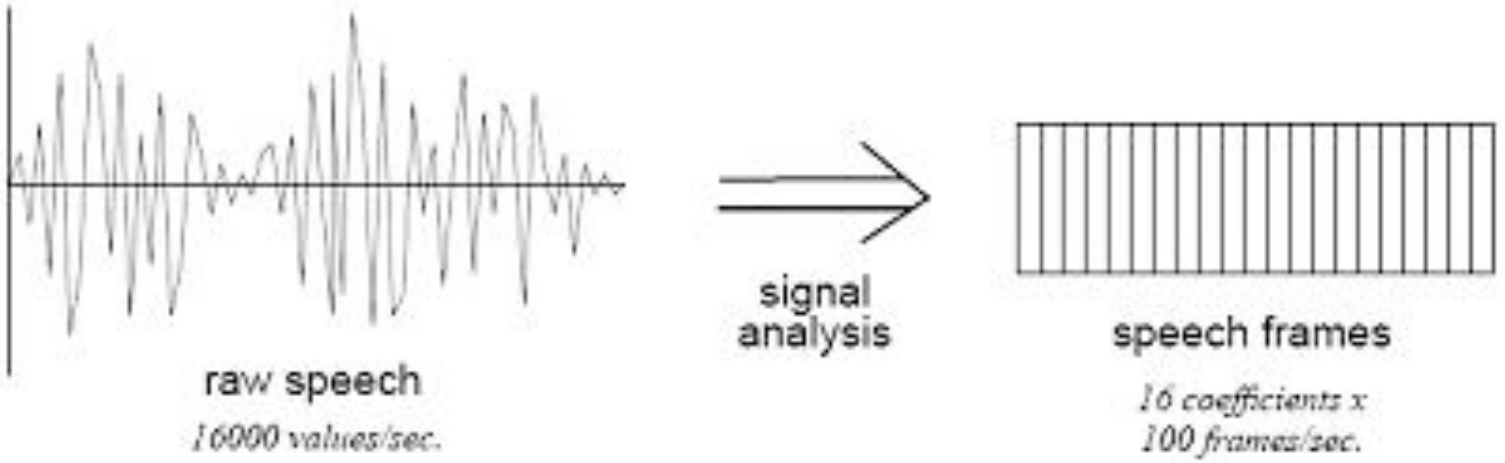


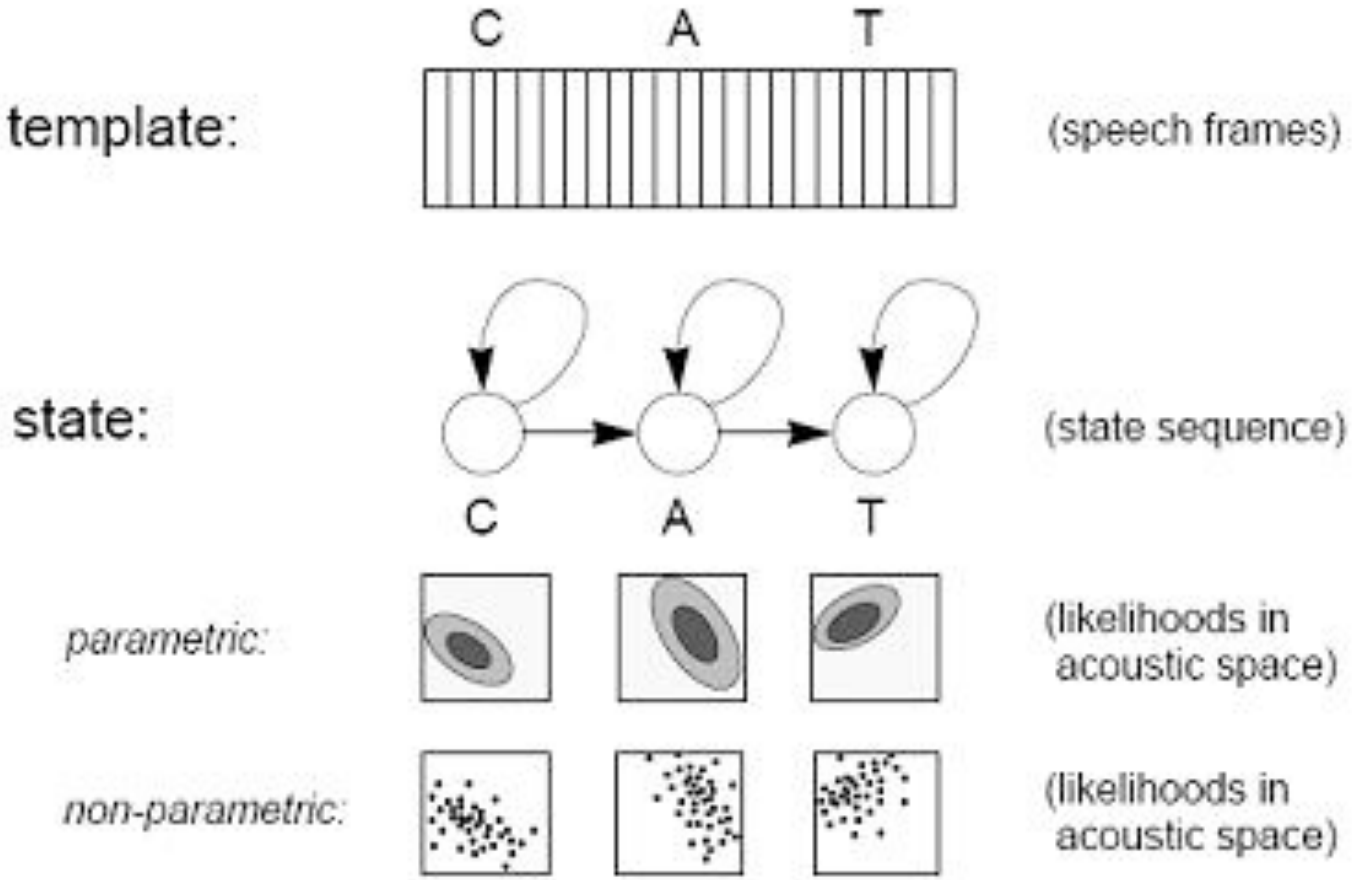
HARD

Speech Recognition

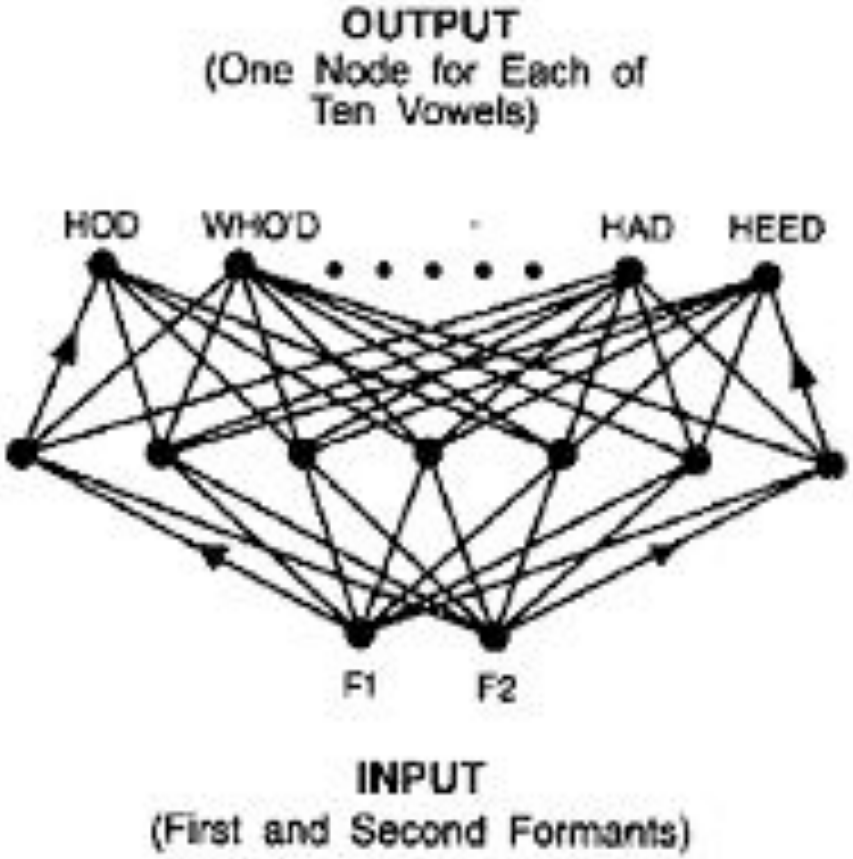


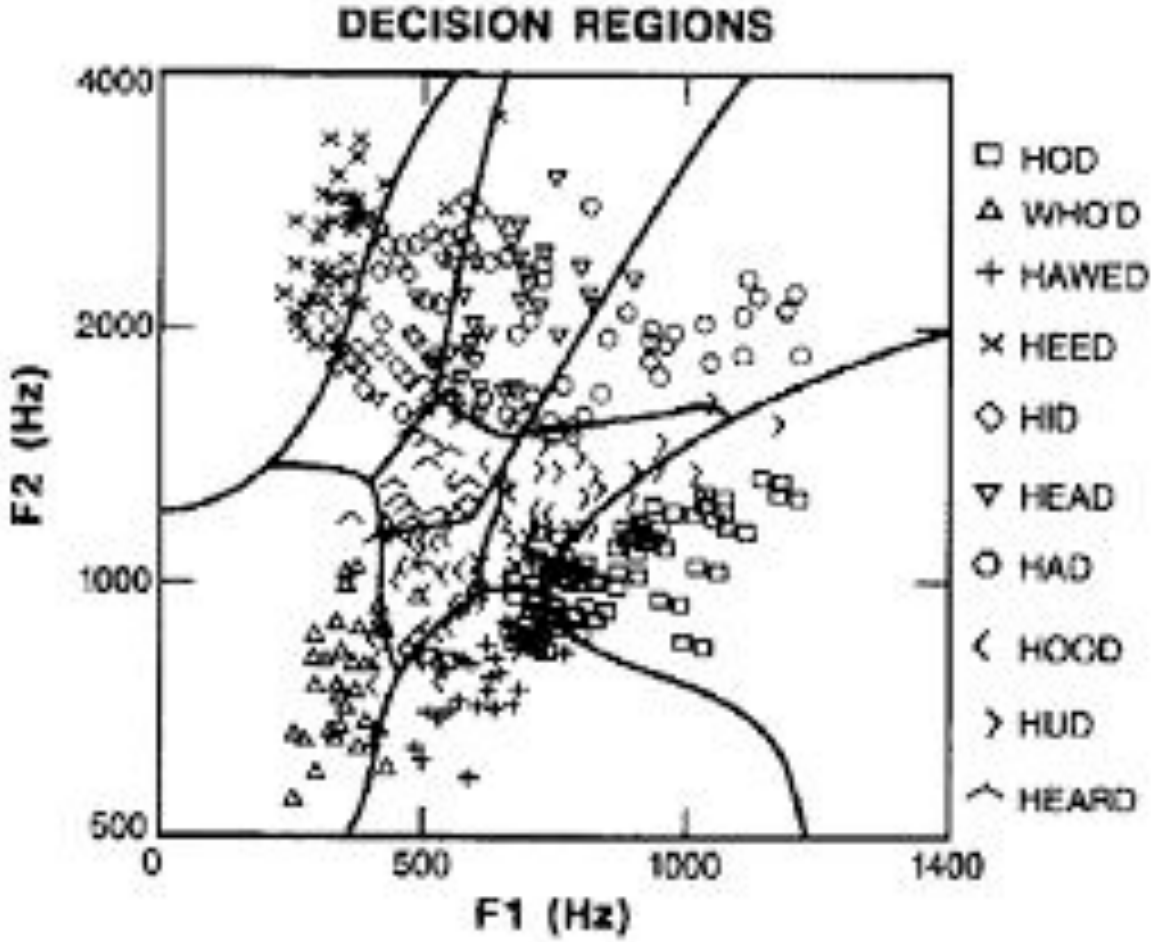
http://www.learnartificialneuralnetworks.com/speechrecognition.html


raw speech
16000 values/sec.
signal analysis
speech frames
16 coefficients x
100 frames/sec.


C A T
template:
(speech frames)
state:
C A T
(state sequence)
parametric:
(likelihoods in acoustic space)
non-parametric:
(likelihoods in acoustic space)

Lots of Machine Learning / Pattern Rec





Decision regions formed by a 2-layer perceptron using backpropagation training and vowel formant data. (From Huang & Lippmann, 1988.)

Language Outline

- Phonetics/phonology/morphology: what words (or subwords) are we dealing with?
- Syntax: What phrases are we dealing with? Which words modify one another?
- Semantics: What's the literal meaning?
- Pragmatics: What should you conclude from what was said? How should you act?

Syntax

A Brief History of AI

- Founded in the 1950s.
- Funded in the 1960s by promising machine translation (esp. Russian). --- Theory: Solve syntax as a program, lookup semantics in dictionary.

What AI Thought Language Was

- Phonetics/phonology/morphology: what words (or subwords) are we dealing with?
- Syntax: What phrases are we dealing with? Which words modify one another?
- Semantics: What's the literal meaning?
- Pragmatics: What should you conclude from what was said? How should you act?

The Plan For Translation

- Build something that parses and generates individual language syntax.
- Automatically morph sentences between languages' syntaxes.
- Use dictionaries to look up replacement words (semantics).

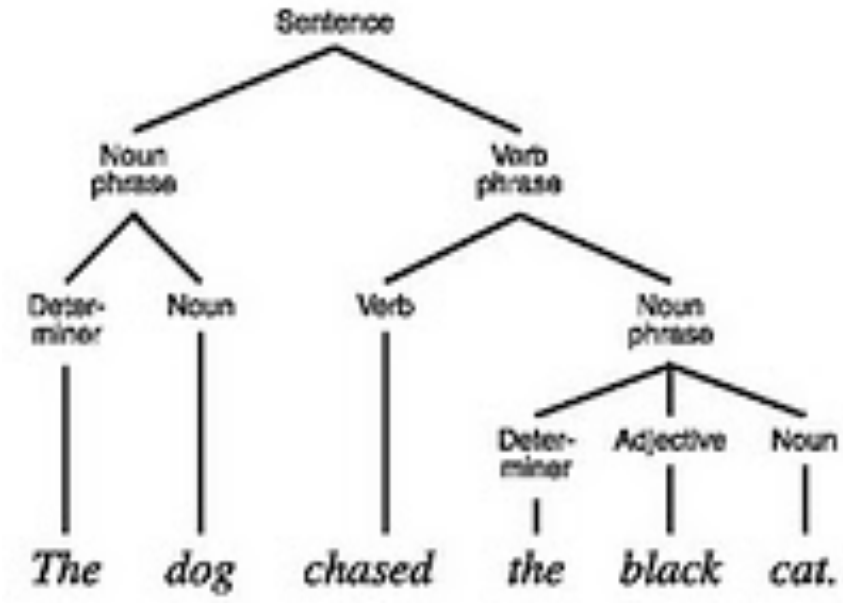
Warning: almost totally doesn't work

Syntax: Chomsky's Grammar(s)

- S → NP + VP
- NP → N | D + NP | ADJ + N | PN
- VP → IV | AUX + VP | TV + NP
- IV → laughed | cried | ...
- AUX → can | will | shall | ... |
- TV → throw | catch | ...
- N → dog | peacock | justice |.
- D → the | a | an
- PN → he | she | they | ...

English! e.g.
SVO vs SOV

Vocabulary:
terminal symbols
closed classes



What to Do With a Grammar: Parse

- Use it to parse a sentence.
  - Ambiguous sentences have multiple parse trees.
  - Ambiguity can came from multiple definitions (remember, plug in semantics last – often FOPL).
  - Other words or context may resolve.

The farmer pulls the cow on the barn.

What to Do With a Grammar: Generate

- Use it to generate a sentence.
  - Associate a probability with every option.
  - Throw dice.
  - Automatic language!

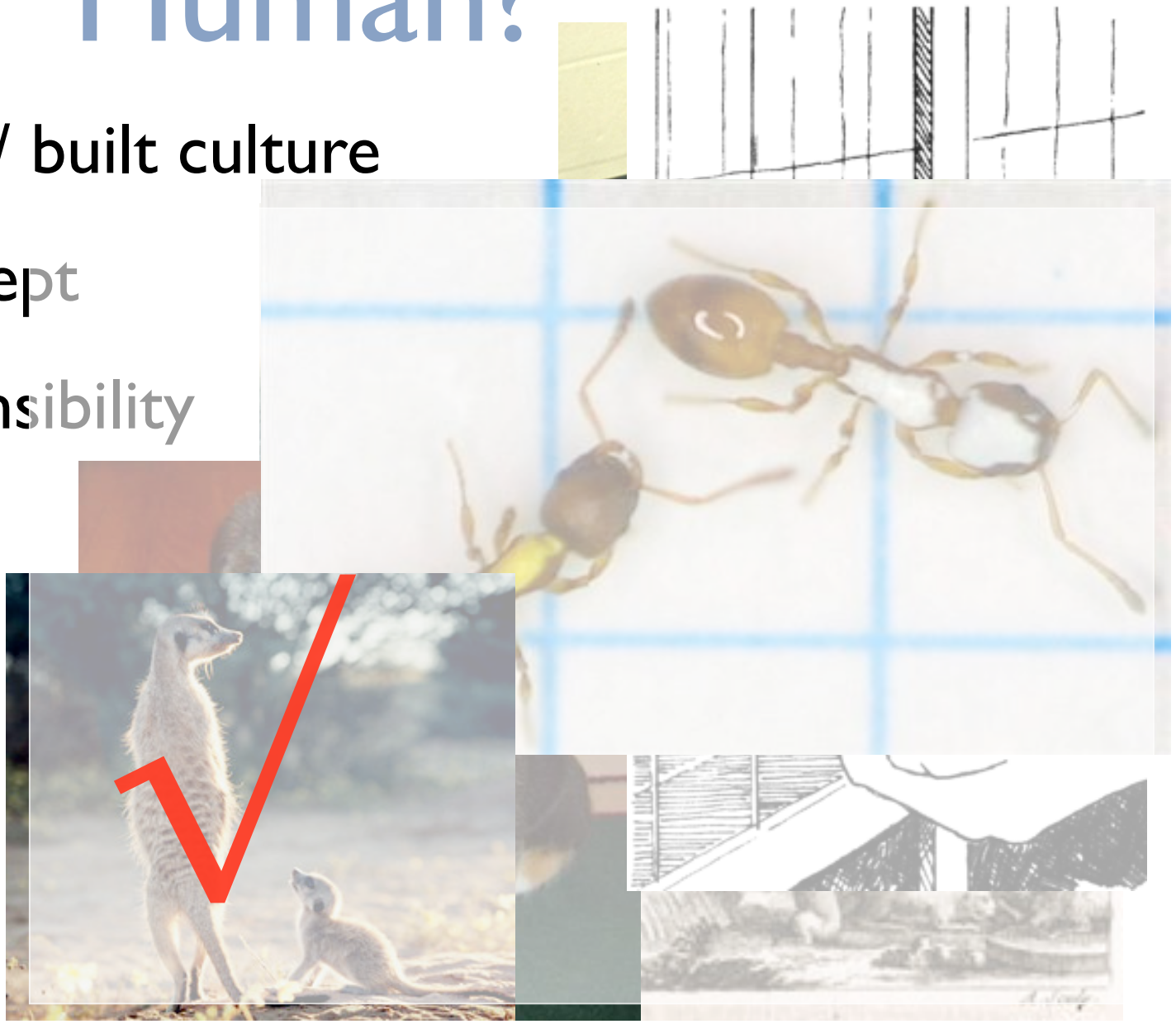
Example

- S → NP + VP
- NP → N | D + NP | ADJ + N | PN
- VP → IV | AUX + VP | TV + NP
- IV → laughed | cried | ...
- AUX → can | will | shall | ... |
- TV → threw | caught | ...
- N → dog | peacock | justice |...
- D → the | a | an

Dog
will
catch
an
peacock
.

Is Language Uniquely Human?

- Tool use / built culture
- Self concept
- Moral sensibility
- Culture
- Teaching
- Language



Compositionality / Recursion

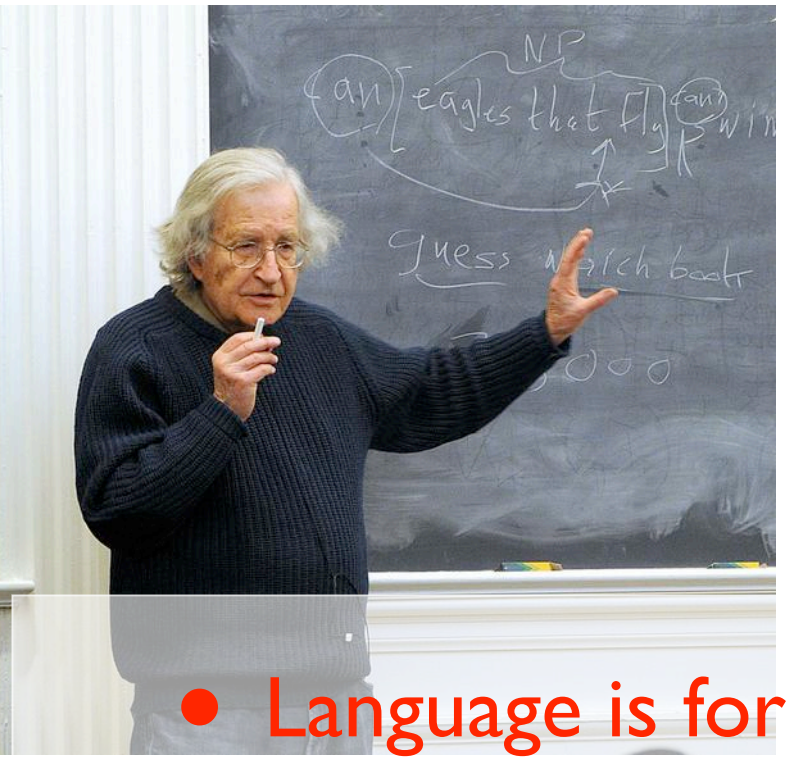
- S → NP + VP
- NP → N | D + NP | ADJ + N | PN
- VP → IV | AUX + VP | TV + NP
- IV → laughed | cried | ...
- AUX → can | will | shall | ... |
- TV → threw | caught | ...
- N → dog | peacock | justice | ...
- D → the | a | an

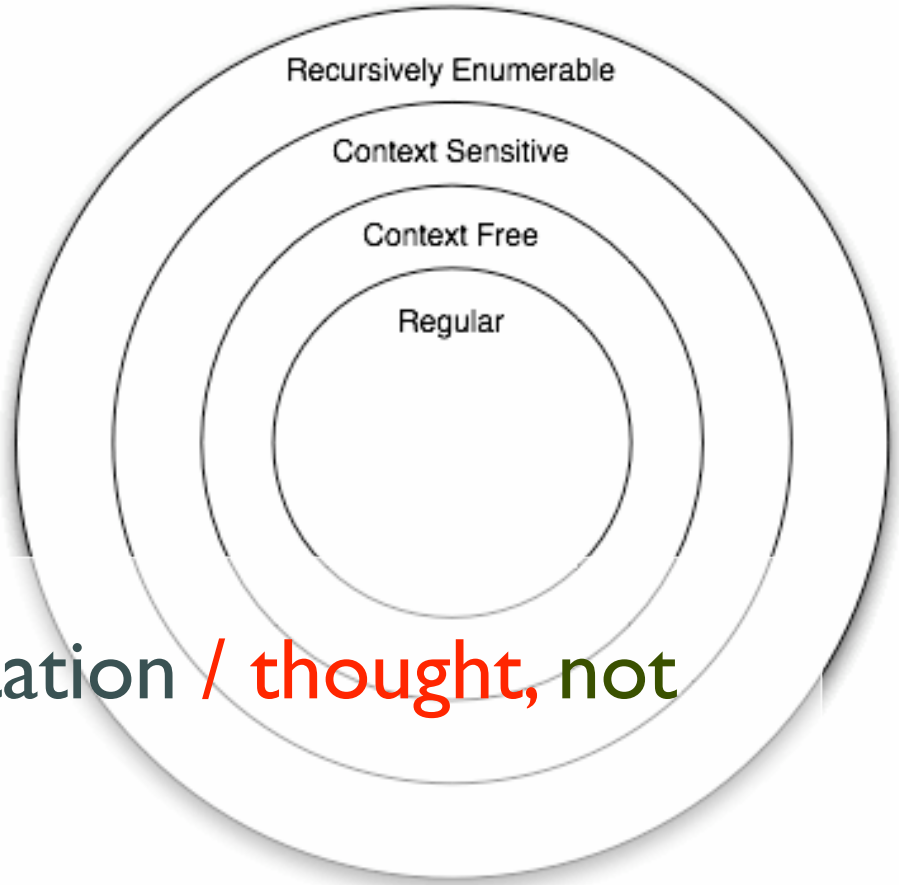
Allows language to be infinitely productive.

What no animal language learner has shown.

(cf. Hauser, Chomsky, & Fitch 2002; Berwick & Chomsky 2015 maybe…)

Chomsky on Cognition





- Language is for computation / thought, not communication.
- Grammars can tell you the limits of human intelligence (e.g. CFG?)

Chomsky's Universal Grammar

- Hypothesis: every human is born with the universal grammar capacity.
- Learns to set parameters from listening (know this is true of phonemes).
- Evidence: Poverty of the stimulus – children don't hear enough negative examples to learn language from scratch.

Critiques of Universal Grammar

- You can learn a stochastic grammar model without many negative examples (Chomsky assumed a deterministic one, Chater & Manning, 2006).
- Many characteristics of the UG evolve in the language naturally in simulation – necessary characteristics of something learnable (Kirby 1999).

Dual replicator theory: Culture & biology both evolve at the same time under each other's influence.

Why are humans special? (Bryson 2008; 2009)

- Humans are the only primate species capable of precise vocal imitation (Fitch 2000; 2007).
- Communicates lots of information, including volume, pitch, timbre and time.
- Allows redundant encoding to preserve important details while others can mutate.
- Allows communication of complex, sustainable behaviour.

Why should temporal imitation matter?

- More information contained in the 'genetic' substrate.
- Allows for more variation while providing redundancy, robustness – assists GAs (Baluja 1992; Weicker & Weicker; 2001; Miglino & Walker 2002).
- Aligns with Wray (2000) on the evolution of language from phrases, Kirby (2000) on cultural selection for language efficacy.

Why Humans are Special (Bryson 2008, 2009)

	temporal imitation	no temporal imitation
second-order representations	people	non-human primates
no second-order representations	birds, seals	most things

Why Humans are Special (Bryson EoL 2010)

	temporal imitation	no temporal imitation
big brains, memories	people	non-human apes
no big brains, memories	birds, seals	most things

Language Outline

- Phonetics/phonology/morphology: what words (or subwords) are we dealing with?
- Syntax: What phrases are we dealing with? Which words modify one another?
- Semantics: What's the literal meaning?
- Pragmatics: What should you conclude from what was said? How should you act?

Pragmatics

Just one slide…

Pragmatics

- What you really mean– requires context.
- Much elaborate work on reference. e.g. "They thought I was going to town but that wasn't what I meant."
- Still doesn't get you to "uh"→ /no don't go in there keep going straight/ (Agre & Chapman 1988).

Which leads into…

Semantics and Grounding

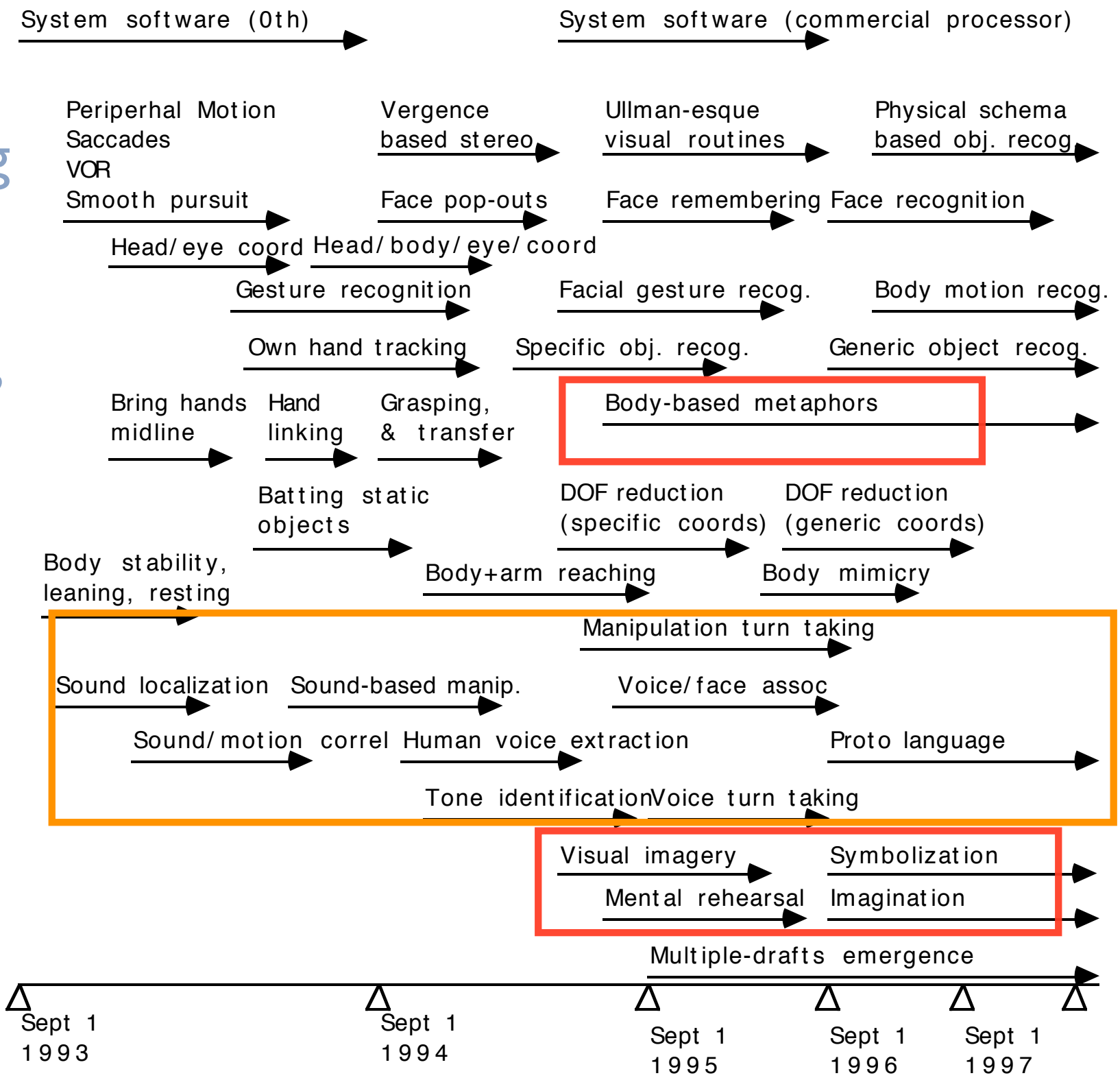
A Brief History of AI

- Founded in the 1950s.
- Funded in the 1960s by promising machine translation (esp. Russian). --- Theory: Solve syntax as a program, lookup semantics in dictionary.
- By 1980s, funders restless. --- Theory: Semantics requires grounding in an embodied system (Harnad 1990, Brooks 1991).
- 1990s(–*now?*): Robots for Language.

Embodiment

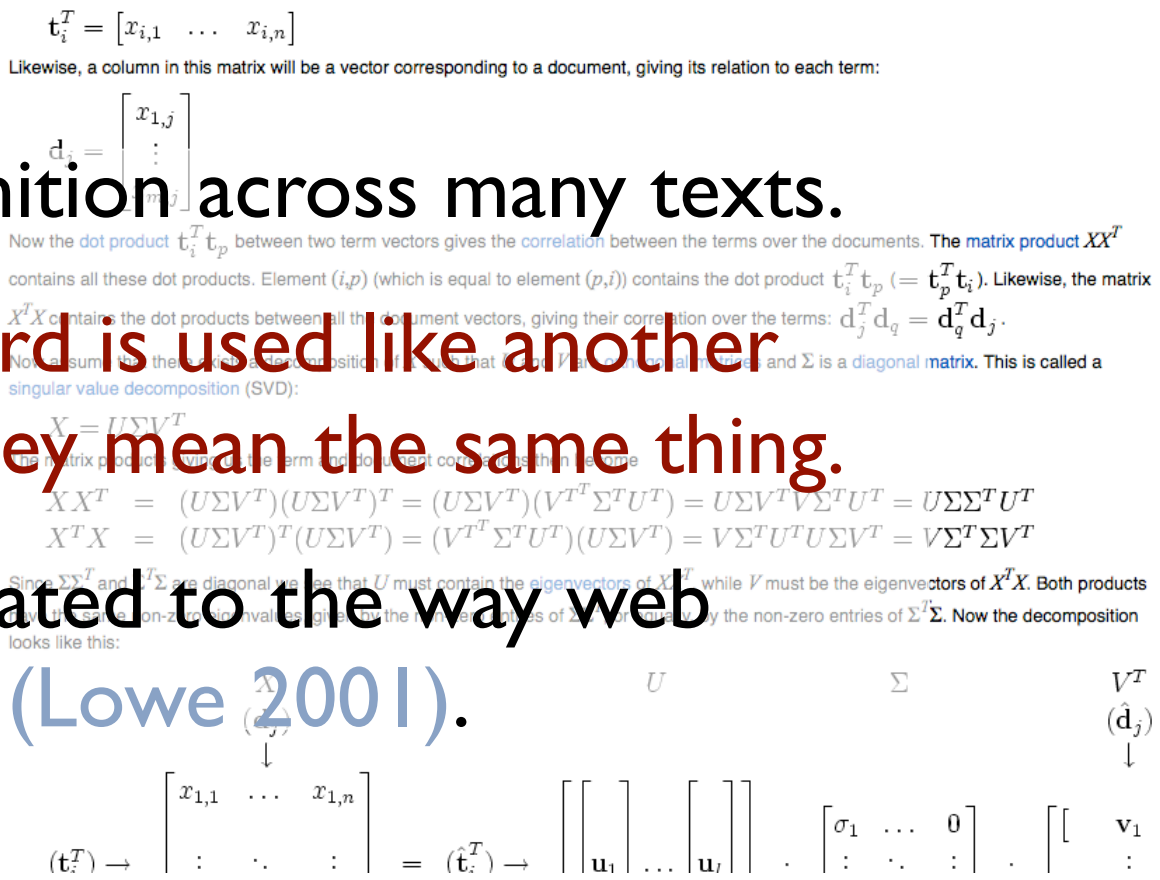
- Hypothesis: NLP has failed so far because semantics isn't grounded in human-like experience.
- E.g. life & career are understood via a metaphor to path which you learn about the hard way in your first few years. (Lakoff & Johnson 1999)
- Funding argument for humanoid robotics.
- Not much positive evidence.

"Building Brains for Bodies", Brooks & Stein (1993), MIT AI lab tech report 1439.



Alternative: Large Corpus Linguistics

- Do pattern recognition across many texts.
- The more one word is used like another word, the more they mean the same thing.
- Mathematically related to the way web pages are indexed (Lowe 2001).



$$\mathbf{t}_i^T = \begin{bmatrix} x_{i,1} & \dots & x_{i,n} \end{bmatrix}$$

Likewise, a column in this matrix will be a vector corresponding to a document, giving its relation to each term:

$$\mathbf{d}_j = \begin{bmatrix} x_{1,j} \\ \vdots \\ x_{m,j} \end{bmatrix}$$

Now the dot product $\mathbf{t}_i^T \mathbf{t}_p$ between two term vectors gives the correlation between the terms over the documents. The matrix product XX^T contains all these dot products. Element (i,p) (which is equal to element (p,i)) contains the dot product $\mathbf{t}_i^T \mathbf{t}_p$ $(= \mathbf{t}_p^T \mathbf{t}_i)$. Likewise, the matrix $X^T X$ contains the dot products between all the document vectors, giving their correlation over the terms: $\mathbf{d}_j^T \mathbf{d}_q = \mathbf{d}_q^T \mathbf{d}_j$.

Now assume that the singular value decomposition of X is $U\Sigma V^T$, where U and V are orthogonal matrices and Σ is a diagonal matrix. This is called a singular value decomposition (SVD):

$$X = U\Sigma V^T$$

The matrix products giving us the term and document correlations then become

$$\begin{aligned} XX^T &= (U\Sigma V^T)(U\Sigma V^T)^T = (U\Sigma V^T)(V^{T^T}\Sigma^T U^T) = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T \\ X^T X &= (U\Sigma V^T)^T(U\Sigma V^T) = (V^{T^T}\Sigma^T U^T)(U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T\Sigma V^T \end{aligned}$$

Since $\Sigma\Sigma^T$ and $\Sigma^T\Sigma$ are diagonal we see that U must contain the eigenvectors of XX^T, while V must be the eigenvectors of $X^T X$. Both products have the same non-zero eigenvalues, given by the non-zero entries of $\Sigma\Sigma^T$, or equivalently by the non-zero entries of $\Sigma^T\Sigma$. Now the decomposition looks like this:

$$\begin{array}{ccc} & X & \\ & (\mathbf{d}_j) & \\ & \downarrow & \\ (\mathbf{t}_i^T) \rightarrow & \begin{bmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \dots & x_{m,n} \end{bmatrix} & = \end{array} \begin{array}{ccc} & U & \\ & & \\ & & \\ (\hat{\mathbf{t}}_i^T) \rightarrow & \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_l \end{bmatrix} & \end{array} \begin{array}{ccc} & \Sigma & \\ & & \\ & & \\ & \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_l \end{bmatrix} & \end{array} \begin{array}{ccc} & V^T & \\ & (\hat{\mathbf{d}}_j) & \\ & \downarrow & \\ & \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_l \end{bmatrix} & \end{array}$$

Wikipedia: Latent Semantic Analysis

Large Corpus Linguistics

- Human semantics can be replicated by statistical learning on large corpra (Finch 1993, Landauer & Dumais 1997, McDonald & Lowe 1998).
- Only information gathered on each word's 'meaning' is what words occur in a small window before and after it.
- Normally just choose 75 fairly frequent words to watch out for.

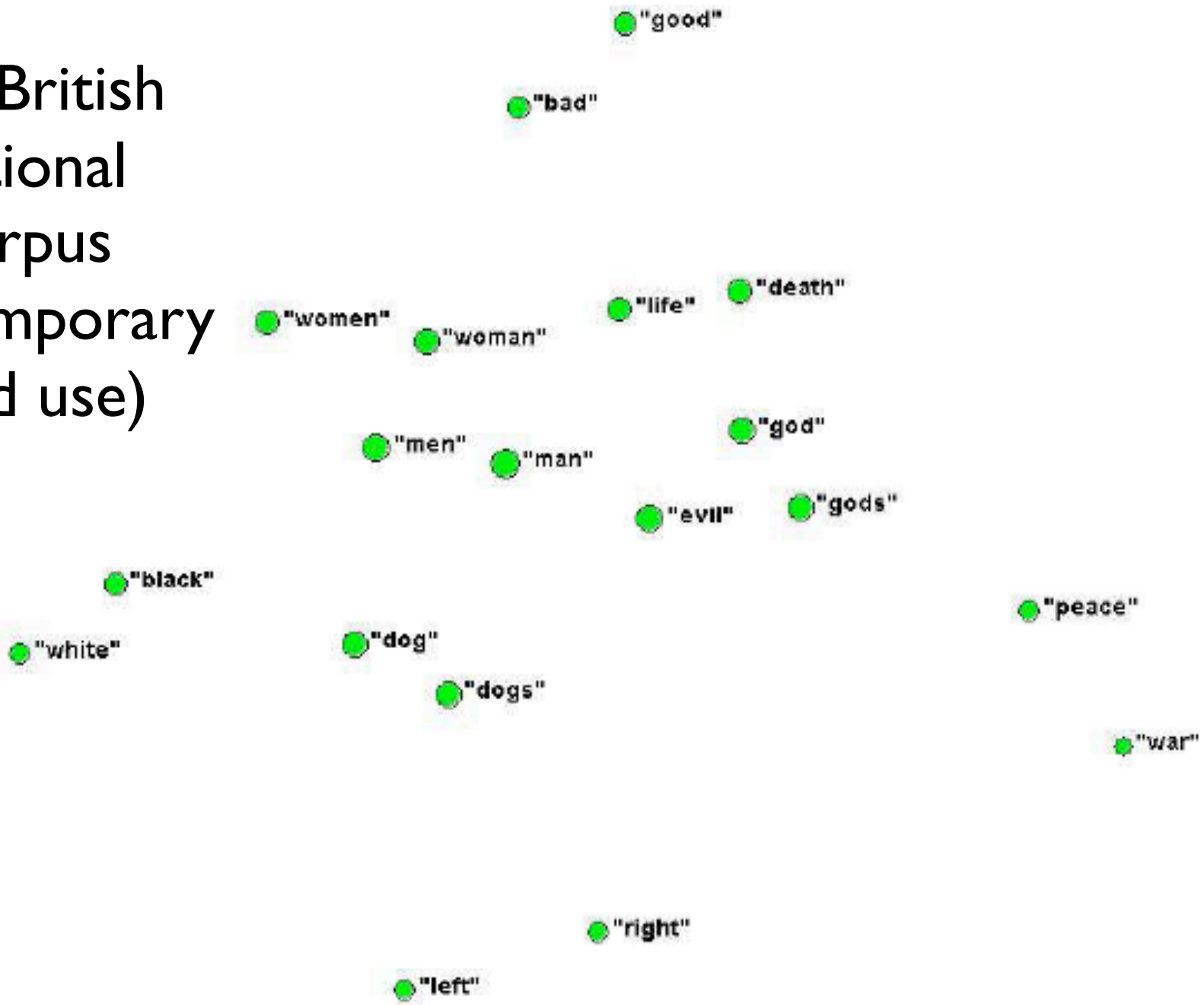
Data to Be Matched

- Semantic Priming – reaction times showing how similar people consider words' meanings to be.
- How quickly you are able to tell that a collection of letters is a real word is dependent on how similar the word's meaning is to words / concepts you have recently been exposed to.

Semantic Priming Replication, visualised with a 2-D projection (Lowe 1998). Analysis for comparison to human data uses similarity measured using 75-D cosines.

Bilovich 2006

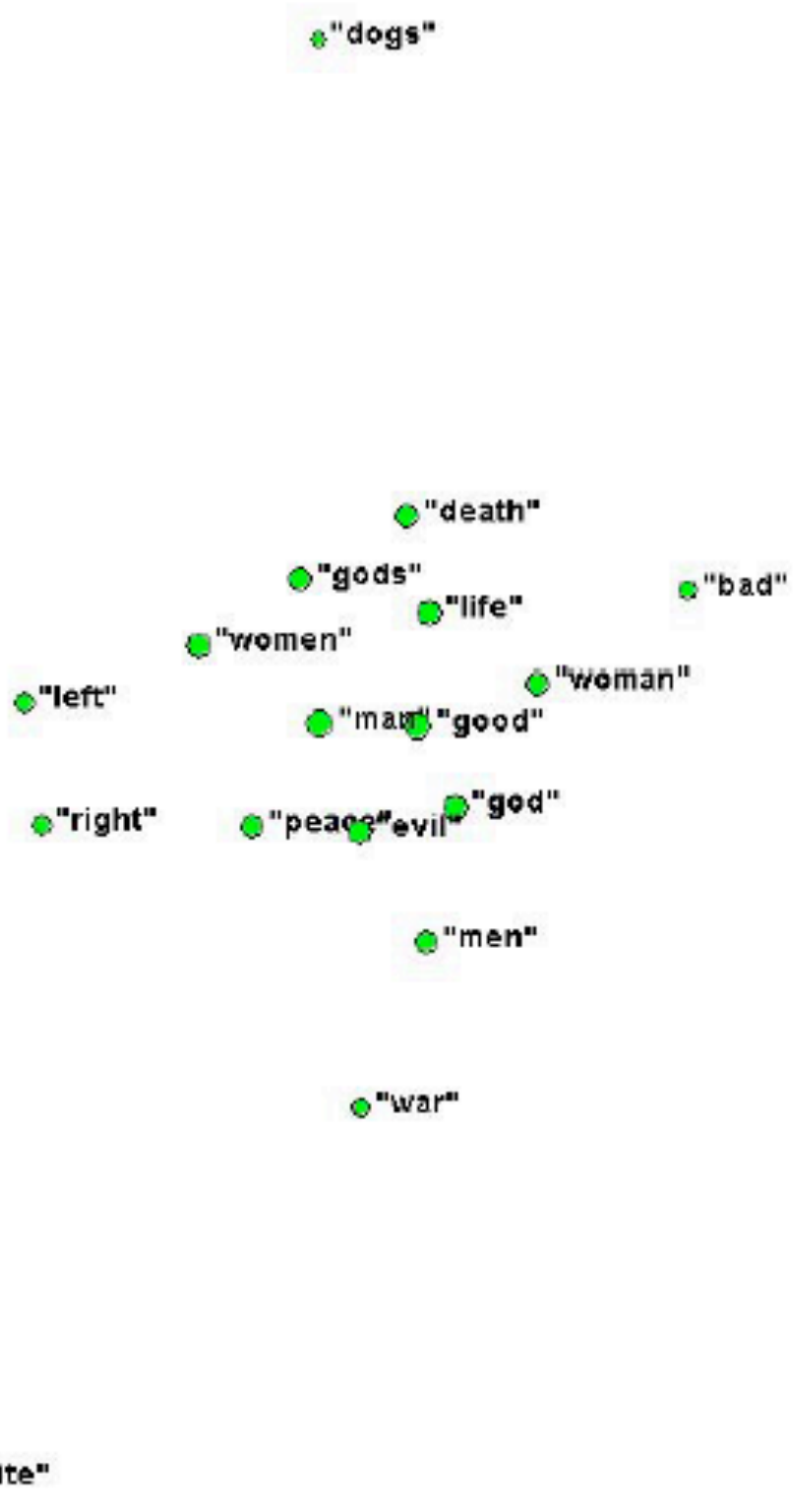
text: British National Corpus (contemporary word use)



Evolution of moral agency terms (Bilovich & Bryson 2008)

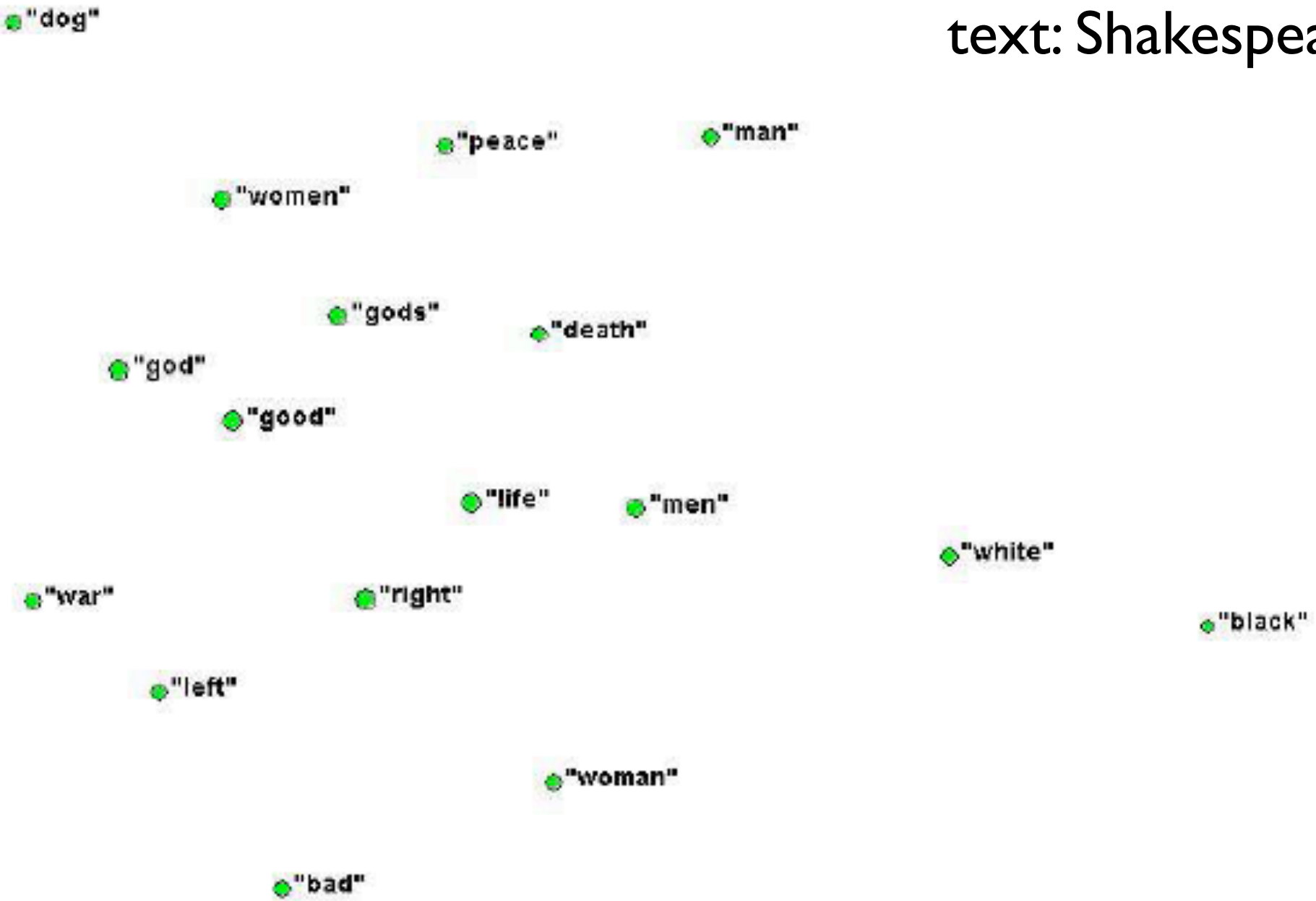
terms from the implicit bias task (Banaji & Greenwald 1994)

text: Bible



Bilovich 2006

text: Shakespeare



Humanlike Biases in Corpus Semantics

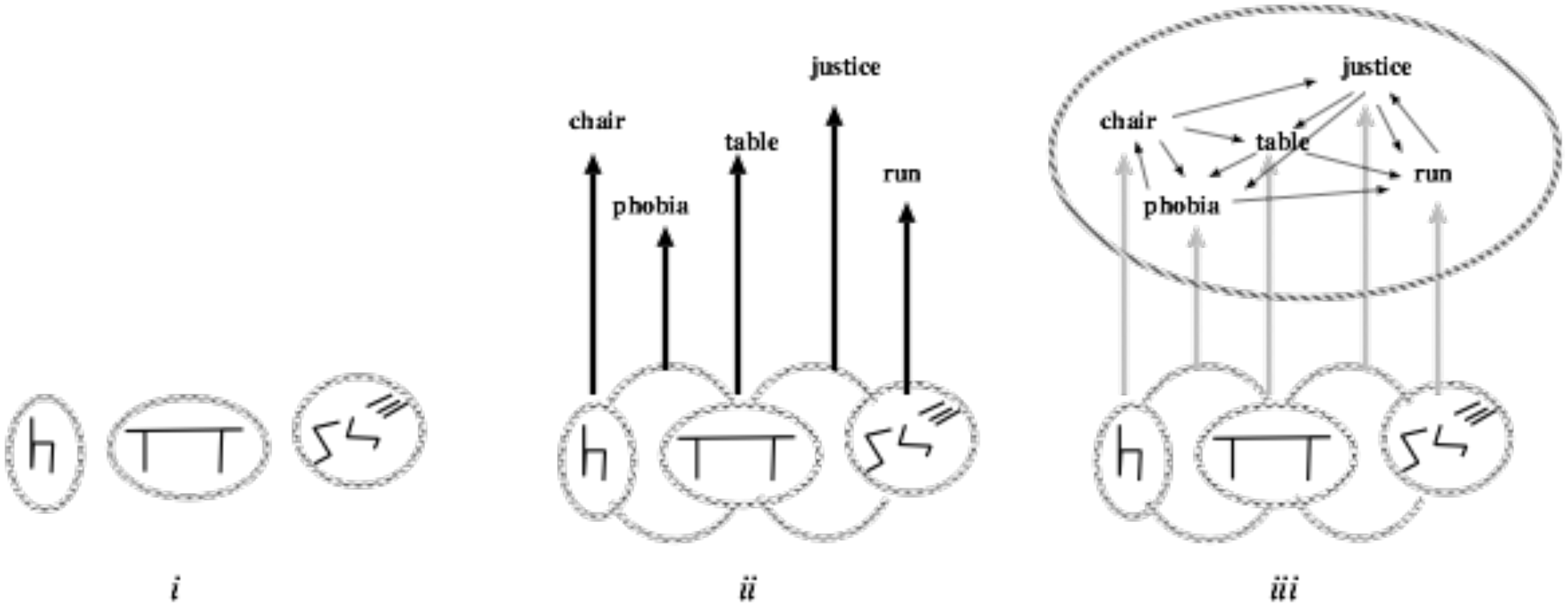
- Bilovich & I did not replicate Banaji (2003).
  - Nearest miss was Shakespeare – (nearly) single author?
- Macfarlane & I (2013) found matches.
- Caliskan, Bryson & Narayanan (2017) matched every general-population text-based implicit bias.

Macfarlane (2013) Results

- Life terms more like pleasant & Death terms more like unpleasant words.
- Elderly & Youth did not go as per Banaji on pleasantness, though did on competence.
- Male terms more like Career & Female terms more like Family.

In preparation; also University of Bath Computer Science technical report.

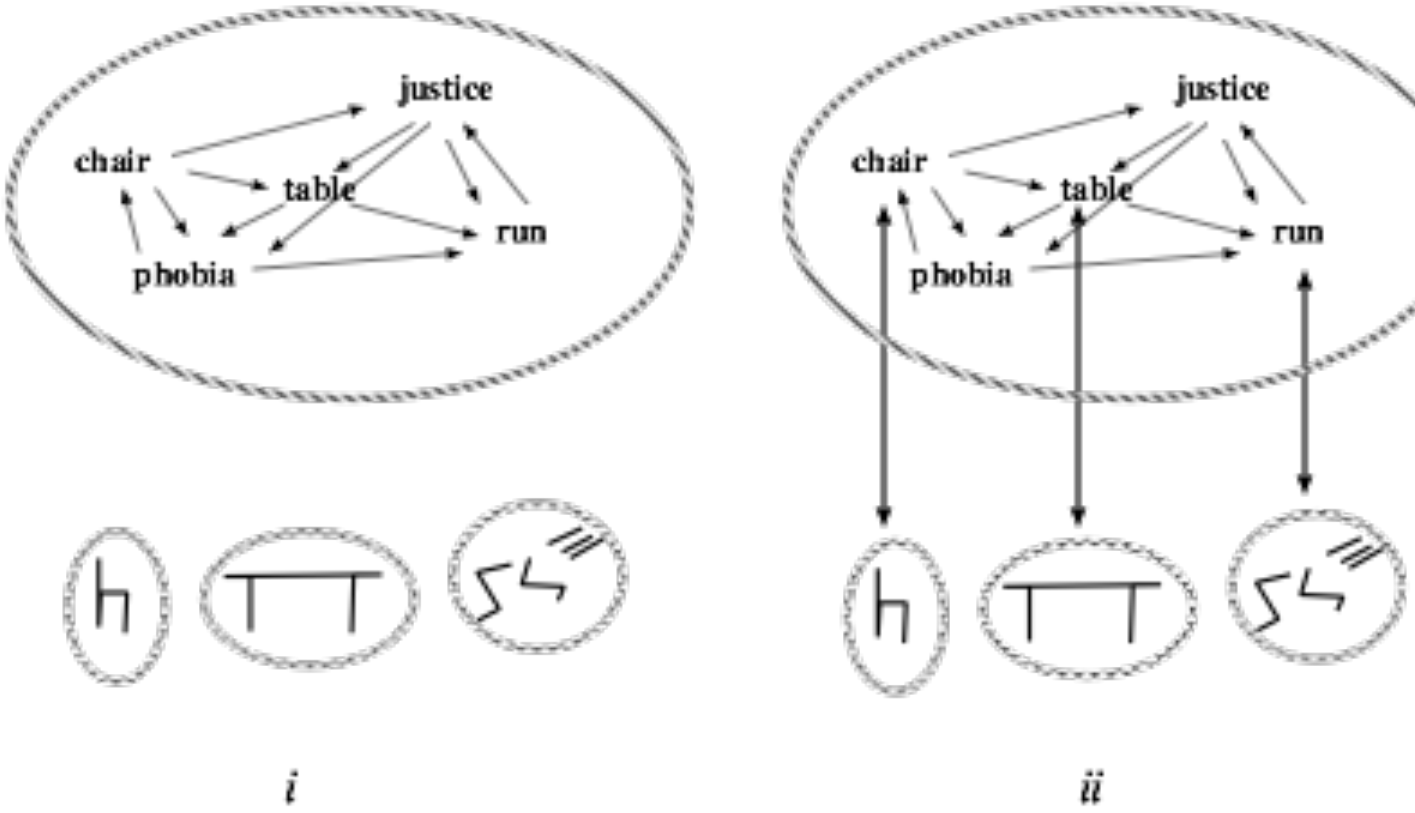
Traditional Theory of Semantics



Ontology

e.g. Deacon (1997) *The Symbolic Species*

Corpus Semantics Allows...



Ontology

more plausible!

Outline

- What is culture for? (computationally)
  - Why are we social?
  - Why do we communicate?
- Language as a special case:
  - Phonetics/phonology/morphology, Syntax, Semantics, & Pragmatics.
- Natural Language Processing (NLP)

What AI Used to Think Language Was

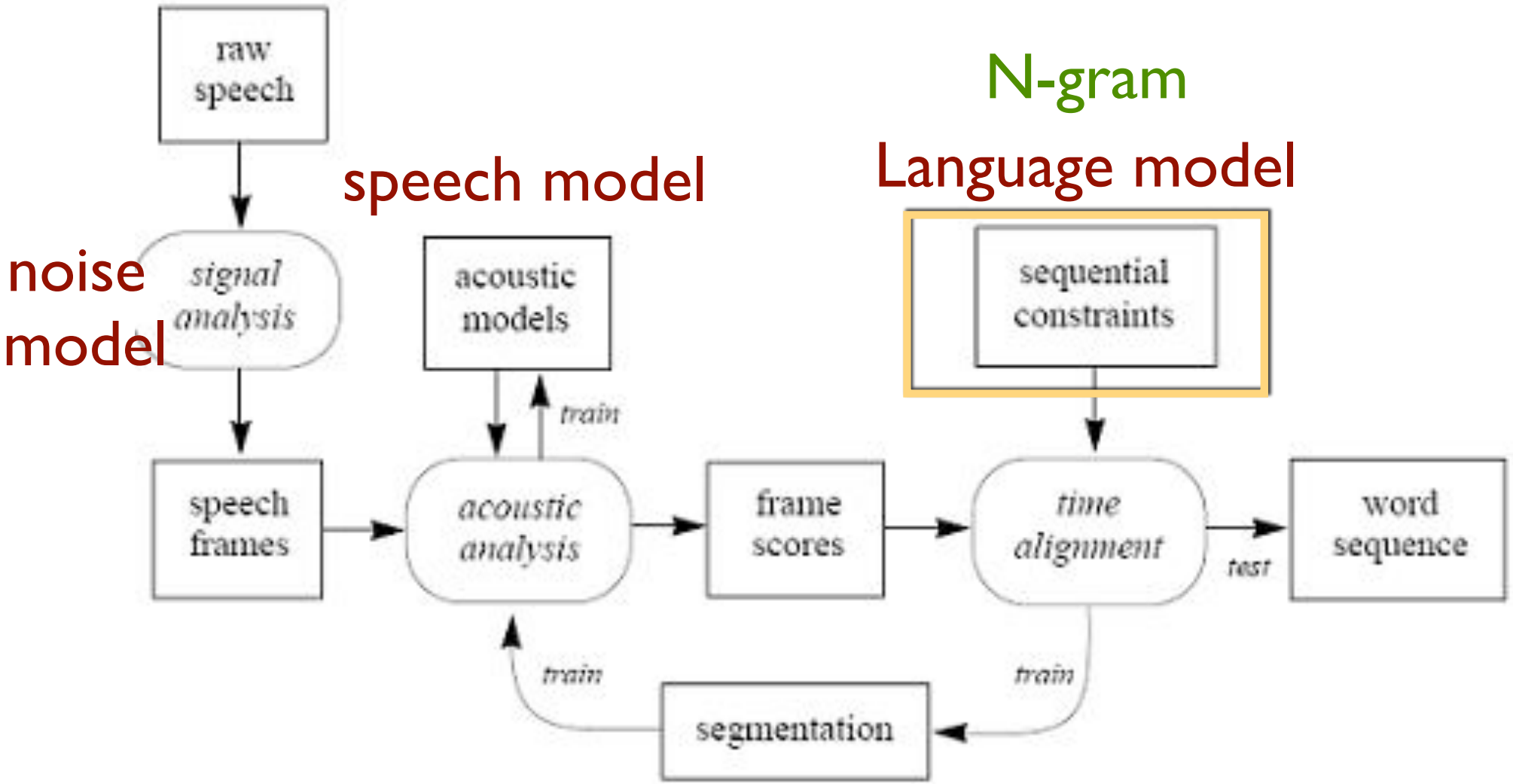
- Phonetics/phonology/morphology: what words (or subwords) are we dealing with?
- Syntax: What phrases are we dealing with? Which words modify one another?
- Semantics: What's the literal meaning?
- Pragmatics: What should you conclude from what was said? How should you act?

N-grams

- Large corpus technique for both language generation and speech recognition.
- Given previous N words, what is a probable following term? Memorise a sliding window through text.
  - Recognition: disambiguates parses.
  - Generation: just press go.

http://johno.jsmf.net/knowhow/ngrams/

Speech Recognition



http://www.learnartificialneuralnetworks.com/speechrecognition.html

Rooter: A Methodology for the Typical Unification of Access Points and Redundancy

Jeremy Stribling, Daniel Aguayo and Maxwell Krohn

ABSTRACT

Many physicists would agree that, had it not been for congestion control, the evaluation of web browsers might never have occurred. In fact, few hackers worldwide would disagree with the essential unification of voice-over-IP and public-private key pair. In order to solve this riddle, we confirm that SMPs can be made stochastic, cacheable, and interposable.

I. INTRODUCTION

Many scholars would agree that, had it not been for active networks, the simulation of Lamport clocks might never have occurred. The notion that end-users synchronize with the

The rest of this paper is organized as follows. For starter we motivate the need for fiber-optic cables. We place ou work in context with the prior work in this area. To ad dress this obstacle, we disprove that even though the much tauted autonomous algorithm for the construction of digital to-analog converters by Jones [10] is NP-complete, objec oriented languages can be made signed, decentralized, an signed. Along these same lines, to accomplish this mission, w concentrate our efforts on showing that the famous ubiquitou algorithm for the exploration of robots by Sato et al. runs i $\Omega((n + \log n))$ time [22]. In the end, we conclude.

II. ARCHITECTURE

accepted to the **World Multiconference on Systemics, Cybernetics and Informatics, 1995.**

http://pdos.csail.mit.edu/scigen/

Krohn cofounded OKCupid as an undergraduate

NATURE | NEWS

Publishers withdraw more than 120 gibberish papers

Conference proceedings removed from subscription databases after scientist reveals that they were computer-generated.

Richard Van Noorden

24 February 2014 | Updated: 25 February 2014

Note: probably more about a) reviewing & b) "academic" incentives esp. in China than NLP.

The publishers Springer and IEEE are removing more than 120 papers from their subscription services after a French researcher discovered that the works were computer-generated nonsense.

Over the past two years, computer scientist Cyril Labbé of Joseph Fourier University in Grenoble, France, has catalogued computer-generated papers that made it into more than 30 published conference proceedings between 2008 and 2013.

14 April 2015

SCIgen Architecture

[–] **wonkypedia** 8 points 16 hours ago

I got inspired by this and created my own. I used very basic Markov chains trained on a bunch of paper abstracts. The results seem pretty good if you have good training data.

What is under the hood on scigen?

permalink

[–] SCIgenAMA [S] 17 points 16 hours ago

Jeremy: we explicitly avoided Markov chains or anything else that was technically challenging, in the service of trying to make the papers as funny as possible. With Markov chains, you might get something syntactically correct, but it is likely to be boring.

With SCIgen, we literally sat around for two weeks and just brainstormed buzzwords, clauses, paragraph structures and other paper elements just based on what we thought would be funny. That's the grammar. Then SCIgen itself just goes through the grammar and makes random choices to fill stuff in. That's why you see things like "a testbed of Gameboys" in the evaluation sections sometimes -- we just thought it would be hilarious.

permalink parent

https://www.reddit.com/r/IAmA/comments/3210ym/at_mit_we_created_scigen_which_generates/

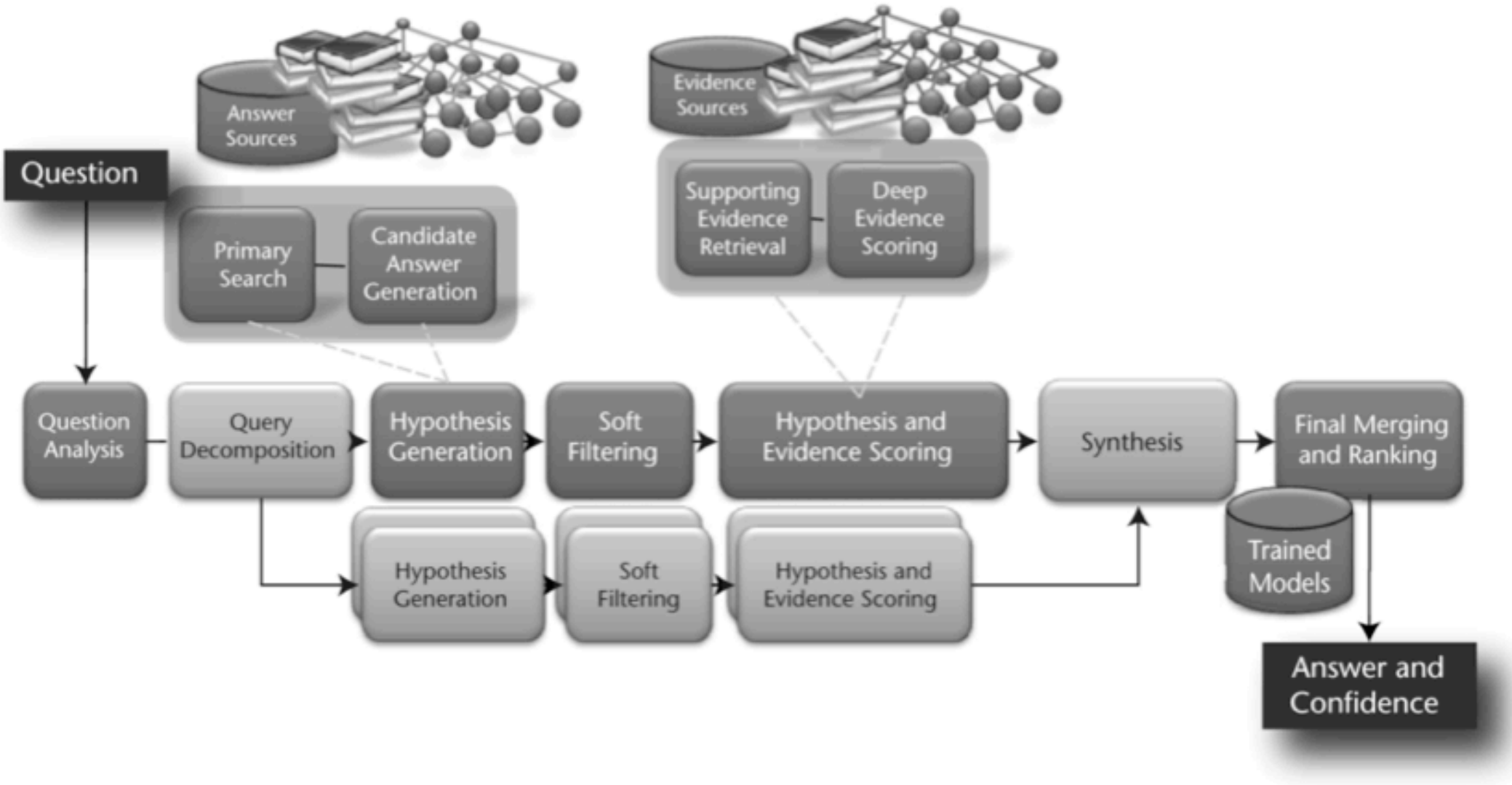
Generally, Still Need 'Real' Natural Language Processing (NLP)

- Negation.
- Referents for "this" and "that".
- Recognising multiple meanings for single word.
- Motivation, meaning tracking, turn taking.
- Ethics (not propagating stereotypes).

Cognitive Systems

Jeopardy vs Watson

Videos via Bath graduate, Dale Lane



(Ferrucci *et al.*, AI Magazine 2010)

Summary

- Culture is a powerful process for sharing intelligence / the output of cognition.
- Language is particularly effective at that.
- NLP is hard, but getting there.

$\Longrightarrow$ AI can use our culture / exploit our cognition.

- cf. ethics & consciousness lectures.

Reminder: NLP in Games

- Template matching.
  - Mentioned in Believability lecture: play with Eliza as homework (M-x doctor on emacs)
- Dialog in narrative context (story telling).

Story Generation with Crowdsourced Plot Graphs

Boyang Li, Stephen Lee-Urban, George Johnston, and Mark O. Riedl

School of Interactive Computing, Georgia Institute of Technology
{boyangli; lee-urban; gjhonston3 riedl}@gatech.edu

paper in AAAI 2013