# An Interdisciplinary Survey of Word Learning Research

Harlan D. Harris
Columbia University
Language and Cognition Lab
harlan@psych.columbia.edu

~~November 2003~~ ~~January~~ ~~February~~ March 2004
Occasional Talks in Speech, Language, and Cognition
Department of Computer Science, Columbia University

# Argument

Learning words in the real world seems like it ought to be really hard. But children become remarkably good at it. Word learning in NLP and AI is not as advanced.

We (linguists and psychologists) now are starting to know enough about word learning to help us (computer scientists) start to build practical systems that use human-like techniques for learning words in grounded and embodied applications.

# Not Talking About...

- an implemented system,
- speech recognition,
- learning grammars,
- formal semantics,
- the web,
- WSJ corpora,
- Bayesian anything.

# Outline

- Theory
- Word Learning in NLP
- Word Learning in AI
- Word Learning in Psychology
- Applications and Discussion

# Outline

- Theory
- Word Learning in NLP
- Word Learning in AI
- Word Learning in Psychology
- Applications and Discussion

# Central Questions

- What does it mean to learn a word?
- What is difficult about learning new words?

# Types of Word Learning

- What words is the new word similar to?
  - *smite* ≈ *hit, kill, attack*
- What is the new word's syntactic properties?
  - *smite* is a Vt, *with*-PP frequent, past *smote*, pp. *smitten*
- What is the new word's semantics?
  - *smite(X, Y)* ≈ HIT(*X, Y*)
- Learning is a process
  - Incomplete/tentative knowledge
  - Production vs. comprehension

# Word Learning is Hard

- Indeterminacy of reference (Quine 1960)
  - Disambiguation is hard
  - Can always find alternative definitions consistent with experience (Weir, in press)


- Disambiguation seems to require significant skills and experience: e.g., joint attention, shared perspective, and plenty of repetition in different contexts (Naigles 2002).

# Theory Sum-Up

- What does it mean to learn a word?
  - Link lexical form with semantic representation
- What is difficult about learning new words?
  - Indeterminacy of reference

# Outline

- Theory
- Word Learning in NLP
  - Learning from Linguistic Context
  - Learning from Semantic Context
- Word Learning in AI
- Word Learning in Psychology
- Applications and Discussion

# Lexical Word Learning Tasks

- Identify/segment words/morphemes in text
- Find POS, subcategorization from syntax
- Find similarity structure in syntax/semantics
  - Latent Semantic Analysis (Landauer and Dumais, 1997) – Multidimensional scaling to extract similarity metric from text
  - Hierarchical concepts -- Hypo-/syno-/hyper-nyms
- Bootstrap from discourse
  - Ehrlich and Rapaport (1997) – Induce logical representations of semantics from syntax heuristics in narrative NLU

# Semantic Word Learning

- Goal: Given paired text/semantics, induce semantics of new words
- Thompson and Mooney (1998, 2003)
  - Natural language and Prolog queries
  - Find common substructures of queries
    - “What is the largest...?”
    - Largest(x, ...)
  - Greedy search for a set of constructions that cover the Prolog set

# Siskind (1996)

Incremental cross-situational learning (Pinker 1989), given sentence and set of possible-meaning predicate representations.

- 3 processes:
  - Use known words to account for known semantics.
  - Maintain version space list of unaccounted-for semantic terms for each unknown word.
  - Look for semantic representations that match the semantic terms identified.

John *took* the *ball.*

CAUSE(John,
GO(ball, TO(John)))

CAUSE(X,
GO(ball, TO(X)))

{CAUSE, GO, ball, TO}

CAUSE(X,GO(Y,TO(X)))

# NLP Sum-Up

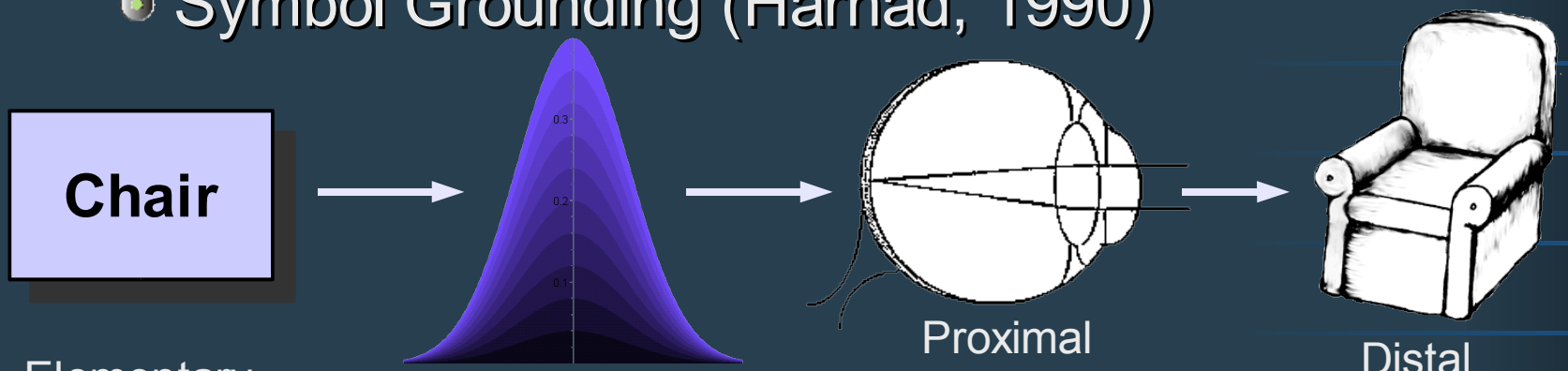
- What does it mean to learn a word?
  - Discern statistical patterns about the word's context and usage
  - Translate between text and a formal semantics
- What is difficult about learning new words?
  - Systems tend to learn syntactic properties, or highly-constrained semantic properties
  - Tasks tend to be analytical and special-purpose, not communicative and general-purpose

# Outline

- Theory
- Word Learning in NLP
- Word Learning in AI
  - Embodied Cognition
  - Grounded Word Learning
- Word Learning in Psychology
- Applications and Discussion

# Embodied Cognition

- Intelligent agents (including people) acting *in the world*, not just on data
  - "This project calls for detailing the myriad ways in which cognition depends upon – is grounded in – the physical characteristics, inherited abilities, practical activity, and environment of thinking agents." (Anderson, 2003)
  - Symbol Grounding (Harnad, 1990)



# Grounded Words

- Some words are grounded transparently in perceptions
  - “blue”, “happy”, “above”, “sharp”, “salty”
- Some words are grounded in complex categories
  - “chair”, “vegetable”, “concerto”, “swim”
- Some words are grounded in relation to other words and concepts
  - “uncle”, “revolt”, “should”, “happier”
- Ungrounded morphemes -- no semantics
  - **It** is raining. I **do** not like rain.

# Grounded Word Learning

- Induce meanings for words by associating with perceptions
  - Roy (1999) – CELL learns object names from audio and visual input

  - Roy (2002) – DESCRIBER learns to describe colored rectangles on a screen (also, Regier 1996)
  - Steels (1999) – Talking Heads system creates new words for objects in a language game.

# AI Sum-Up

- What does it mean to learn a word?
  - Ground symbols in raw percepts, communications
- What is difficult about learning new words?
  - Systems tend to work from first principles
  - Tasks limited to games, observing correlations
  - Extremely small-scale

# Outline

- Theory
- Word Learning in NLP
- Word Learning in AI
- Word Learning in Psychology
  - Background
  - Biases and heuristics
  - Statistics
  - Language and Thought
- Applications and Discussion

# Rates of Word Learning

- First produced word around 13 months
- Rate of word learning grows roughly linearly (Bloom 2001)
- 10 words/day for school-age children
- 20,000 – 100,000 words for adults
- Productive vocabulary significantly (5-9 months) lags comprehension vocabulary (Goodman 2001)
  - Siskind (1996) – Find terms first, build usable predicates second.

# First Words and Biases

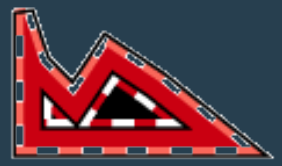
- Objects (kitty), people (mama), relations (up), social (hi, no)
- Initial words – easy words? Inherantly easy? Innate abilities?
- Later words – become easy? Developing innate abilities? Acquired new abilities?
- Biases both theoretically and practically necessary for (word) learning.
- What are the restrictions on the hypothesis space that allow people to solve the mapping problem?

# Object Bias

- Heuristic: New words refer to whole objects, not parts, actions, attributes, relations (Markman 1989, Golinkoff et al. 1995)
  - Train: dax hopping
  - Test: dax standing vs. wug hopping
- *Gavagai* means Rabbit
- But, less than half of early words are object names. (Bloom et al. 1993)
- Bias towards learning object labels correctly, not towards learning object labels.

# Shape Bias

This is a dax.



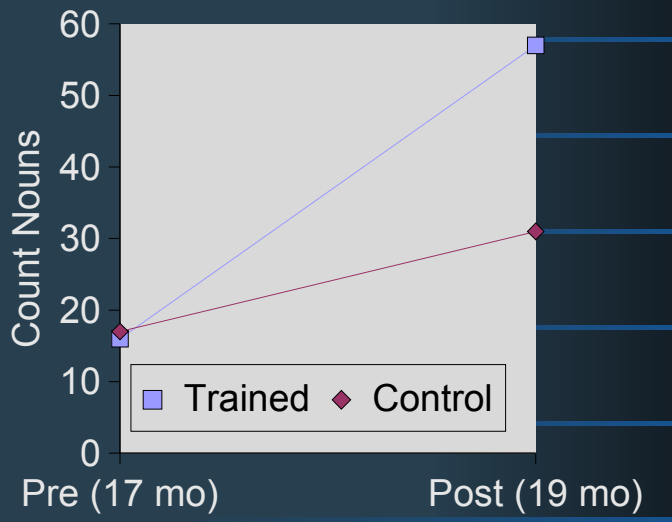
Show me the dax.

(from Linda Smith's web page)

- Generalize to shape, regardless of material or size. (Landau, Smith, and Jones, 1988)
- Only true for count nouns/object names, not for similarity judgements!

# Learning Biases

- Smith et al. (2000)'s “Associative Crane”
  - Learn a few count nouns through extensive observation, trial and error.
  - Note that shape seems to be most relevant.
  - For new words, try generalizing based on shape first.
- Taught infants 8 shape-extendable objects in lab...
- show huge effects in real-world learning!




- Meta-learning / automatic bias learning in ML (e.g. Baxter 2000, Vilalta and Drissi 2002)

# Verb Learning

- Verbs have complex syntax as well as complex semantics
- Idiosyncratic verb constructions (Tomasello 1992, Akhtar and Tomasello 1997)
  - “Mommy break” but not “Break cup”
  - Different patterns for different verbs
- Verbs are first memorized, then generalized
- Related to Representational Redescription (Karmiloff-Smith 1992) and Inductive Logic Programming

# Attention and Social Cues

- Social-Pragmatic approach to language acquisition (Tomasello 2001)
  - Goal is to decipher communicative intentions, not to solve mapping problems
  - E.g., verbs used as imperatives, not to describe world (Tomasello & Kruger 1992)
  - Knowledge of speaker's attention important to learn meaning (Bloom 2000; Yu, Ballard & Aslin 2003)

# Statistics

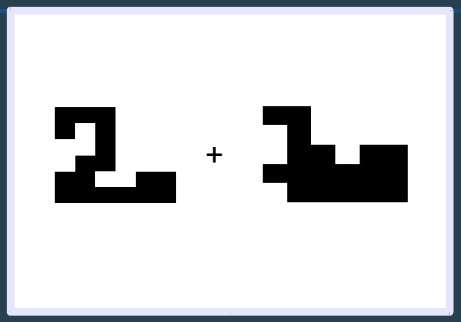
- Saffran et al. (1996) et al.
  - Infants and adults can learn to recognize word-like items presented in an unsegmented sequence:
    - pidokubatigepidokuterami...
  - Conditional probs. only clue to item boundaries
    - P(*do* | *pi*) >> P(*ba* | *ku*)
  - Any stimulus/modality tested works
  - Suggested that may be important component of word identification/learning (Saffran 2001 shows that infants view novel items in English frame as more salient than non-English frame)

# Lexical Status of Statistical Chunks

- (Magnuson and Harris, WIP)

- What processes and experiences lead to new words being added to the lexicon?
- Under what conditions does statistical word segmentation contact the lexicon?
- How important is reference (semantics) in word learning?
- What sorts of exposure to proto-words affect measures of familiarity (frequency)?

# Lexical Status Experiment Design

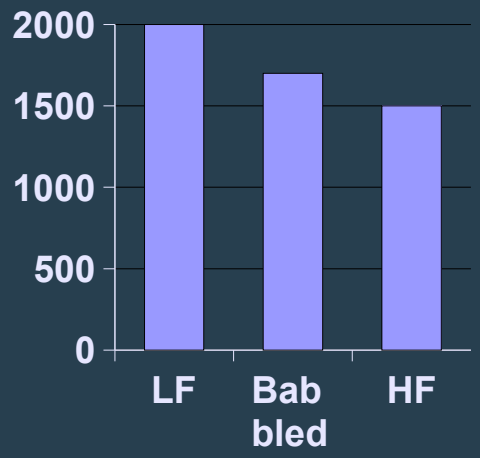


Learn Items

Babbling

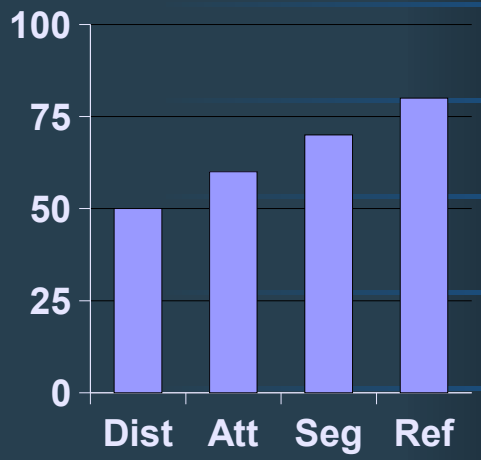
8 total CVCVs
2 HF
6 LF (2 in babbling)

Distracted | Attention | Segmented | Reference



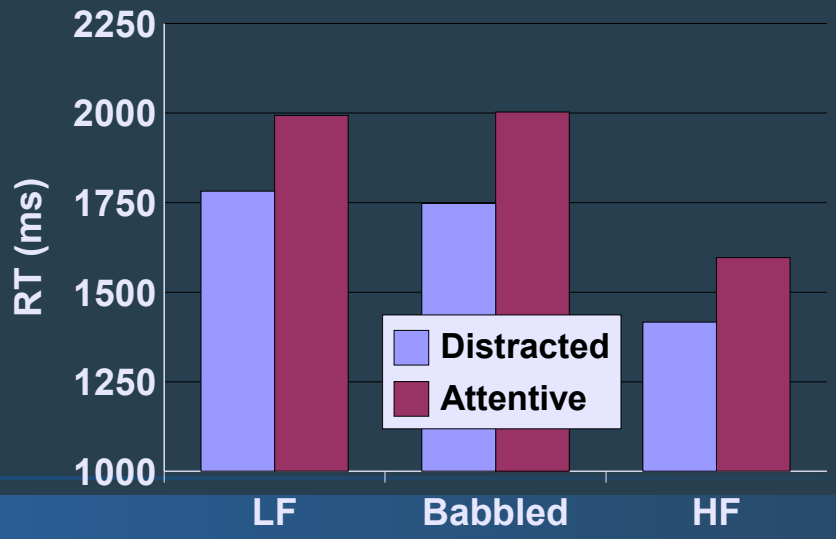

Test Items

Test Babbling



# Preliminary Results

- 13 subjects, ½ of design, Distracted and Attentive conditions only...
- 63% on Test Babbling task – are segmenting and remembering at least some of the items across a 2-3 minute distraction ($p < .001$)
- No evidence for contact w/ lexicon so far...



# Division of Dominance

- (Gentner and Boroditsky 2001)
- Cognitive Dominance: “concepts arise from the cognitive-perceptual sphere and are simply named by language.”
  - tend to be more open-class linguistically, easily individuated / learned
- Linguistic Dominance: “clumping is not pre-ordained, and language has a say in how the bits get conflated into concepts.”
  - tend to be more closed-class linguistically, hard to individuate / learn

proper names

concrete nouns

kinship terms

verbs

spacial preps.

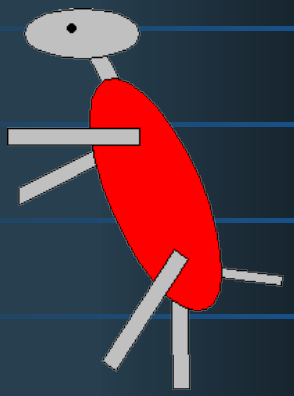
dets. conjs.

# Proper Names

- What is the meaning of a proper name?
  - A label/referent only? (“light semantics”)
  - Everything known about the referent? (“heavy semantics”)
- Psychology has prefered the former (Hollis and Valentine 2001), but formal semantics has recently come to favor the latter (Hurford 2003)
- Exp.: How do expectations from linguistic context affect the rate of learning identical concepts? How do those expectations affect attention to non-distinguishing features?

# Proper Names Details

- (Harris and Magnuson, WIP)
- Train subjects on names (proper, common) for aliens, measure learning rate, attention to relevant attributes.
- Vary linguistic context and category structure:
  - “This is a mark” vs. “This is Mark”
  - 4 different individual aliens vs. 4 groups of 4 similar aliens
- Extremely preliminary unreliable results:
  - Learn slower in PN linguistic context
  - Pay attention to more attributes when learning individuals



# Psych Sum-Up

- What does it mean to learn a word?
  - Be able to use the word in productive communication
- What is difficult about learning new words?
  - Ambiguity, cognitive and articulatory restrictions
  - But it gets much easier

# Outline

- Definitions
- Word Learning in NLP
- Word Learning in AI
- Word Learning in Psychology
- Applications and Discussion
  - Shared environments
  - Word-learning by useful agents

# Shared Embodied Environments

- Real World (Robotics)
  - Steels, Roy, etc.
- Virtual Environments
  - Training, education, simulation
  - Partially hand-grounded symbols
  - Schuler (2001) – Environment-based disambiguation
    - Simulated agents repairing a simulated jet engine
    - Semantics from environment disambiguate (e.g., PP-attachment)

# Word Learning Methods

- Observation -- “Barb, pass me that shiny thing.”
  - Paired meaning-language
  - Lots of ambiguity
- Instruction -- “This thing is shiny.”
  - Clearer semantics, less ambiguity
- Definitions -- “ 'Shiny' means reflecting light.”
  - Cheap, typically ungrounded
- Language games -- “Is this shiny? No? Is that?”
  - Grounded, emergent

# Conclusions

- Building intelligent systems that learn words in embodied environments will be challenging, but valuable.
- Language acquisition research provides interesting inspiration and valuable constraints on efficient lexical learning processes:
  - Incremental, cross-situational learning
  - Innate and acquired biases
  - Meta-learning of relevant features
  - Variety of learning methods
  - Shared reference, social/pragmatic cues

# Learning Concepts and Words

- Whorf (1956) – Linguistic Determinism
  - “We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world... [are organized] largely by the linguistic systems in our minds.”
- Neo-Whorfianism – language has significant influences on cognition, but the cross-linguistic differences are generally minor and unrelated to general intelligence.