Semantic Parsing: Past, Present, and Future

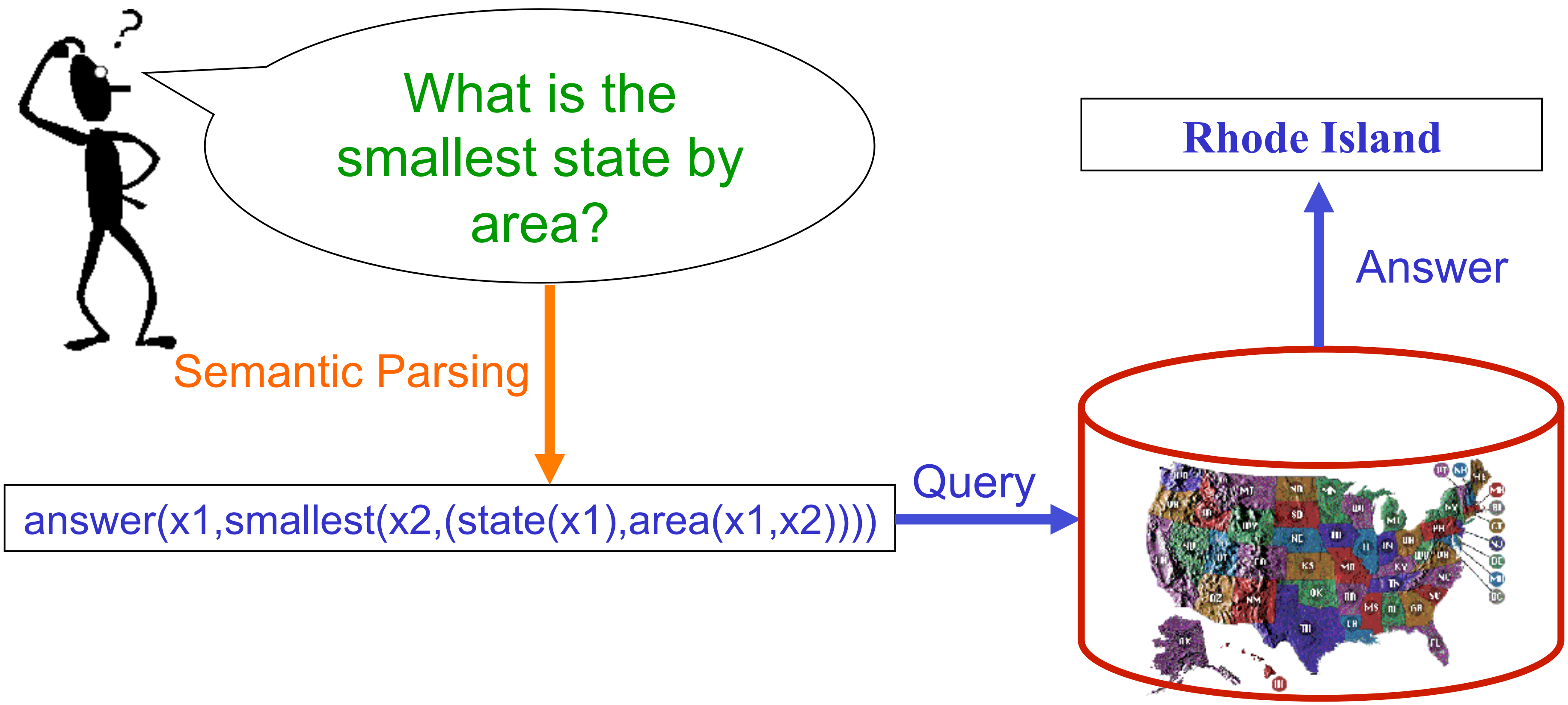
Raymond J. Mooney
Dept. of Computer Science
University of Texas at Austin

What is Semantic Parsing?

- Mapping a natural-language sentence to a detailed representation of its complete meaning in a fully formal language that:
  - Has a rich ontology of types, properties, and relations.
  - Supports automated reasoning or execution.

Geoquery:
A Database Query Application

- Query application for a U.S. geography database containing about 800 facts [Zelle & Mooney, 1996]

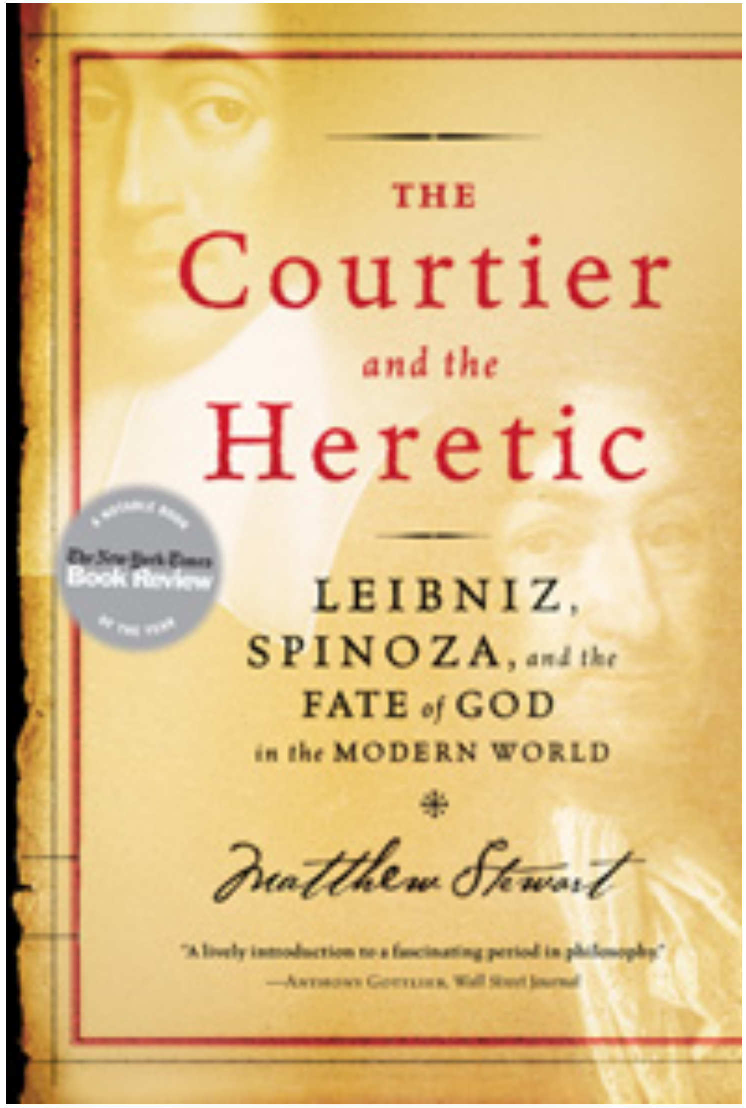


Prehistory 1600's

- Gottfried Leibniz (1685) developed a formal conceptual language, the *characteristica universalis*, for use by an automated reasoner, the *calculus ratiocinator*.

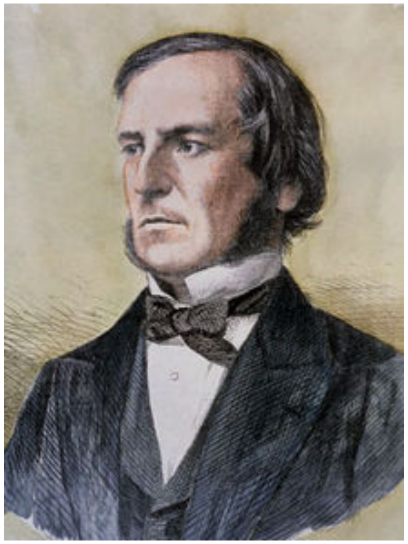
"The only way to rectify our reasonings is to make them as tangible as those of the Mathematicians, so that we can find our error at a glance, and when there are disputes among persons, we can simply say: Let us calculate, without further ado, to see who is right."

Interesting Book on Leibniz



Prehistory 1850's

- George Boole (*Laws of Thought*, 1854) reduced propositional logic to an algebra over binary-valued variables.



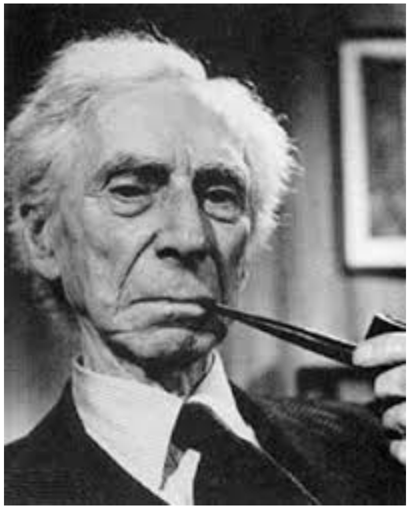
- His book is subtitled "*on Which are Founded the Mathematical Theories of Logic and Probabilities*" and tries to formalize *both* forms of human reasoning.

Prehistory 1870's

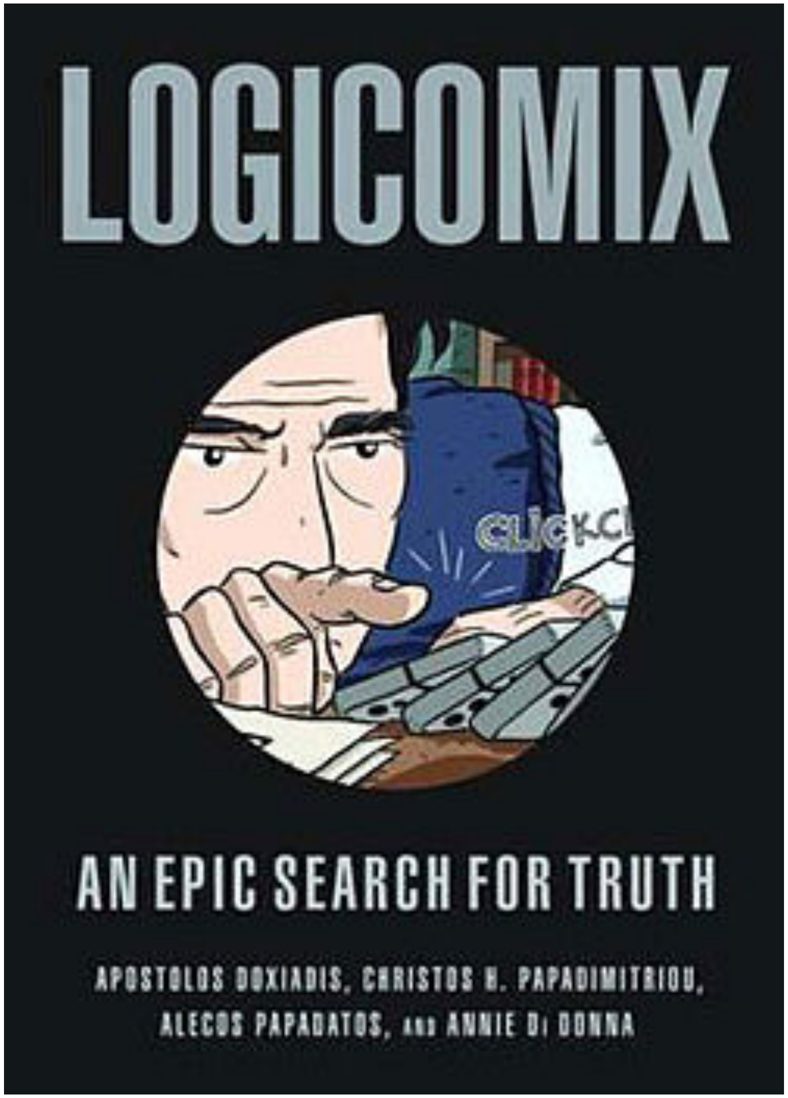
- Gottlob Frege (1879) developed *Begriffsschrift* (concept writing), the first formalized quantified predicate logic.

Prehistory 1910's

- Bertrand Russell and Alfred North Whitehead (*Principia Mathematica*, 1913) finalized the development of modern first-order predicate logic (FOPC).

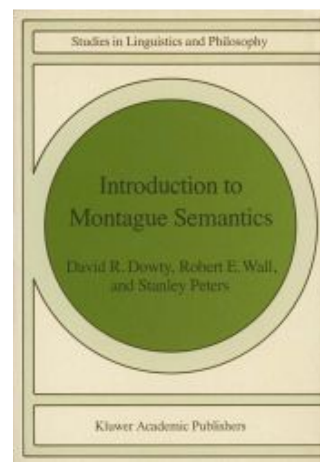


Interesting Book on Russell



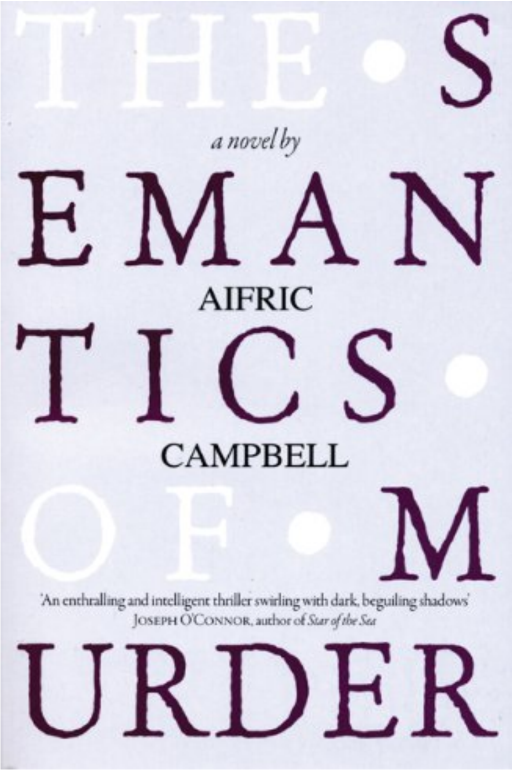
History from Philosophy and Linguistics

- Richard Montague (1970) developed a formal method for mapping natural-language to FOPC using Church's *lambda calculus* of functions and the fundamental principle of *semantic compositionality* for recursively computing the meaning of each syntactic constituent from the meanings of its sub-constituents.
- Later called "Montague Grammar" or "Montague Semantics"



Interesting Book on Montague

- See Aifric Campbell's (2009) novel *The Semantics of Murder* for a fictionalized account of his mysterious death in 1971 (homicide or homoerotic asphyxiation??).

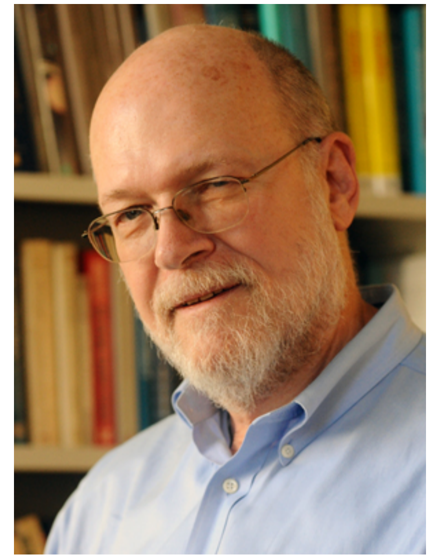


Early History in AI

- Bill Woods (1973) developed the first NL database interface (LUNAR) to answer scientists' questions about moon rooks using a manually developed Augmented Transition Network (ATN) grammar.

Early History in AI

- Dave Waltz (1975) developed the next NL database interface (PLANES) to query a database of aircraft maintenance for the US Air Force.
- I learned about this early work as a student of Dave's at UIUC in the early 1980's.



(1943-2012)

Early Commercial History

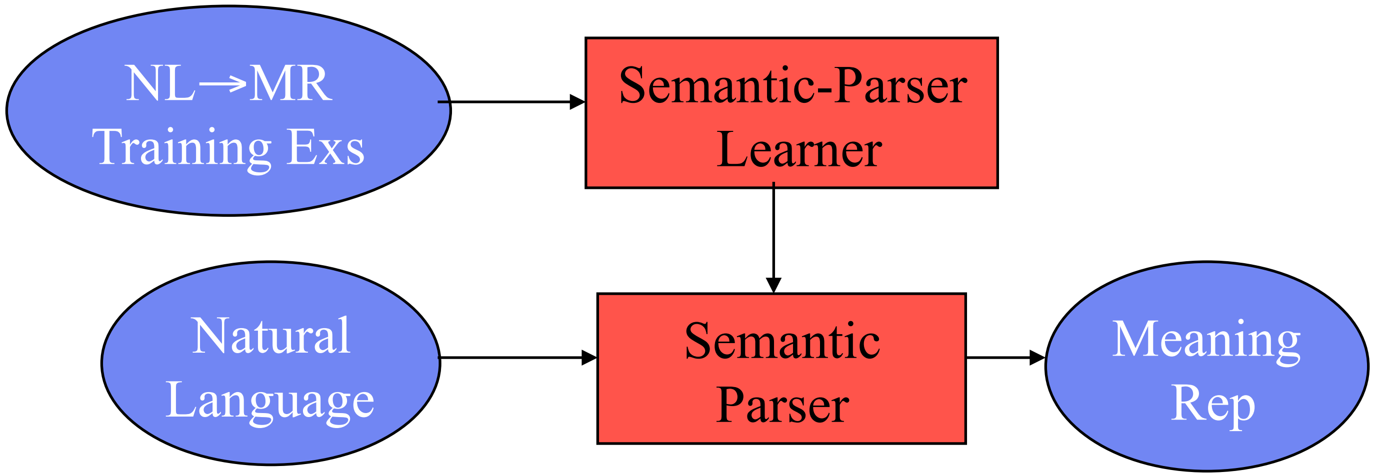
- Gary Hendrix founded Symantec ("semantic technologies") in 1982 to commercialize NL database interfaces based on manually developed *semantic grammars*, but they switched to other markets when this was not profitable.
- Hendrix got his BS and MS at UT Austin working with my former UT NLP colleague, Bob Simmons (1925-1994).

1980's: The "Fall" of Semantic Parsing

- Manual development of a new semantic grammar for each new database did not "scale well" and was not commercially viable.
- The failure to commercialize NL database interfaces led to decreased research interest in the problem.

Learning Semantic Parsers

- Manually programming robust semantic parsers is difficult due to the complexity of the task.
- Semantic parsers can be learned automatically from sentences paired with their formal meaning representations (MRs).



History of Learning Semantic Parsers

- I started working on learning semantic parsers in 1992 and by 2010 had 6 PhD's who finished their thesis on the topic.

- There was also work in the 1990's on learning semantic parsers for ATIS at BBN and elsewhere (Miller et al., 1994; Kuhn & DeMori, 1995).

Different Learning Approaches My Former Students Explored

- Inductive Logic Programming (CHILL, WOLFIE, COCKTAIL)



Lappoon Rupert Tang

John Zelle Cindy Thompson

- Probabilistic Synchronous Grammars (WASP)

John Yuk Wah Wong

- SVMs with String Kernels (KRISP)

Rohit Kate

- Integration with statistical syntactic parsing (SCISSOR, SYNSEM)

Ruifang Ge

Semantic Parsing Renaissance

- In 2005, Luke Zettlemoyer started developing a CCG-based approach to learning semantic parsers.

- I met Luke at AAAI-05, where he informed me that his interest in semantic parsing originated from learning of my work from his undergrad advisor at NC-State, UT alum James Lester.

Present Foci of Semantic Parsing

- Reducing supervision from full MR's to query-answer pairs, other types of weak supervision, or unsupervised learning.
- Scaling up to broader-coverage domains, such as querying Freebase.
- Grounded learning that connects language to perception and action.

Learning from Queries and Answers

- Instead of training on gold-standard MRs, just train on gold-standard answers:
  - “What is the state with the smallest area?”$\Rightarrow$ “Rhode Island”
- Using learning methods for structured output (Clarke et al., 2010).
- Using “dependency based” latent logical forms (Liang et al., 2011).

Learning from Mixed-Initiative Dialog

- Learn to parse harder sentences by using user interaction to break them down into simpler components through "clarification dialogs" (Artzi & Zettlemoyer, 2011).

SYSTEM: how can i help you? (OPEN TASK)
USER: i would like to fly from atlanta georgia to london england on september twenty fourth in the early evening i would like to return on october first departing from london in the late morning
SYSTEM: leaving what city? (ASK:x:from(fl; x))
USER: atlanta georgia
SYSTEM: leaving atlanta. (CONFIRM:from(fl; ATL)) going to which city? (ASK:x:to(fl; x))
USER: london
SYSTEM: arriving in london england. (CONFIRM:to(fl;LON)) what date would you like to depart atlanta? (ASK:x:from(fl; ATL) ^ departdate (fl; x))
USER: september twenty fourth in the early evening
[conversation continues]

Unsupervised Learning

- Use relational clustering of words and phrases to automatically induce a “latent” set of semantic predicates for types and relations from dependency-parsed text. (Poon & Domingos, 2008; Titov & Klementiev, 2011)

Scaling Up

- Several recent projects have focused on scaling up to databases with large ontologies/schemas like Freebase.
  - Use standard schema-matching techniques to extend the lexicon (Cai & Yates, 2013).
  - Augment a CCG parser with on-the-fly ontology matching (Kwiatkowski et al., 2013).
  - Learn to automatically add “bridging” predicates to the query (Berant at al., 2013).

Grounded Semantic Parsing

- Produce meaning representations that can be automatically executed in the world (real or simulated) to accomplish specific goals.
- Learn only from language paired with the ambiguous "real-world" context in which it is naturally used.

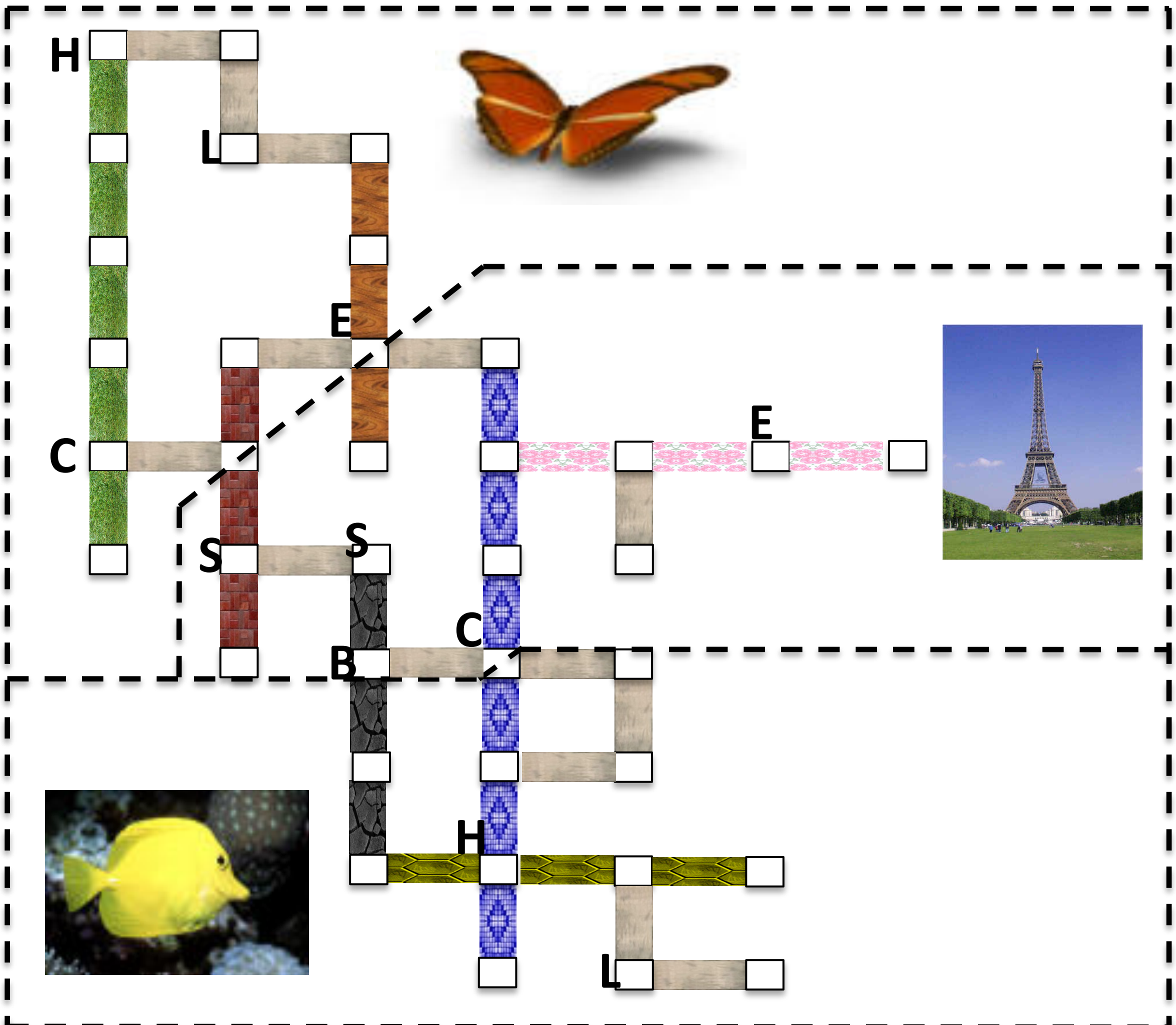
See my AAAI-2013 Keynote Invited Talk
on "Grounded Language Learning"
on videolectures.net

Learning to Follow Directions in a Virtual Environment

- Learn to interpret navigation instructions in a virtual environment by simply observing humans giving and following such directions (Chen & Mooney, AAAI-11).

- **Eventual goal:** Virtual agents in video games and educational software that automatically learn to take and give instructions in natural language.

Sample Virtual Environment
(MacMahon, et al. AAAI-06)



H – Hat Rack

L – Lamp

E – Easel

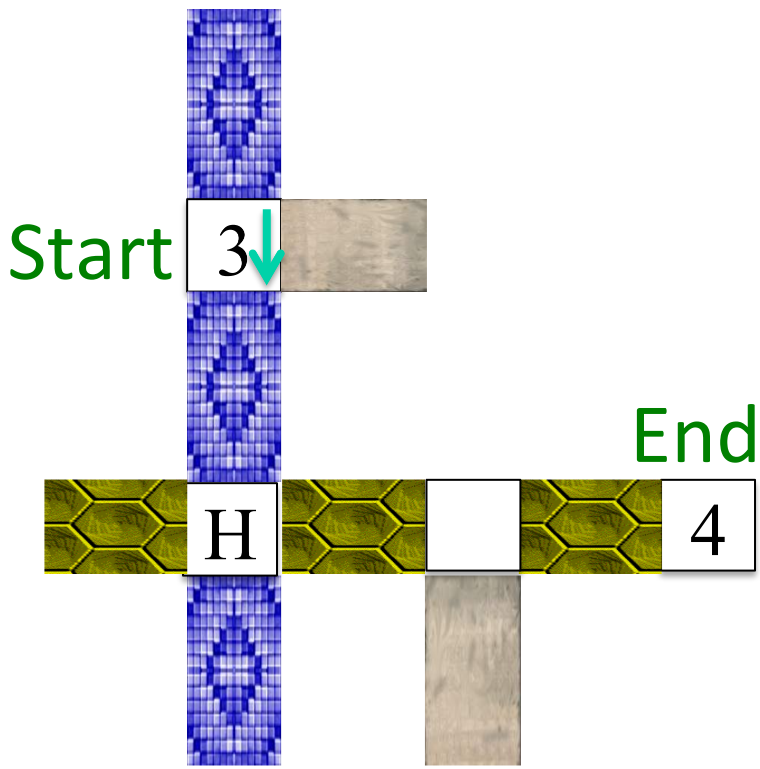
S – Sofa

B – Barstool

C - Chair

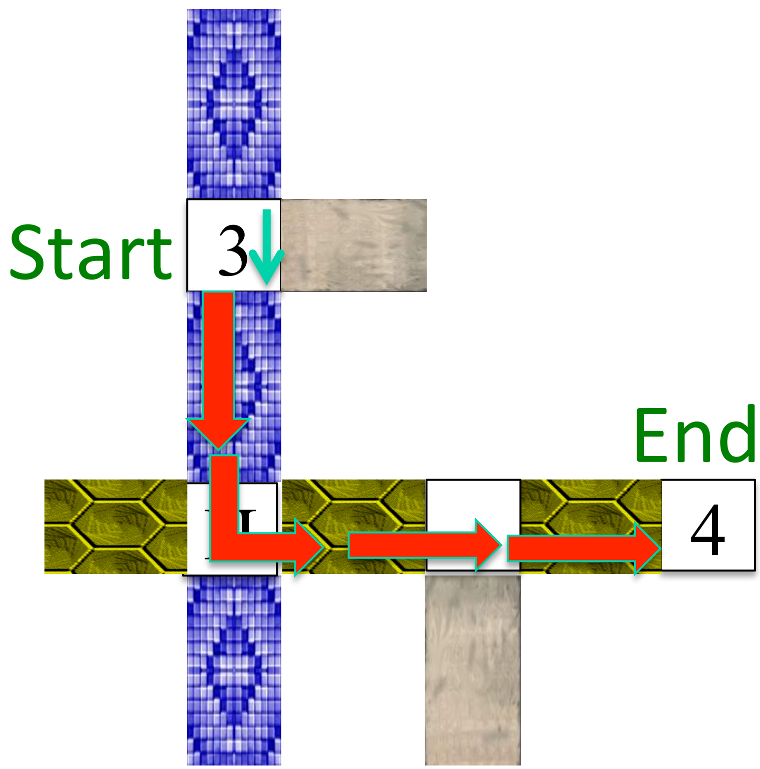
Sample Navigation Instructions

•Take your first left. Go all the way down until you hit a dead end.



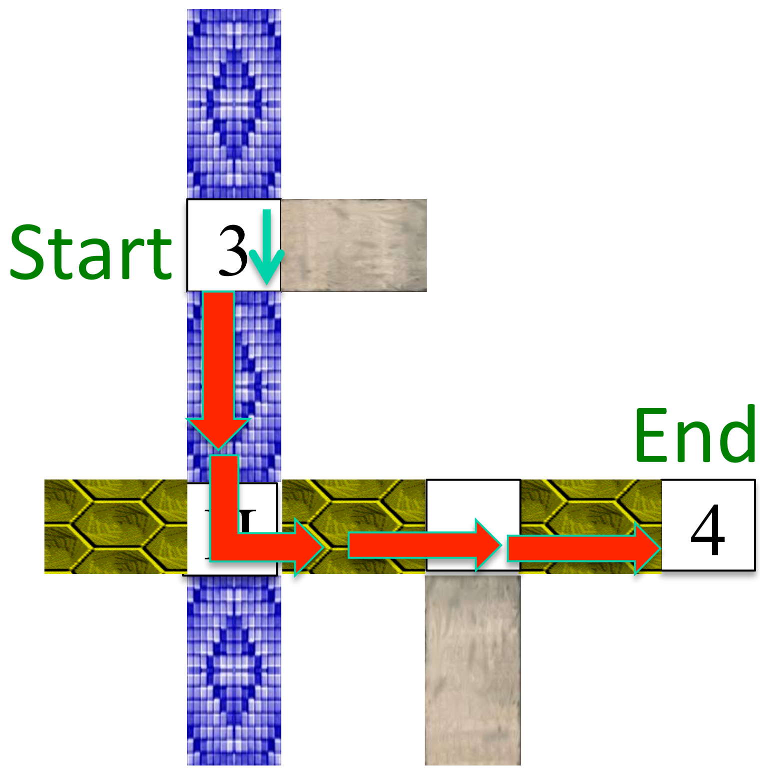
Sample Navigation Instructions

•Take your first left. Go all the way down until you hit a dead end.



Observed primitive actions:
Forward, Left, Forward, Forward

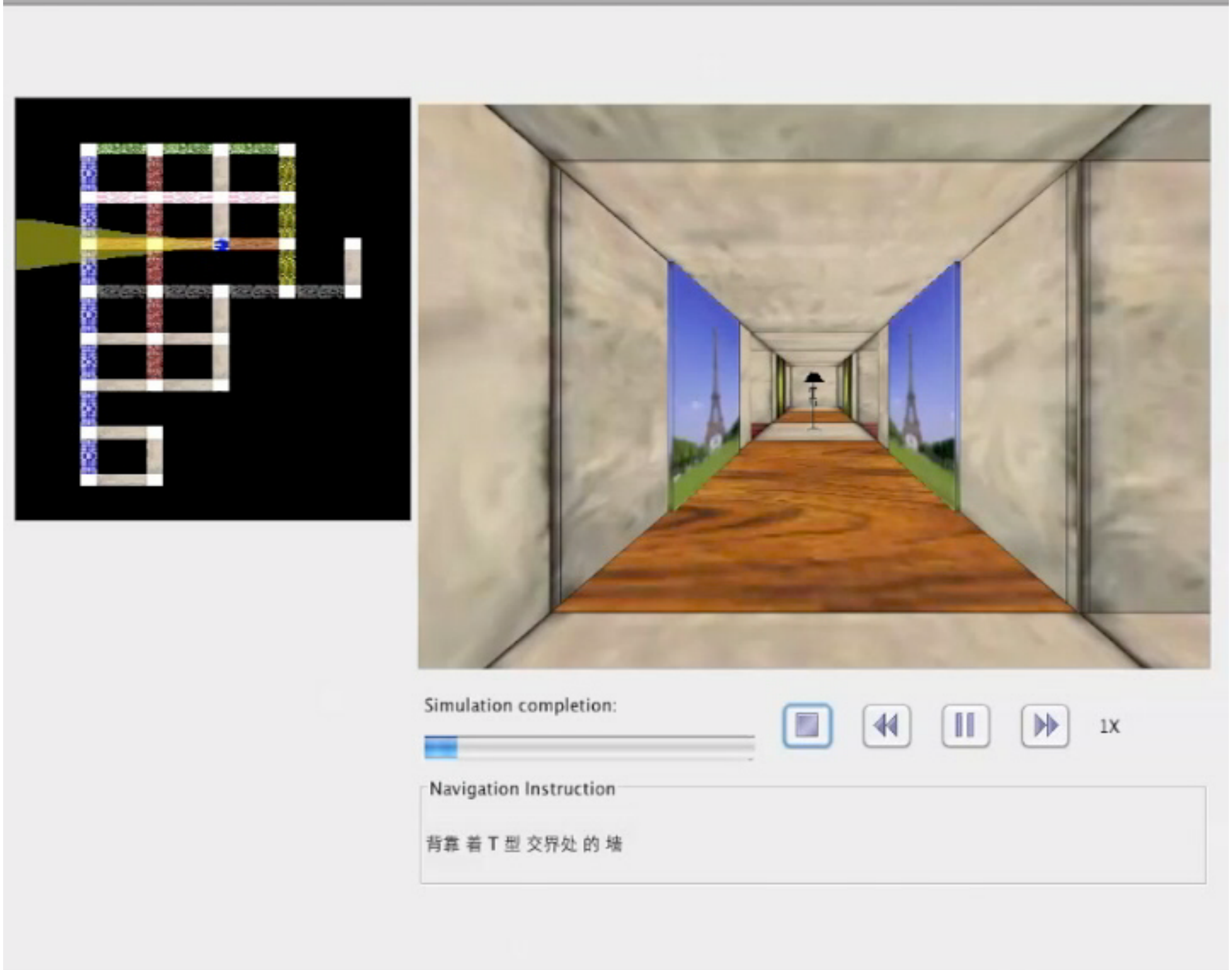
Sample Navigation Instructions



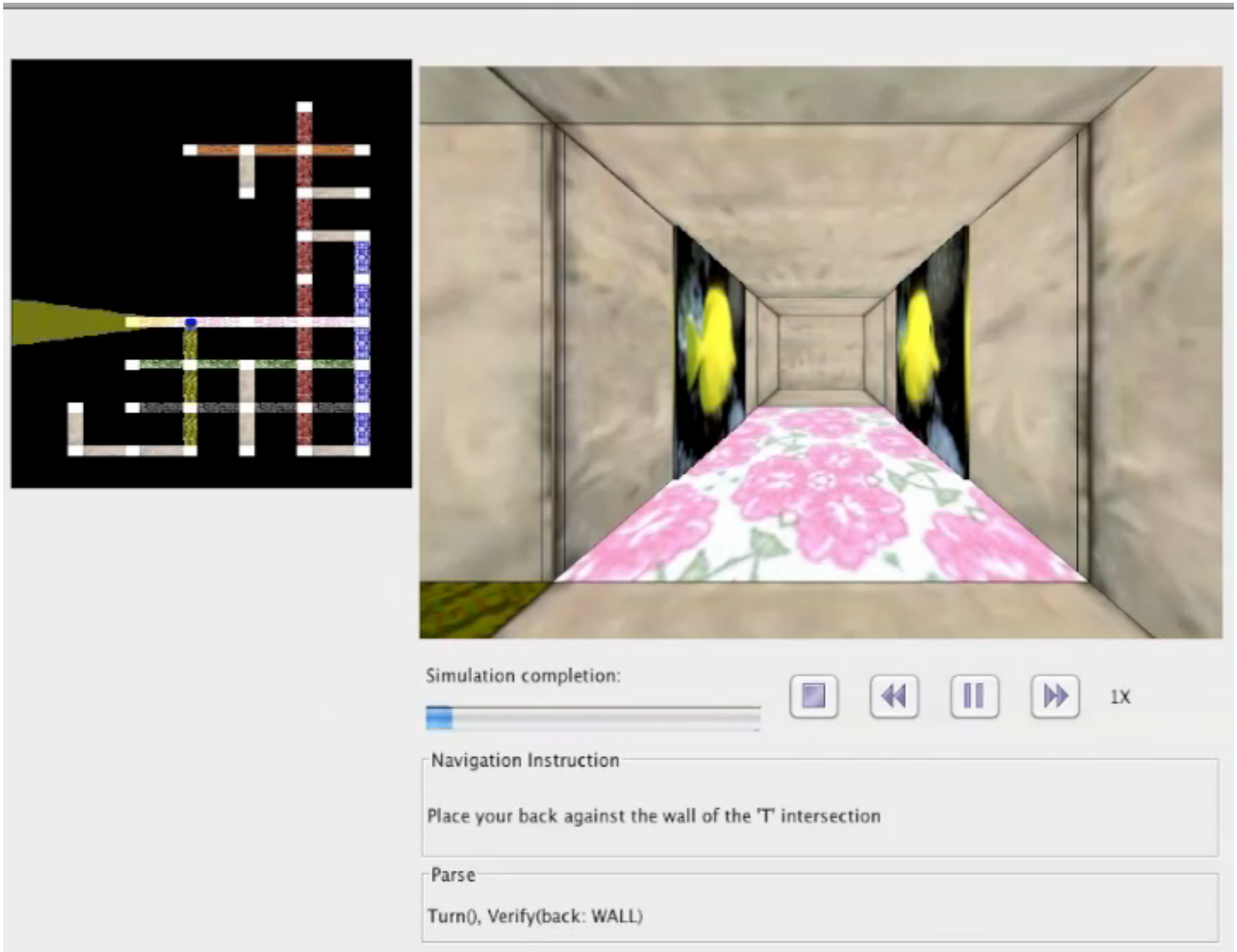
Observed primitive actions:
Forward, Left, Forward, Forward

- Take your first left. Go all the way down until you hit a dead end.
- Go towards the coat hanger and turn left at it. Go straight down the hallway and the dead end is position 4.
- Walk to the hat rack. Turn left. The carpet should have green octagons. Go to the end of this alley. This is p-4.
- Walk forward once. Turn left. Walk forward twice.

Observed Training Instance in Chinese



Executing Test Instance in English
(after training in English)



Navigation-Instruction Following Evaluation Data

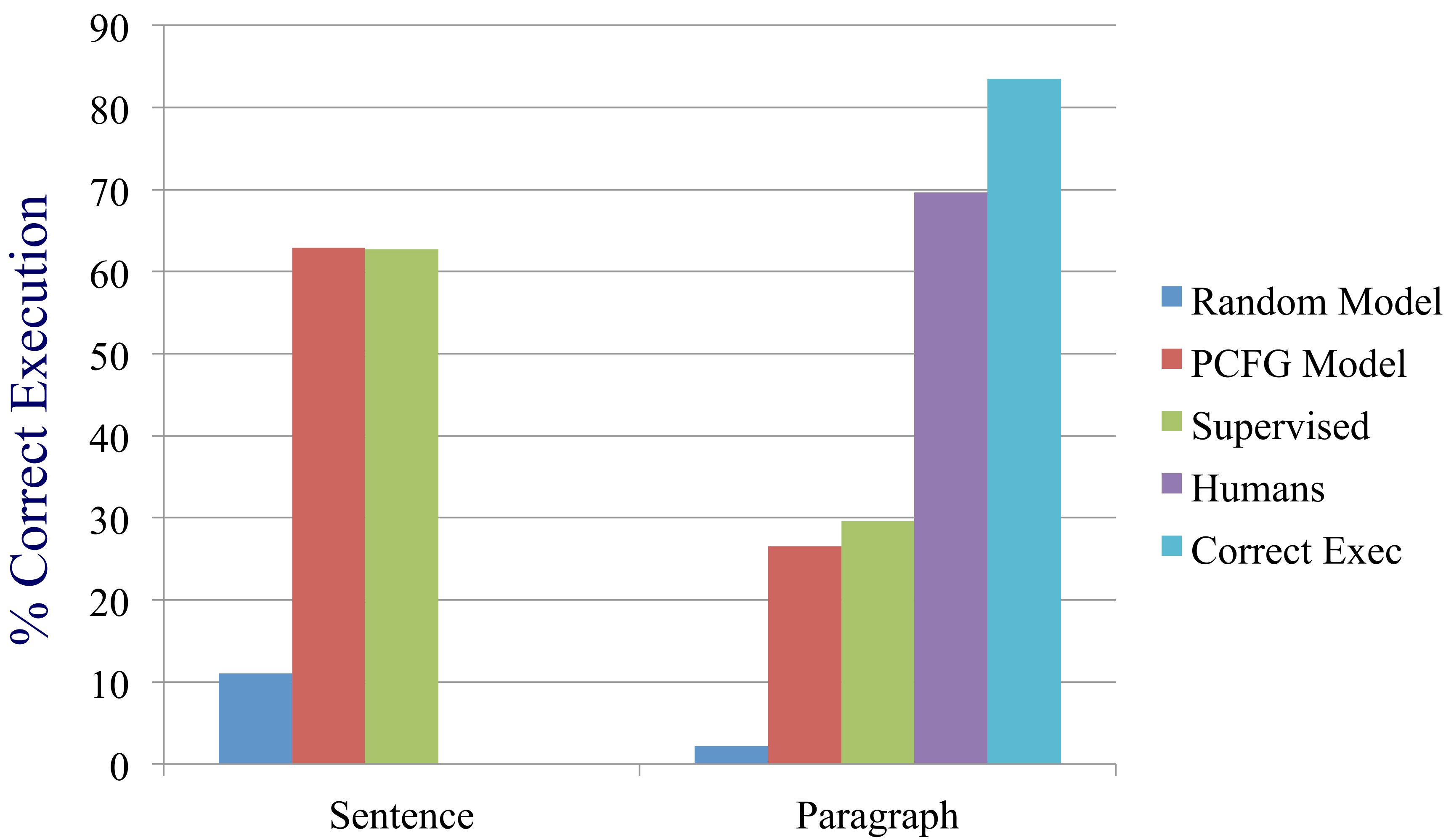
- 3 maps, 6 instructors, 1-15 followers/direction

	Paragraph	Single-Sentence
# Instructions	706	3,236
Avg. # sentences	5.0 (±2.8)	1.0 (±0)
Avg. # words	37.6 (±21.1)	7.8 (±5.1)
Avg. # actions	10.4 (±5.7)	2.1 (±2.4)

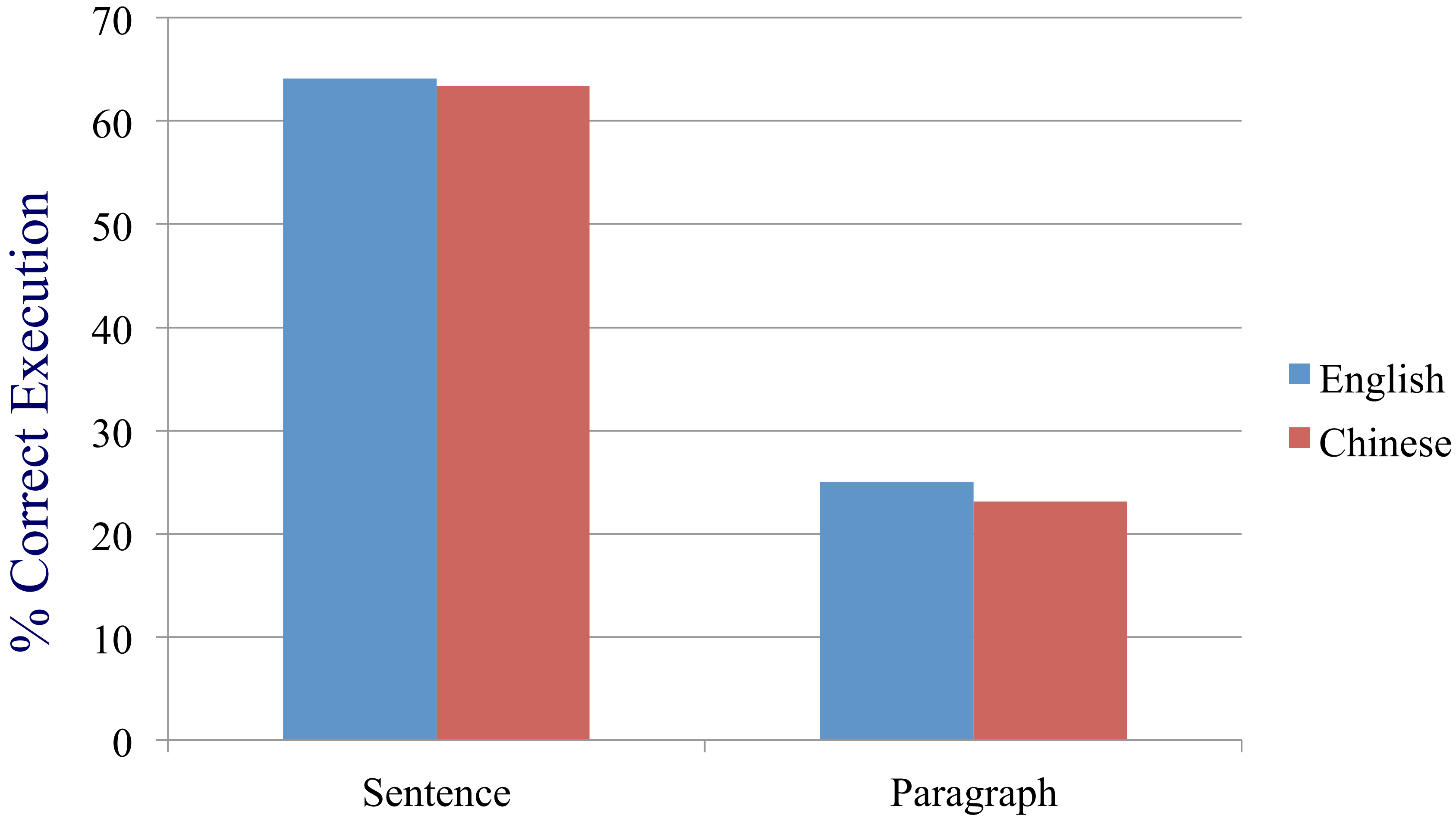
End-to-End Execution Evaluation

- Test how well the system follows new directions in novel environments.
  - Leave-one-map-out cross-validation.
- **Strict metric**: Correct iff the final position exactly matches goal location.
- **Lower baseline**:
  - Simple probabilistic generative model of executed plans without language.
- **Upper bounds**:
  - Supervised semantic parser trained on gold-standard plans.
  - Human followers.
  - Correct execution of instructions.

End-to-End Execution Results
English



End-to-End Execution Results
English vs. Mandarin Chinese



Other Work on Grounded Semantic Parsing

- See the final three talks of the workshop:
  - Asking for Help Using Inverse Semantics
    Stefanie Tellex
  - Computing with Natural Language
    Percy Liang
  - Grounded Semantic Parsing
    Luke Zettlemoyer

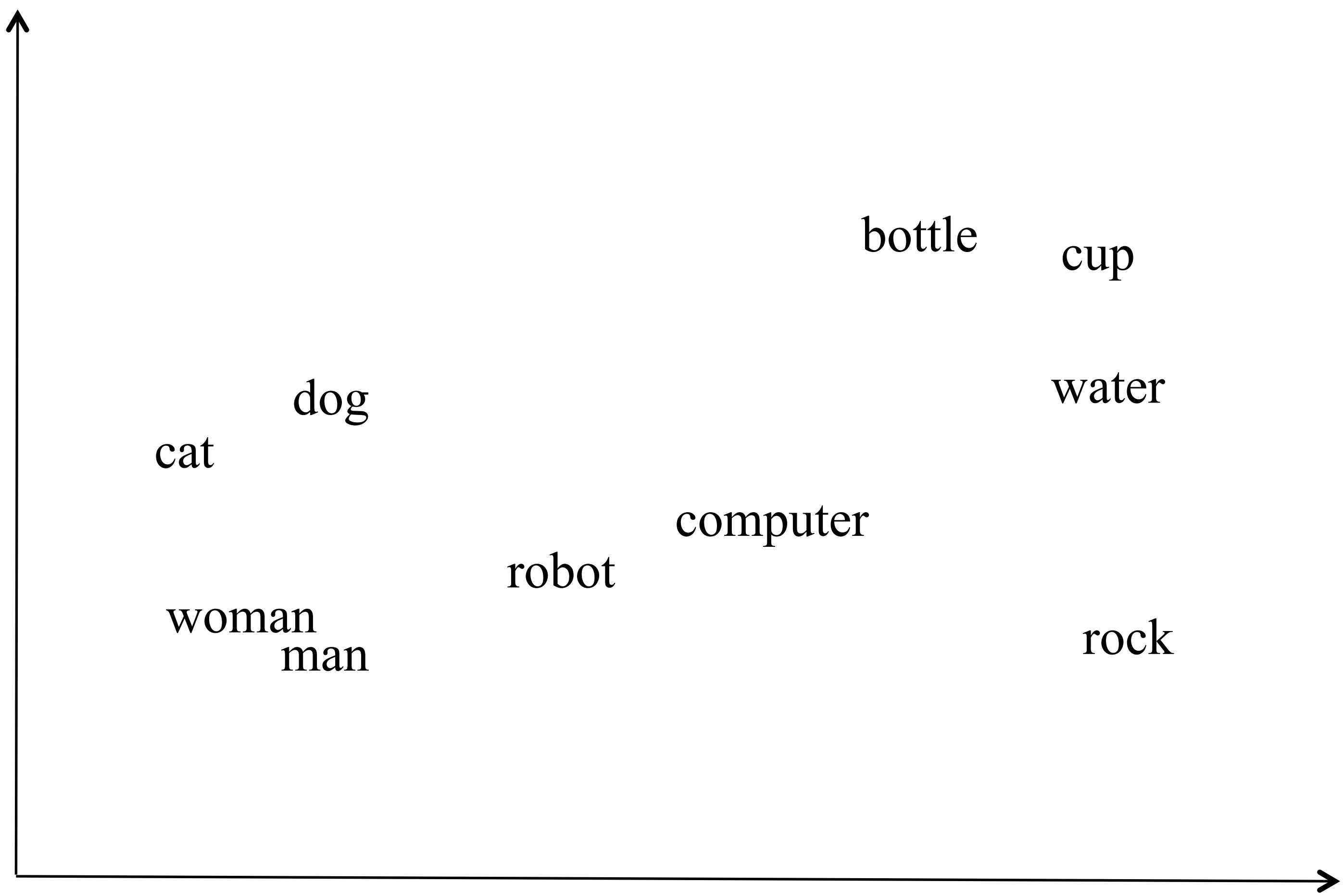
Future:
Integrating Logical and Distributional Semantics

- Standard semantic parsing requires being given or creating a fixed ontology of properties and relations with binary truth-values.
- Developing a broad-coverage ontology is difficult.
- Does not account for the “graded” (non-binary) nature of linguistic meaning.

Distributional (Vector-Space) Lexical Semantics

- Represent word meanings as points (vectors) in a (high-dimensional) Euclidian space.
- Dimensions encode aspects of the context in which the word appears (e.g. how often it co-occurs with another specific word).
- Semantic similarity defined as distance between points in this semantic space.
- Many specific mathematical models for computing dimensions and similarity
  - 1st model (1990): Latent Semantic Analysis (LSA)

Sample Lexical Vector Space



Issues with Distributional Semantics

- How to compose meanings of larger phrases and sentences from lexical representations? (many recent proposals…)
- None of the proposals for compositionality capture the full representational or inferential power of FOPC (Grefenstette, 2013).

"You can't cram the meaning of a whole %&!$# sentence into a single $&!#* vector!"

Using Distributional Semantics with Standard Logical Form

- Recent work on unsupervised semantic parsing and Lewis and Steedman (2013) automatically create an ontology from distributional information but do not allow gradedness and uncertainty in the final semantic representation and inference.

Formal Semantics for Natural Language using Probabilistic Logical Form

- Represent the meaning of natural language in a formal ***probabilistic*** logic (Beltagy et al., 2013, 2014).
  - Markov Logic Networks (MLNs)
  - Probabilistic Similarity Logic (PSL)

"Montague meets Markov"

Markov Logic

(Richardson & Domingos, 2006)

- Set of weighted clauses in first-order predicate logic.
- Larger weight indicates stronger belief that the clause should hold.
- MLNs are templates for constructing Markov networks for a given set of constants

MLN Example: Friends & Smokers

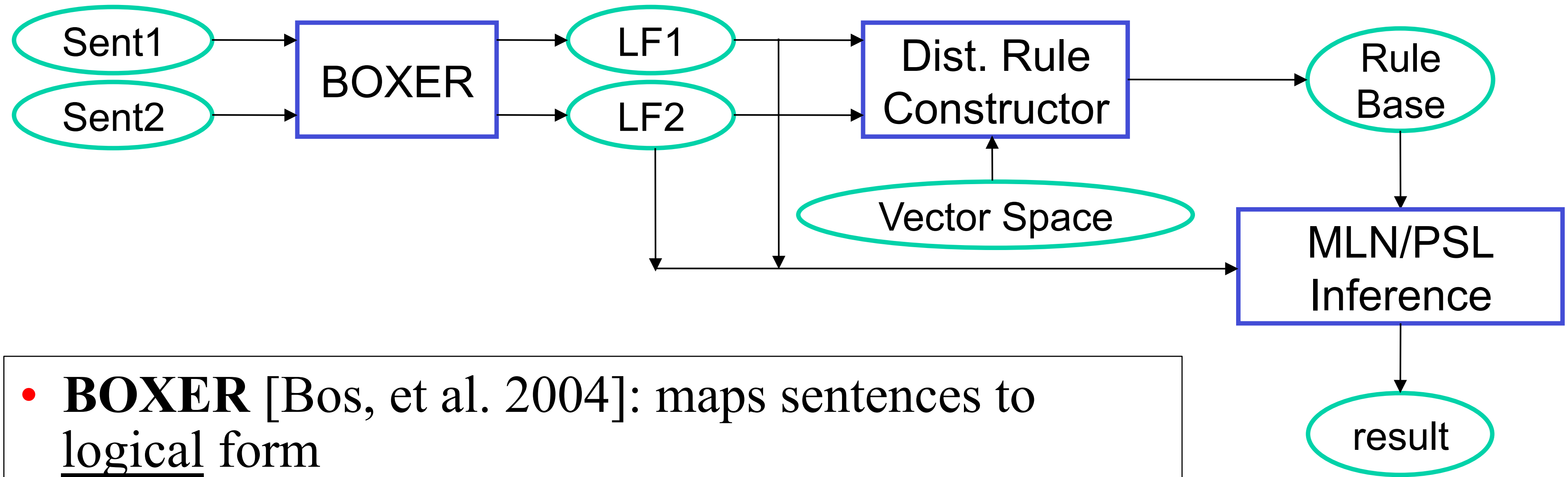
1.5	$\forall x \; Smokes(x) \Rightarrow Cancer(x)$
1.1	$\forall x, y \; Friends(x, y) \Rightarrow \big(Smokes(x) \Leftrightarrow Smokes(y)\big)$

Markov Logic Inference

- Infer probability of a particular query given a set of evidence facts.
  - P(Cancer(Anna) | Friends(Anna,Bob),Smokes(Bob))

System Architecture

(Garrette et al. 2011, 2012; Beltagy et al., 2013, 2014)



- **BOXER** [Bos, et al. 2004]: maps sentences to logical form
- **Distributional Rule constructor**: generates relevant soft inference rules based on distributional similarity
- **MLN/PSL**: probabilistic inference
- **Result**: degree of entailment or semantic similarity score (depending on the task)

Sample RTE Problem

T: "A man is slicing a pickle."

$\exists x,y,z(man(x) \wedge slice(y) \wedge Agent(x,y) \wedge pickle(z) \wedge Patient(z,y))$

H: "A guy is cutting a cucumber."

$\exists x,y,z(guy(x) \wedge cut(y) \wedge Agent(x,y) \wedge cucumber(z) \wedge Patient(z,y))$

Compute P(H | T) in Markov Logic

Distributional Lexical Rules

- For every pair of words (*a, b*) where *a* is in *T* and *b* is in *H* add a soft rule relating the two.

$$\forall x\ (a(x) \rightarrow b(x)) \quad wt(a,b)$$

$$wt(a,b) = log\left(\frac{\cos(\vec{a}, \vec{b})}{1 - \cos(\vec{a}, \vec{b})}\right) - prior$$

$$\forall x\ (man(x) \rightarrow guy(x)) \quad wt(man,guy)$$

$$\forall x\ (slice(x) \rightarrow cut(x)) \quad wt(slice,cut)$$

$\vdots$

$\vdots$

For Details See Our Poster:

Beltagy, I., Erk, K., and Mooney, R.J., "Semantic Parsing using Distributional Semantics and Probabilistic Logic"

Conclusions

- Past: Semantic parsing has a long, rich history.
- Present: There is blossoming of recent work, particularly in reducing supervision, scaling up, and grounding.
- Future: It's bright, particularly for integrating distributional and logical semantics.

Thanks to Yoav, Tom, and Jonathan for organizing this exciting workshop!