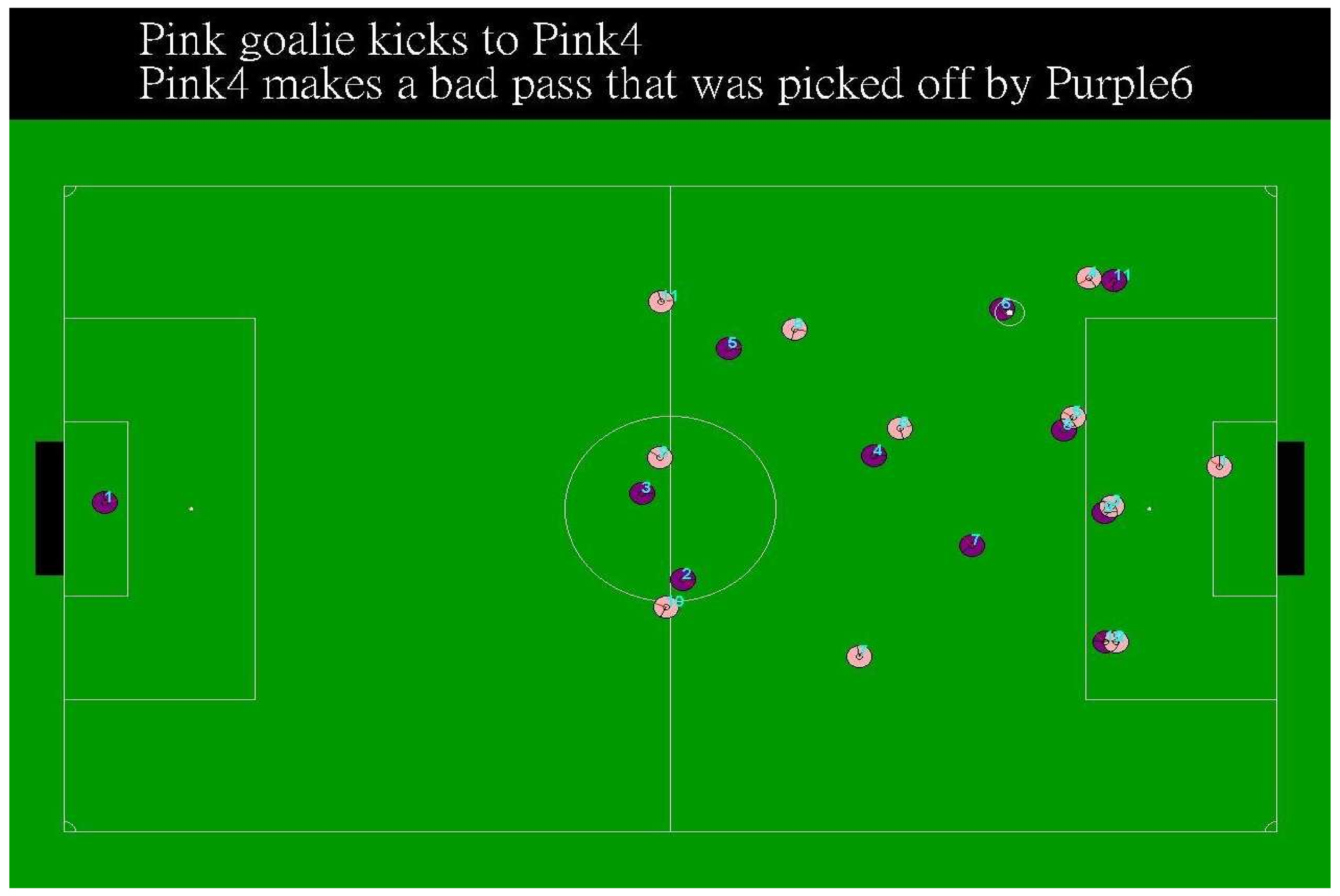
Learning to Sportscast: A Test of Grounded Language Acquisition

David L. Chen and Raymond J. Mooney
The University of Texas at Austin
{dlcc,mooney}@cs.utexas.edu

Learning to Sportscast



Motivations

- Constructing annotated corpora for training semantic parsers is difficult
- Children acquire language through exposure to linguistic input in the context of a rich, relevant, perceptual environment

Goals

- Learn to ground semantics of language [3, 5]
- Learn language through correlated linguistic and visual input

Tasks

- Learn to sportscast by observing sample human sportscasts
- Build a function to map between natural language (NL) and meaning representations (MRs)

Sportscasting Data

- A rule-based system is used to extract symbolic representations of game events from the simulation game states. These events constitute the MRs
- Human commentaries are recorded from a text box with timestamps
- Each comment is paired with all the events that occurred 5 second or less before the comment was made
- Collected data on four games with an average of 2613 MRs and 509 NL sentences for each game

Challenges

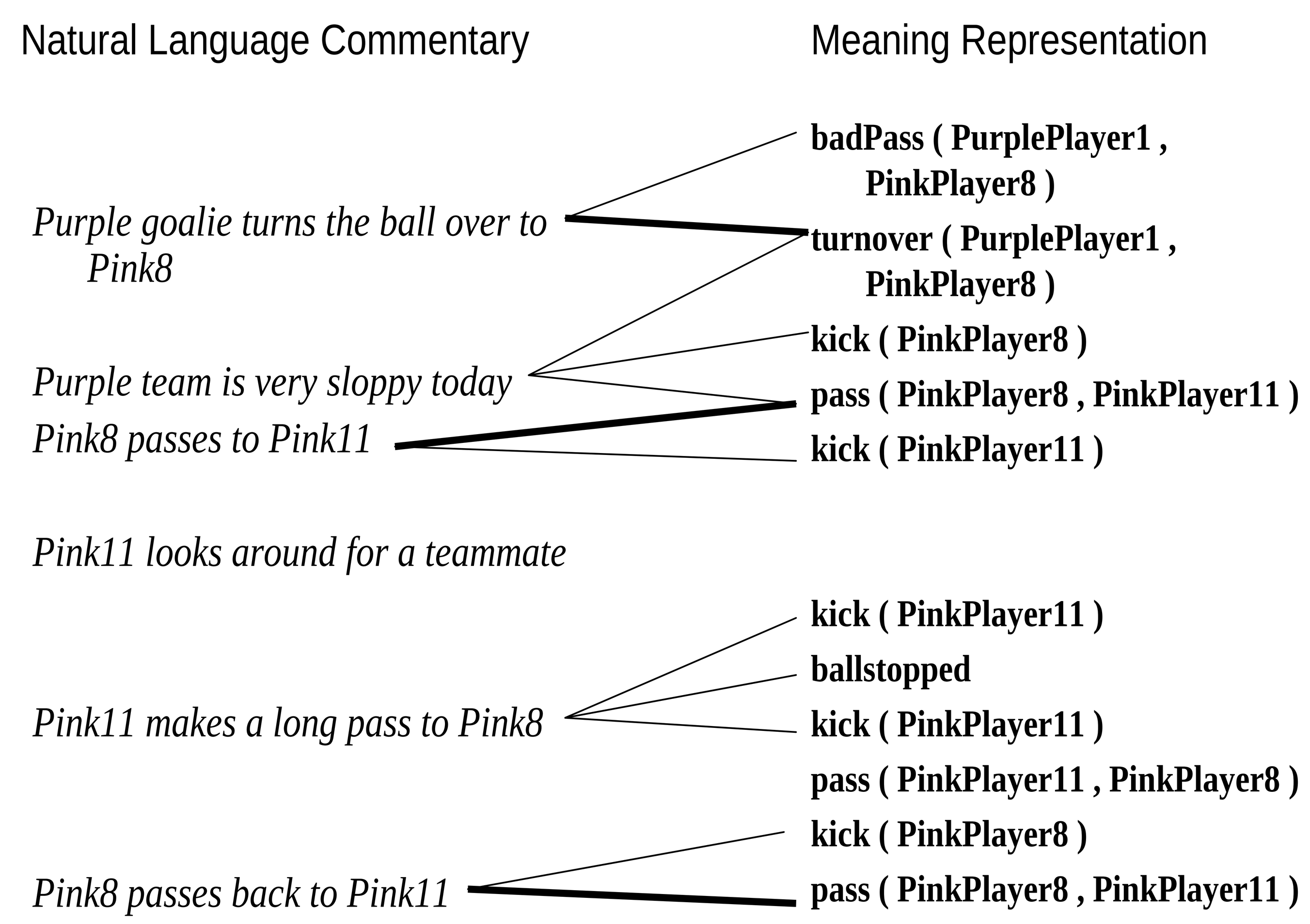
- The training data is highly ambiguous because each commentary is usually associated with several MRs
- Some NL sentences do not correspond to any MRs

Semantic Parsers

- Semantic parsers map NL sentences to MRs
- We experiment with two semantic parser learners
  - KRISP: Uses SVMs with string kernels [1, 2]
  - WASP: Uses synchronous context-free grammar (SCFG) [4]

Sample Data Trace

- The lines indicate the MRs that are associated with each NL sentence
- Bold lines indicate correct NL/MR pairs
- Not all sentences have correct pairings



Tactical Generation

Learning how to generate a NL sentence from a MR

Algorithm Skeleton

1. Train a semantic parser using all possible NL/MR pairings
2. Use the learned semantic parser to evaluate the likelihood of each NL/MR pairing and select the most likely MR for each sentence
3. Re-train the semantic parser using the disambiguated training data and iterate until termination condition

Various Systems

1. KRISPER: Uses KRISP to learn the semantic parser [2]
2. WASPER: Uses WASP to learn the semantic parser
3. KRISPER-WASP: Similar to KRISPER but trains using WASP with the final disambiguated data
4. WASPER-GEN: Uses WASP's language generator to evaluate the likelihood of a NL/MR pair instead of a semantic parser

Strategic Generation

Learning which MRs to talk about

Simple Algorithm

- Treat as a classification problem using only event type as feature
- Estimate how often an event type is commented on by using the disambiguated data from the tactical generation step

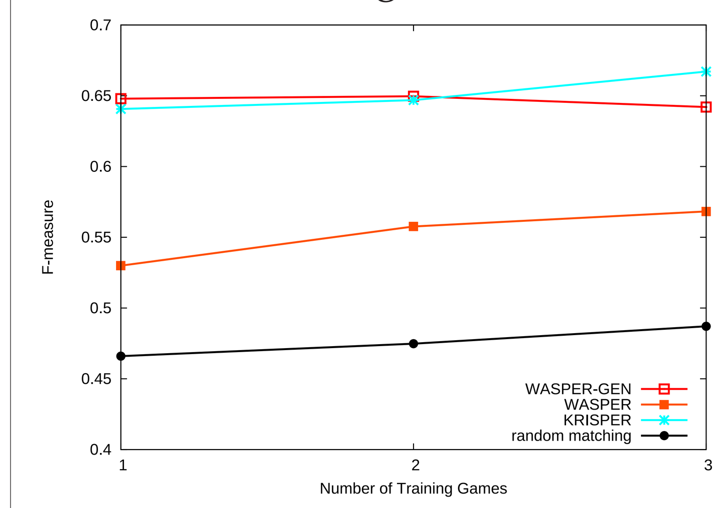
Iterative Generation Strategy Learning (IGSL)

- Directly estimates the likelihood of an event being commented on without learning a semantic parser
- Uses events not associated with any commentaries as negative evidence

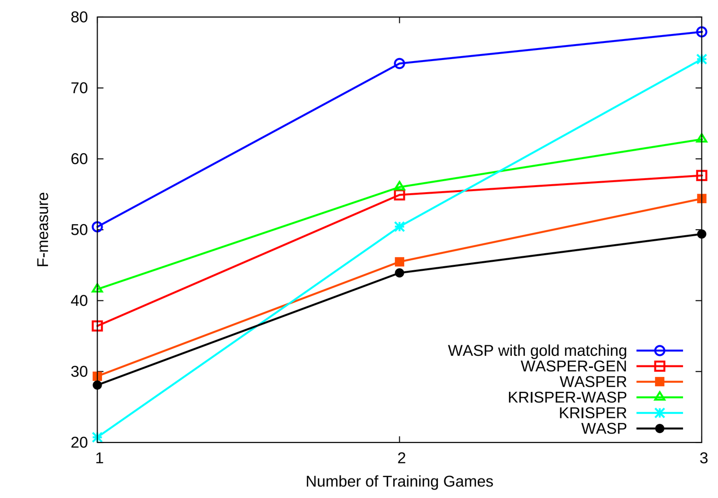
Experimental Evaluation

- WASP and WASP with gold matching are the lower and upper baselines
- Each system is evaluated on four tasks when applicable
  1. Matching: Ability to find correct NL/MR pairs
  2. Parsing: Translate from NL to MR
  3. Generation: Translate from MRto NL
  4. Strategic: Predict which MRs are described

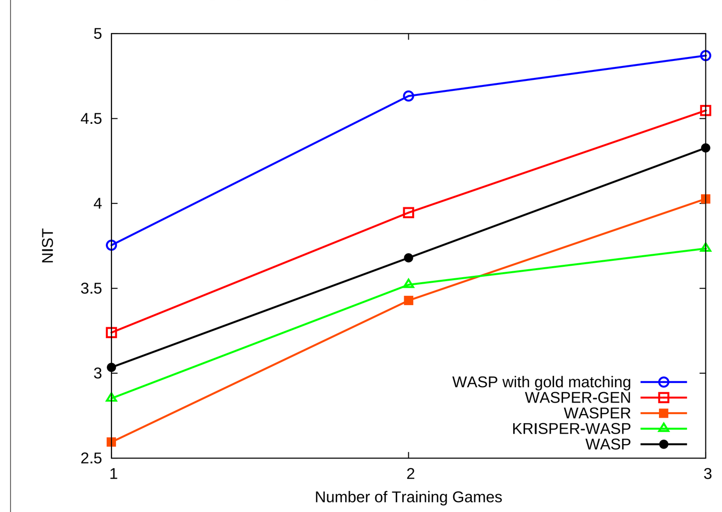
Matching Results



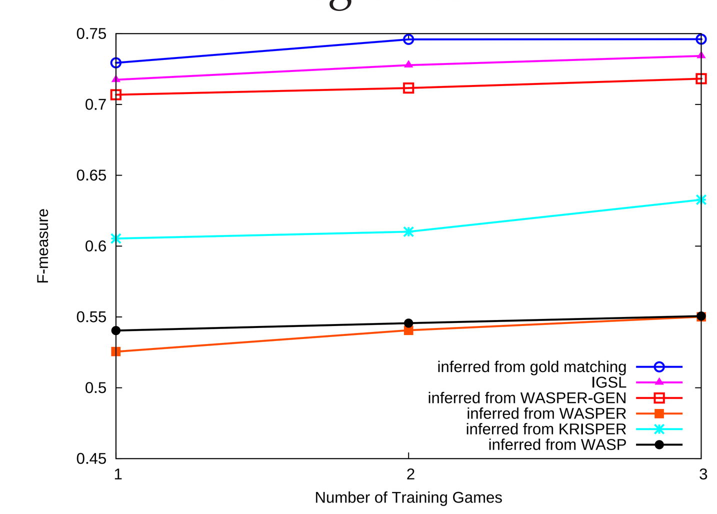
Parsing Results



Generation Results



Strategic Results



Human Evaluation

Setup

- Four fluent English speakers as judges
- Eight commented game clips were evaluated by each judge, half of the clips were commented by a human, the other half by our system
- Judges were not told which game clips were commented by human
- Each category is scored from 1 to 5, with 5 being the best

Results

	English Fluency	*Semantic Correctness*	*Sportscasting Ability*
Human	3.938	4.25	3.625
Machine	3.438	3.563	2.938

References

[1] R. J. Kate and R. J. Mooney. Using string-kernels for learning semantic parsers. In *ACL-06*, pages 913–920, Sydney, Australia, July 2006.

[2] Rohit J. Kate and Raymond J. Mooney. Learning language semantics from ambiguous supervision. In *AAAI-2007*, pages 895–900, 2007.

[3] Deb Roy. Learning visually grounded words and syntax for a scene description task. *Computer Speech and Language*, 16(3):353–385, 2002.

[4] Y. W. Wong and R. J. Mooney. Learning synchronous grammars for semantic parsing with lambda calculus. In *ACL-07*, pages 960–967, 2007.

[5] Chen Yu and Dana H. Ballard. On the integration of grounding language and learning objects. In *AAAI-2004*, pages 488–493, 2004.