# Using Perceptual Context to Ground Language


David Chen

Joint work with Joohyun Kim, Raymond Mooney

Department of Computer Sciences,

University of Texas at Austin



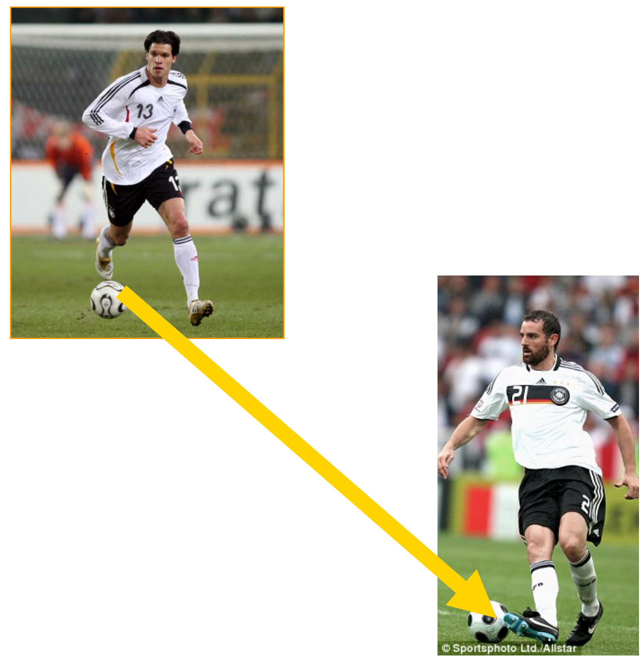
2009 IBM Statistical Machine Learning and Its Application(SMILe) Open House

October 8th and 9th, 2009

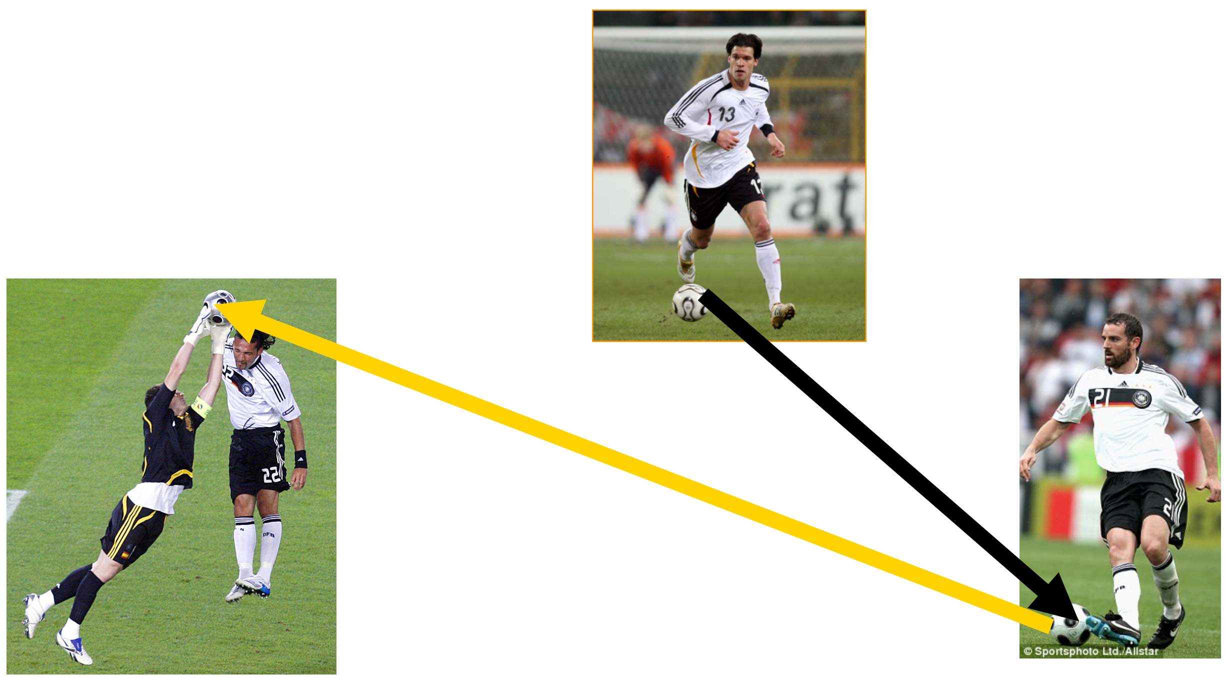
# Learning Language from Perceptual Context

- Children do not learn language from annotated corpora.
- The natural way to learn language is to perceive language in the context of its use in the physical and social world.

# Challenge

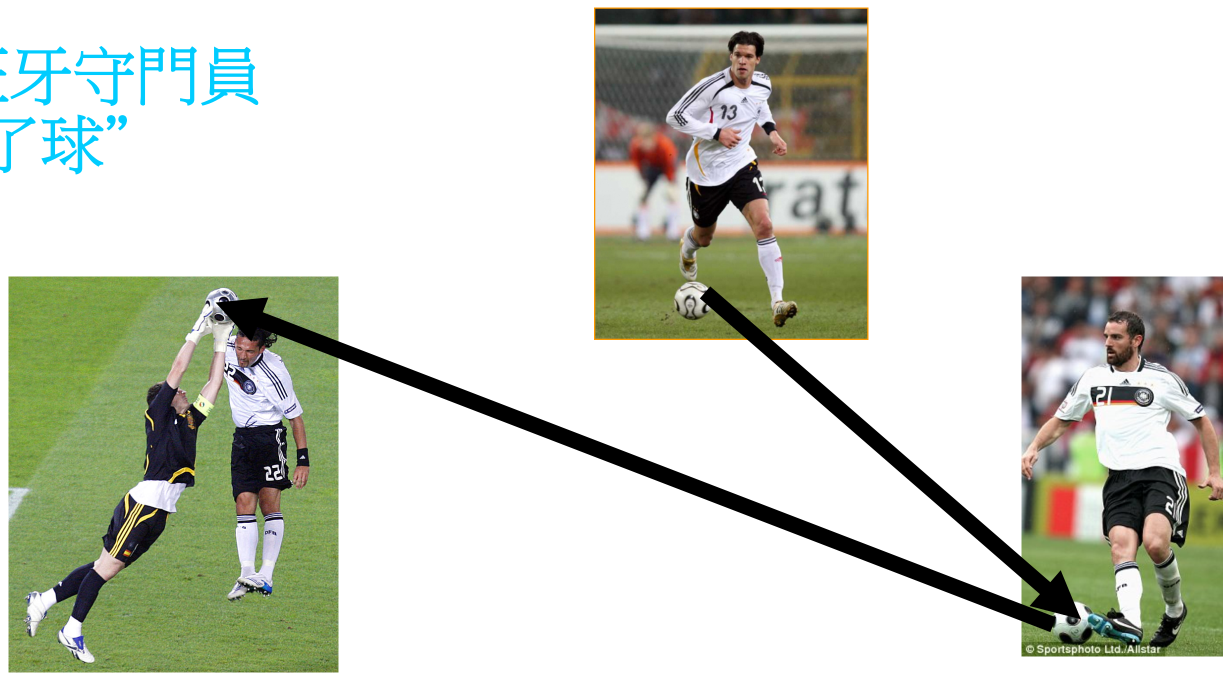


# Challenge



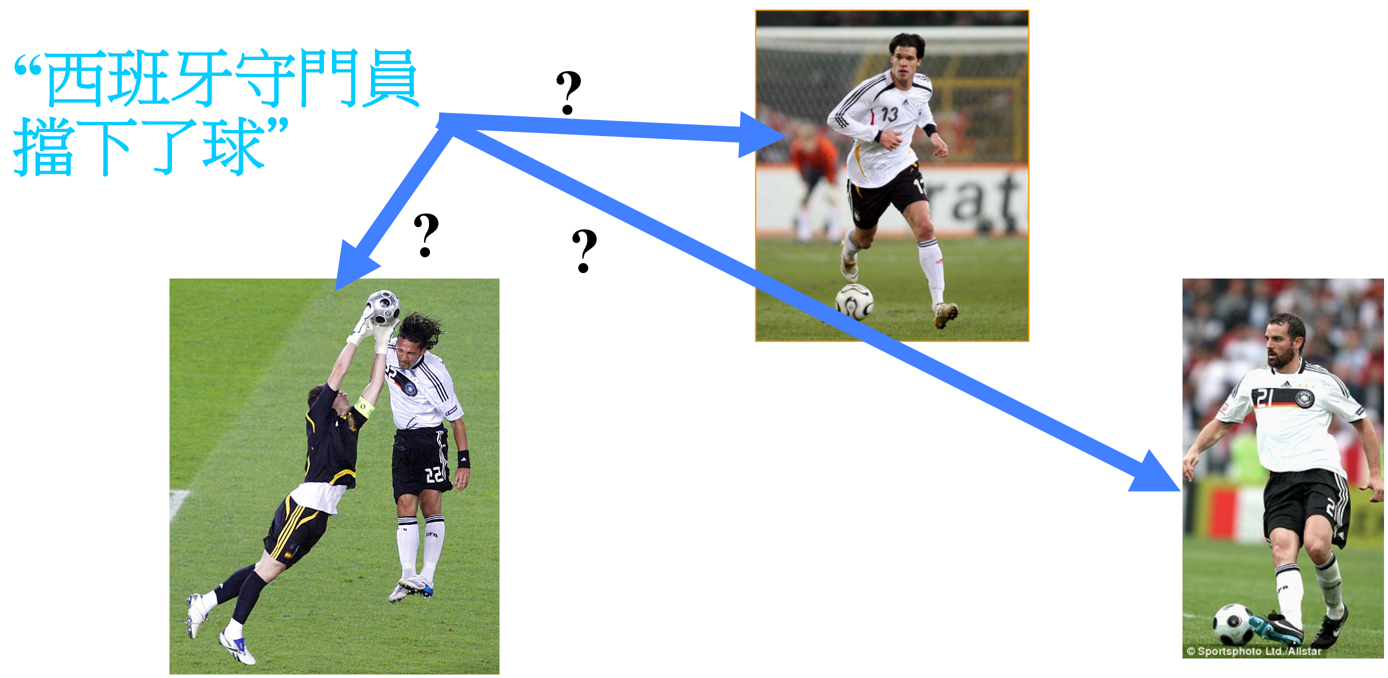
# Challenge

“西班牙守門員
擋下了球”



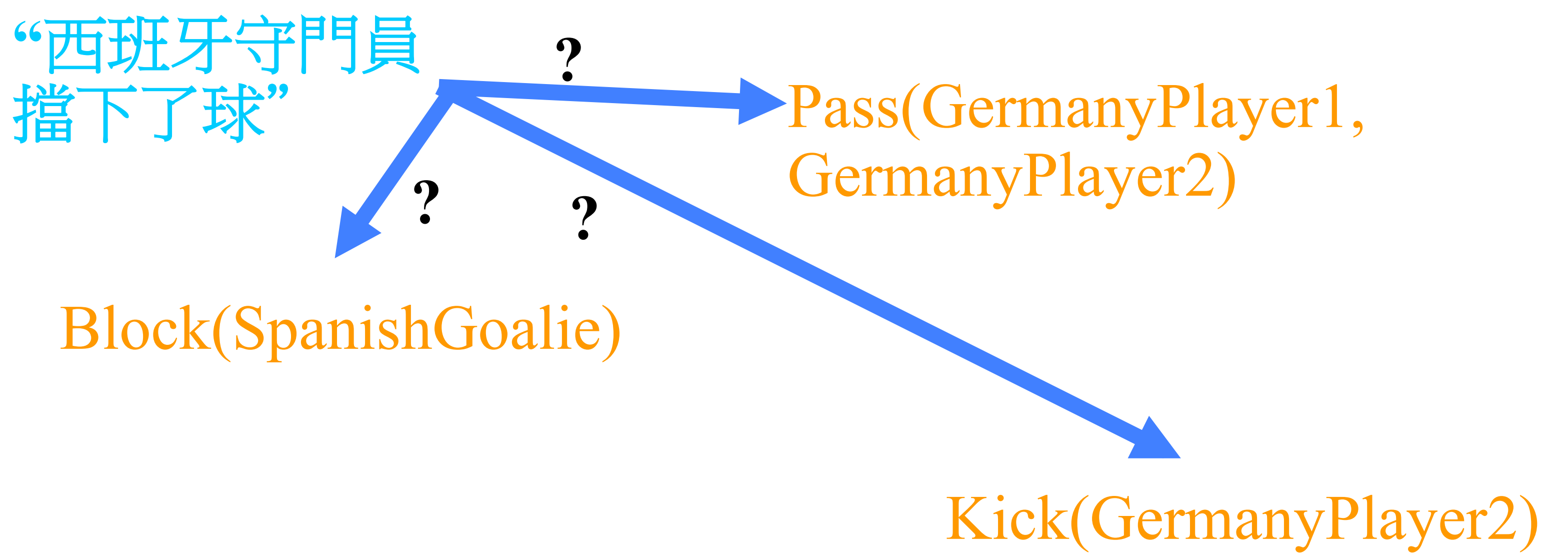
# Challenge

A linguistic input may correspond to many possible events



# Challenge

A linguistic input may correspond to many possible events



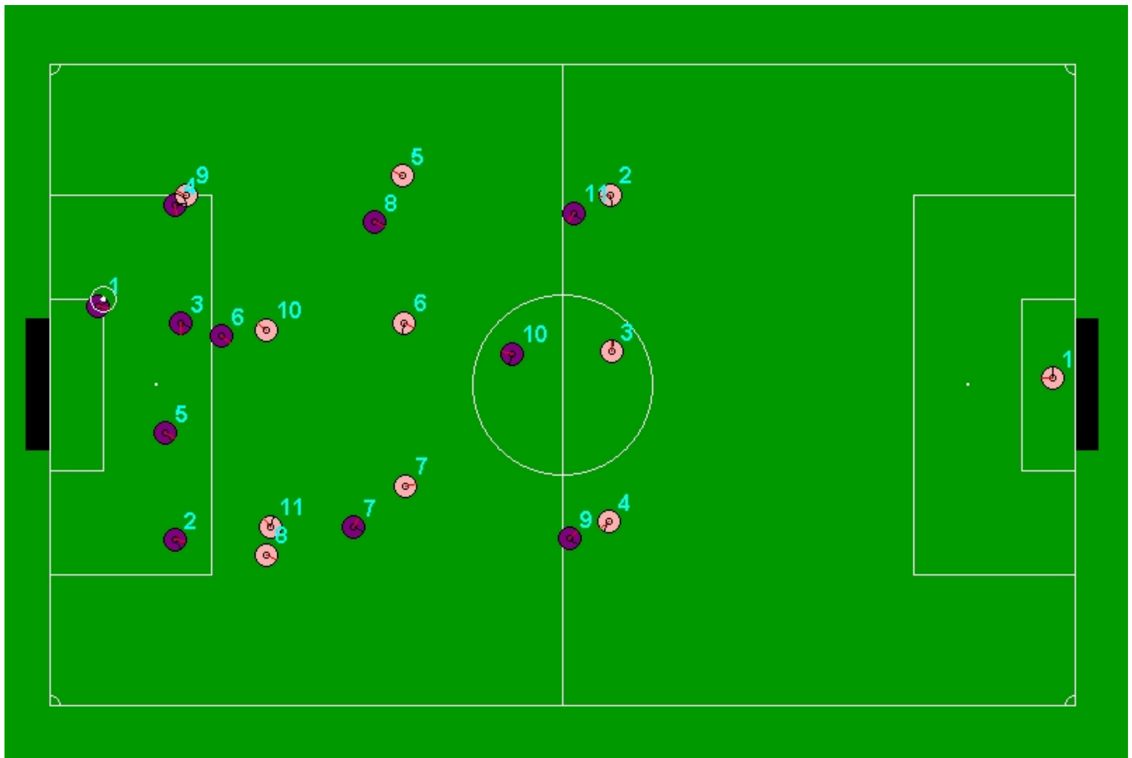
# Overview

- **Sportscasting task**
- Tactical generation
- Human evaluation

# Tractable Challenge Problem: Learning to Be a Sportscaster

- **Goal**: Learn from realistic data of natural language used in a representative context while avoiding difficult issues in computer perception (i.e. speech and vision).
- **Solution**: Learn from textually annotated traces of activity in a simulated environment.
- **Example**: Traces of games in the Robocup simulator paired with textual sportscaster commentary.

# Robocup Simulation League



Purple goalie blocked the ball

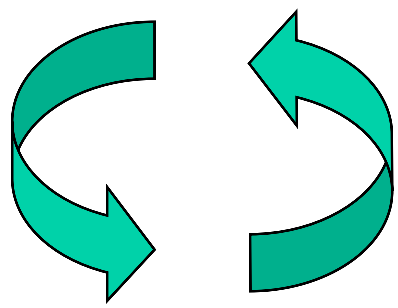
# Learning to Sportscast

- Learn to sportscast by observing sample human sportscasts
- Build a function that maps between natural language (NL) and meaning representation (MR)
  - NL: Textual commentaries about the game
  - MR: Predicate logic formulas that represent events in the game

# Mapping between NL/MR

NL: “Purple3 passes the ball to Purple5”

Semantic Parsing
(NL → MR)



Tactical Generation
(MR → NL)

MR: Pass ( Purple3, Purple5 )

# Robocup Sportscaster Trace

| Natural Language Commentary | Meaning Representation |
| --- | --- |
| | badPass ( Purple1, Pink8 ) |
| Purple goalie turns the ball over to Pink8 | turnover ( Purple1, Pink8 ) |
| | kick ( Pink8) |
| Purple team is very sloppy today | pass ( Pink8, Pink11 ) |
| Pink8 passes the ball to Pink11 | kick ( Pink11 ) |
| Pink11 looks around for a teammate | kick ( Pink11 ) |
| | ballstopped |
| Pink11 makes a long pass to Pink8 | kick ( Pink11 ) |
| | pass ( Pink11, Pink8 ) |
| | kick ( Pink8 ) |
| Pink8 passes back to Pink11 | pass ( Pink8, Pink11 ) |

# Robocup Sportscaster Trace

**Natural Language Commentary**

Purple goalie turns the ball over to Pink8

Purple team is very sloppy today

Pink8 passes the ball to Pink11

Pink11 looks around for a teammate

Pink11 makes a long pass to Pink8

Pink8 passes back to Pink11

**Meaning Representation**

badPass ( Purple1, Pink8 )

turnover ( Purple1, Pink8 )

kick ( Pink8)

pass ( Pink8, Pink11 )

kick ( Pink11 )

kick ( Pink11 )

ballstopped

kick ( Pink11 )

pass ( Pink11, Pink8 )

kick ( Pink8 )

pass ( Pink8, Pink11 )

# Robocup Sportscaster Trace

**Natural Language Commentary**

Purple goalie turns the ball over to Pink8

Purple team is very sloppy today

Pink8 passes the ball to Pink11

Pink11 looks around for a teammate

Pink11 makes a long pass to Pink8

Pink8 passes back to Pink11

**Meaning Representation**

badPass ( Purple1, Pink8 )

turnover ( Purple1, Pink8 )

kick ( Pink8)

pass ( Pink8, Pink11 )

kick ( Pink11 )

kick ( Pink11 )

ballstopped

kick ( Pink11 )

pass ( Pink11, Pink8 )

kick ( Pink8 )

pass ( Pink8, Pink11 )

# Robocup Sportscaster Trace

**Natural Language Commentary**

Purple goalie turns the ball over to Pink8

Purple team is very sloppy today

Pink8 passes the ball to Pink11

Pink11 looks around for a teammate

Pink11 makes a long pass to Pink8

Pink8 passes back to Pink11

**Meaning Representation**

P6 ( C1, C19 )

P5 ( C1, C19 )

P1( C19 )

P2 ( C19, C22 )

P1 ( C22 )

P1 ( C22 )

P0

P1 ( C22 )

P2 ( C22, C19 )

P1 ( C19 )

P2 ( C19, C22 )

# Robocup Data

- Collected human textual commentary for the 4 Robocup championship games from 2001-2004.
  - Avg # events/game = 2,613
  - Avg # English sentences/game = 509
  - Avg # Korean sentences/game = 499
- Each sentence matched to all events within previous 5 seconds.
  - Avg # MRs/sentence = 2.5 (min 1, max 12)
- Manually annotated with correct matchings of sentences to MRs (for evaluation purposes only).

# Overview

- Sportscasting task
- **Tactical generation**
- Human evaluation

# Tactical Generation

- Learn how to generate NL from MR
- Example:

  Pass(Pink2, Pink3) → “Pink2 kicks the ball to Pink3”

- Two steps
  1. Disambiguate the training data
  2. Learn a language generator

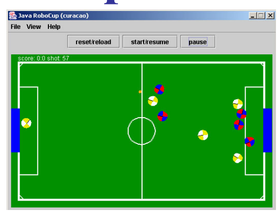
# WASP: **W**ord **A**lignment-based **S**emantic **P**arsing

- Uses statistical machine translation techniques
  - Synchronous context-free grammars (SCFG) [Wu, 1997; Melamed, 2004; Chiang, 2005]
  - Word alignments [Brown et al., 1993; Och & Ney, 2003]
- SCFG supports both:
  - Semantic Parsing: NL $\rightarrow$ MR
  - Tactical Generation: MR $\rightarrow$ NL

# WASPER: **WASP** with **EM**-like **R**etraining

**Sportscaster**

**Robocup Simulator**




Purple7 loses the ball to Pink2

Pink2 kicks the ball to Pink5

Pink5 makes a long pass to Pink8

Pink8 shoots the ball

Pass ( Purple5, Purple7 )
Turnover ( purple7 , pink2 )
Kick ( pink2 )
Pass ( pink2 , pink5 )
Kick ( pink5 )

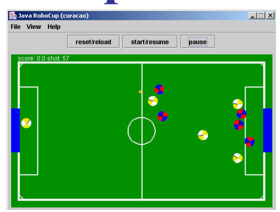
Pass ( pink5 , pink8)
Ballstopped
Kick ( pink8 )

Ambiguous Training Data

# WASPER: **WASP** with **EM**-like **R**etraining

Sportscaster

Robocup Simulator




Purple7 loses the ball to Pink2

Pink2 kicks the ball to Pink5

Pink5 makes a long pass to Pink8

Pink8 shoots the ball

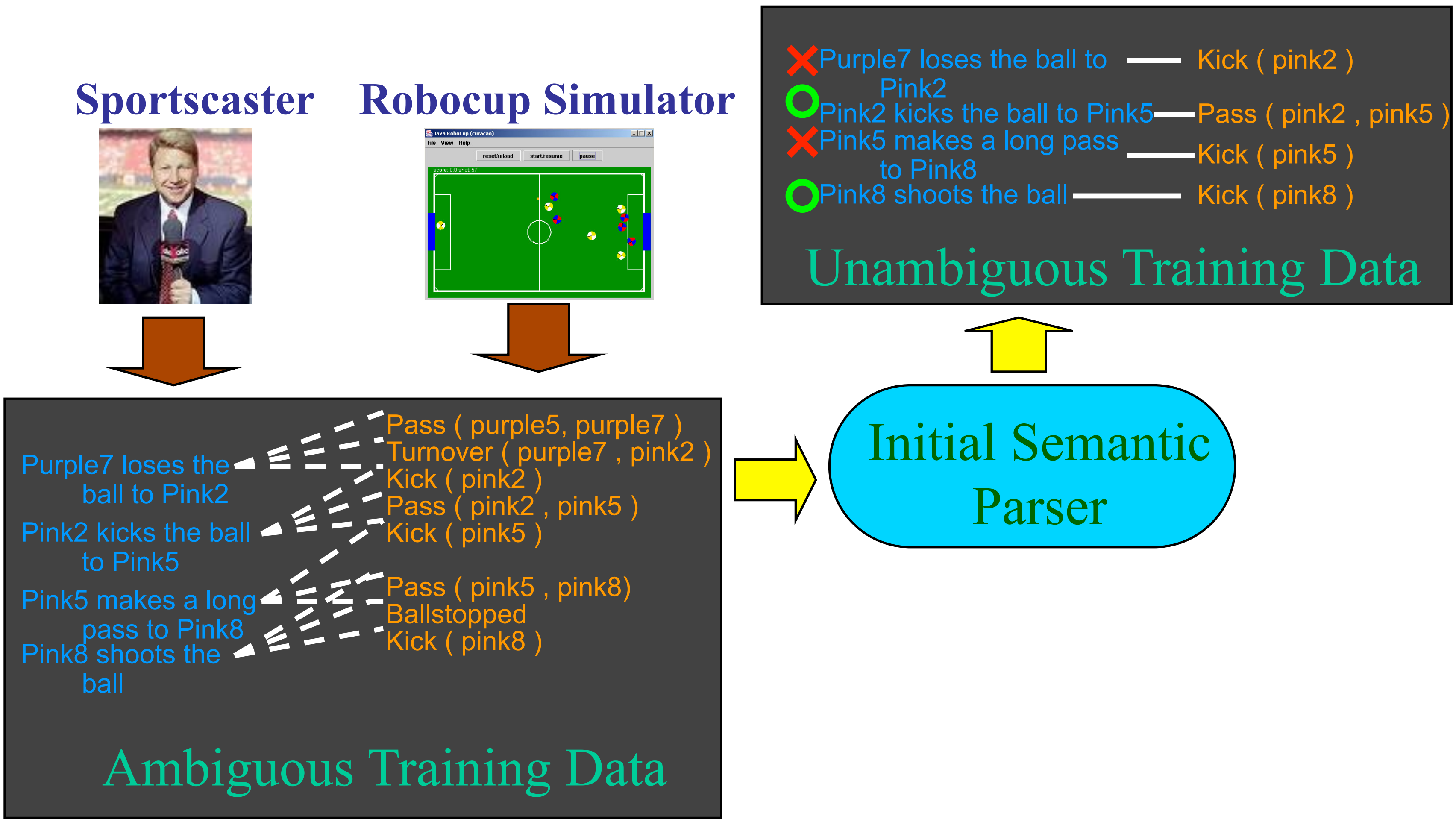
Pass ( Purple5, Purple7 )
Turnover ( purple7 , pink2 )
Kick ( pink2 )
Pass ( pink2 , pink5 )
Kick ( pink5 )
Pass ( pink5 , pink8)
Ballstopped
Kick ( pink8 )

Ambiguous Training Data

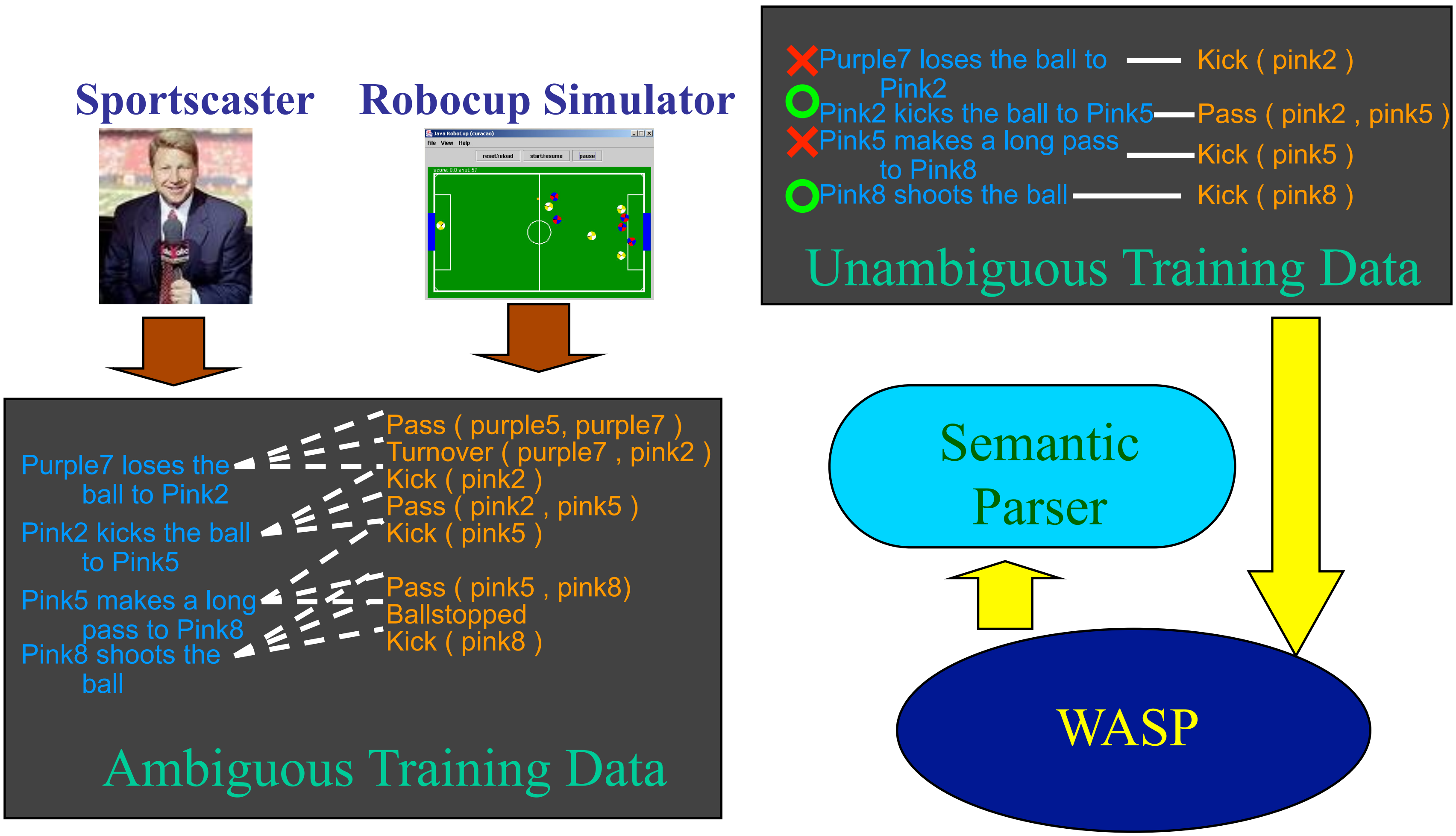
WASP

Initial Semantic Parser

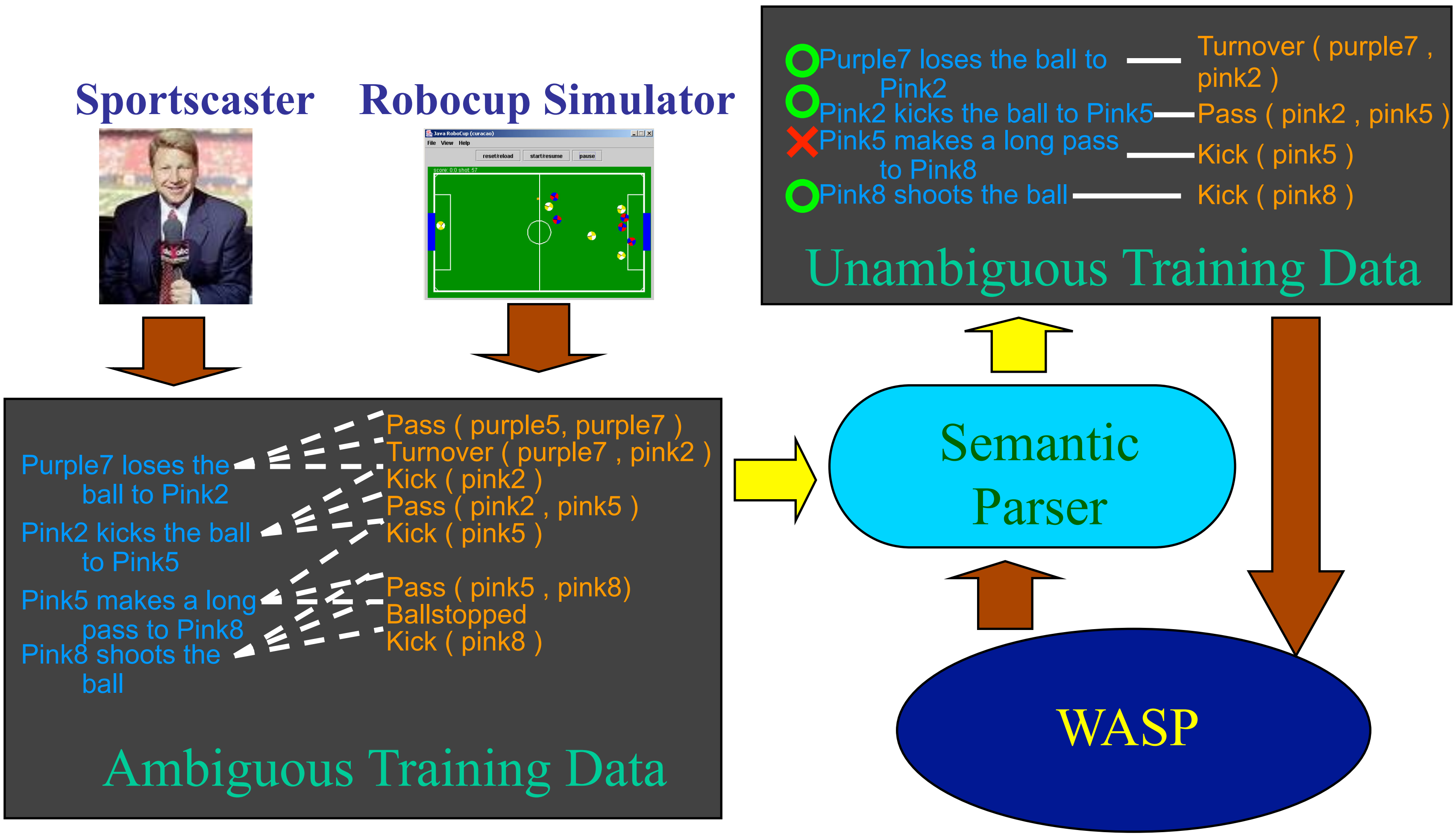
# WASPER: **WASP** with **EM**-like **R**etraining



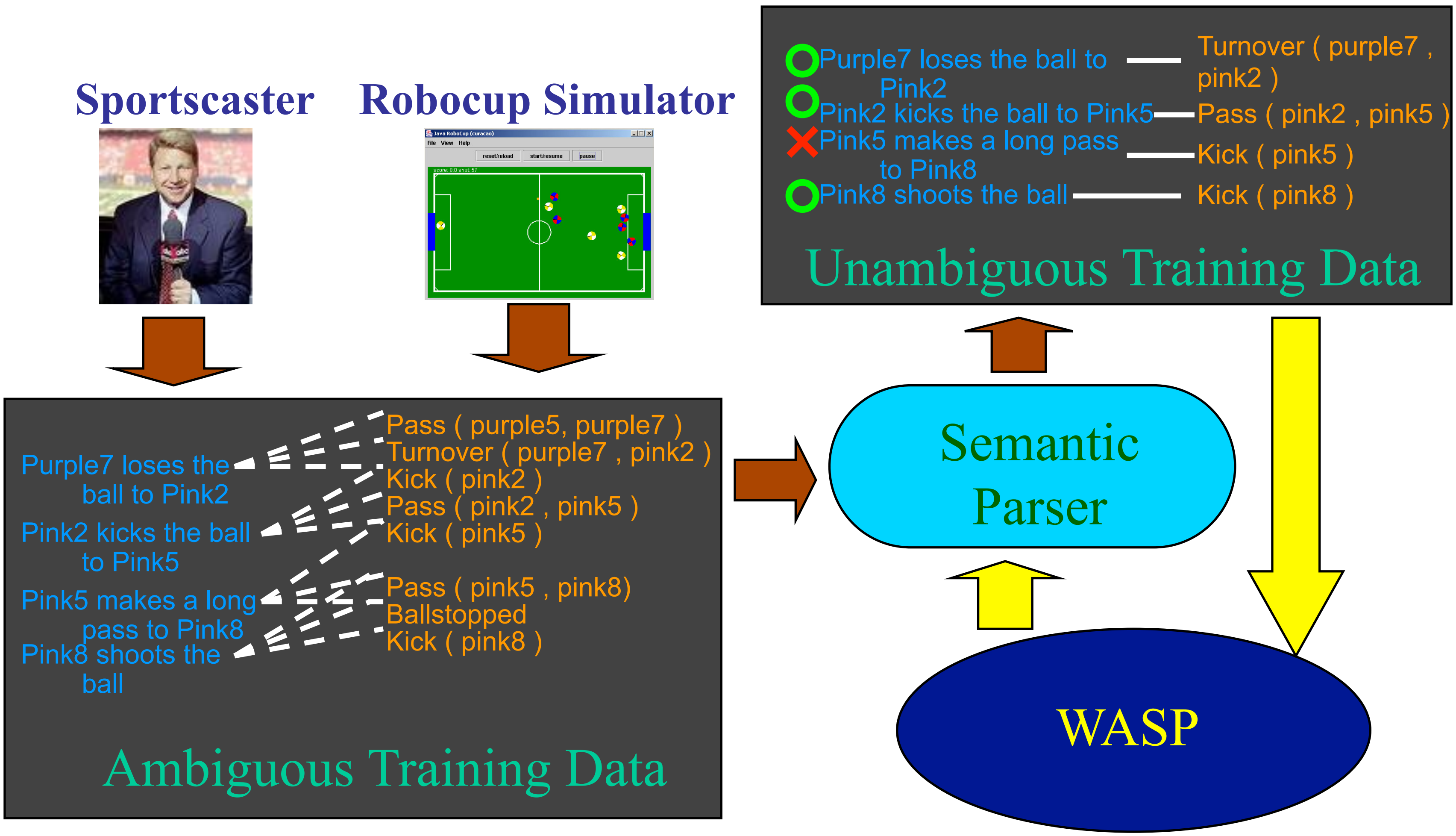
# WASPER: **WASP** with **EM**-like **R**etraining



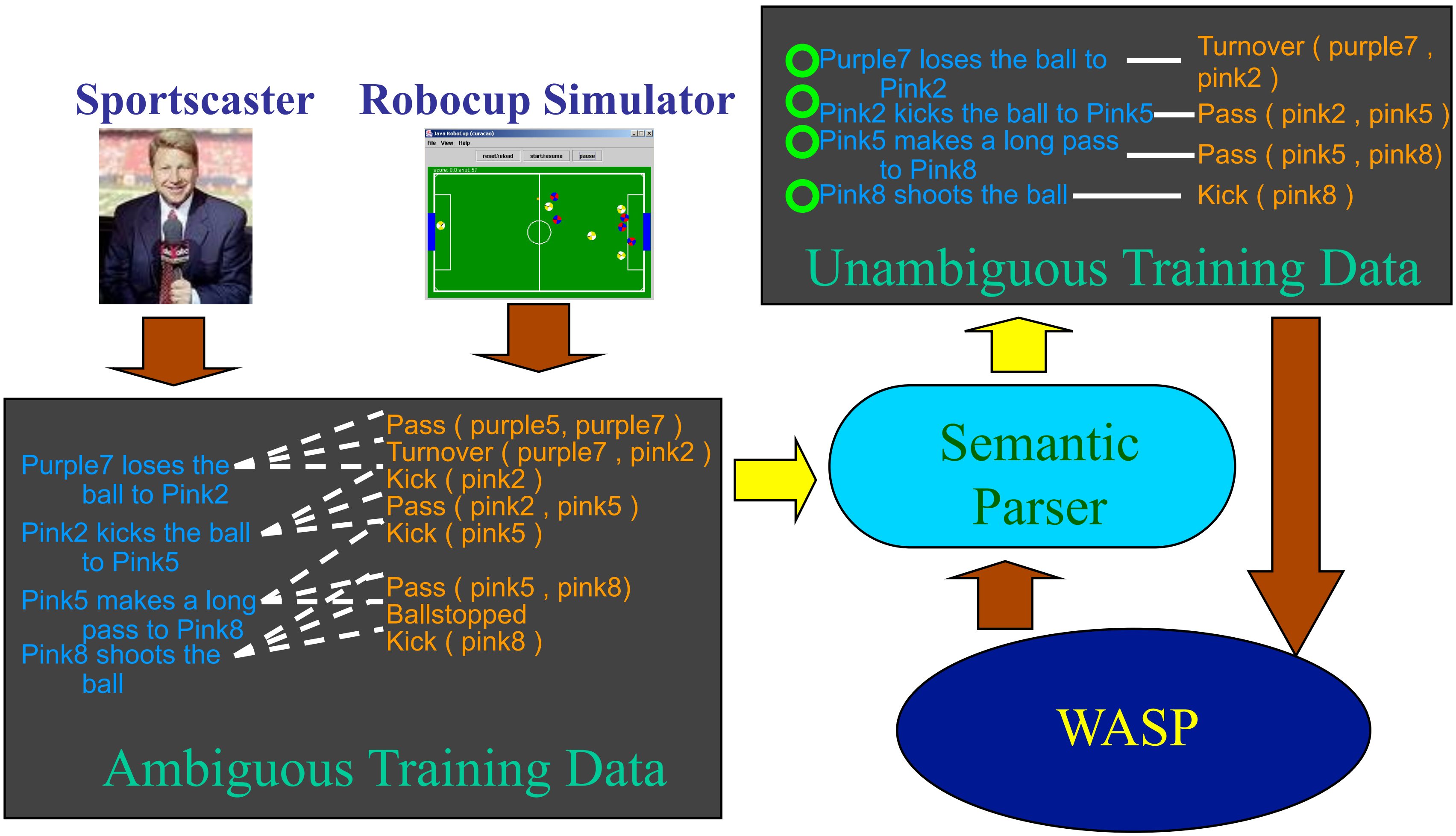
# WASPER: **WASP** with **EM**-like **R**etraining



# WASPER: **WASP** with **E**M-like **R**etraining



# WASPER: **WASP** with **EM**-like **R**etraining



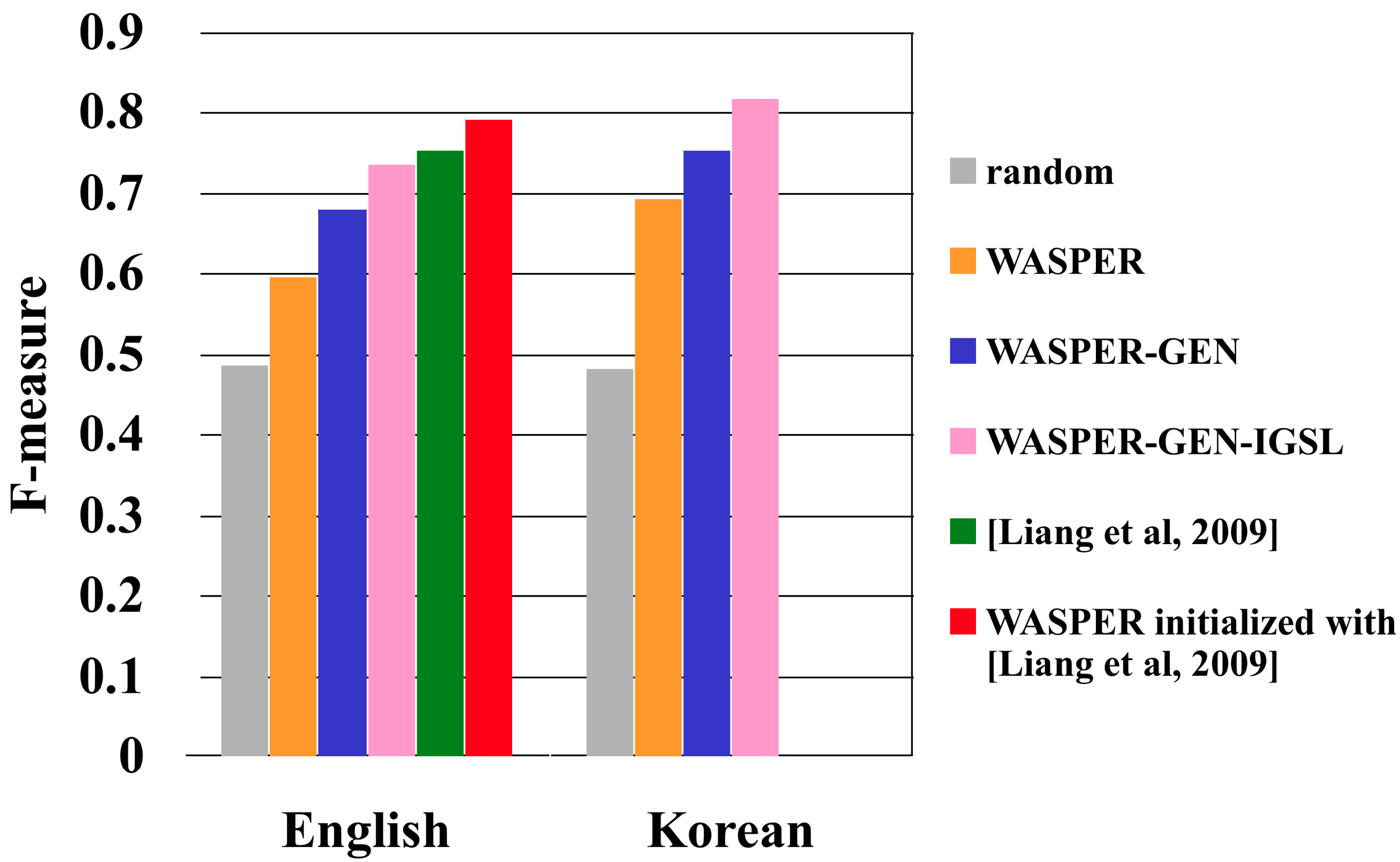
# Additional Systems

- WASPER-GEN
  - Uses tactical generator instead of semantic parser
- WASPER-GEN-IGSL
  - Same as WASPER-GEN except uses Iterative Generation Strategy Learning (IGSL) to initialize the first iteration
- WASP with random matching (lower baseline)
- WASP with gold matching (upper baseline)

# Matching

- 4 Robocup championship games from 2001-2004.
  - Avg # events/game = 2,613
  - Avg # English sentences/game = 509
  - Avg # Korean sentences/game = 499
- Leave-one-game-out cross-validation
- Metric:
  - **Precision**: % of system's annotations that are correct
  - **Recall**: % of gold-standard annotations produced
  - **F-measure**: Harmonic mean of precision and recall

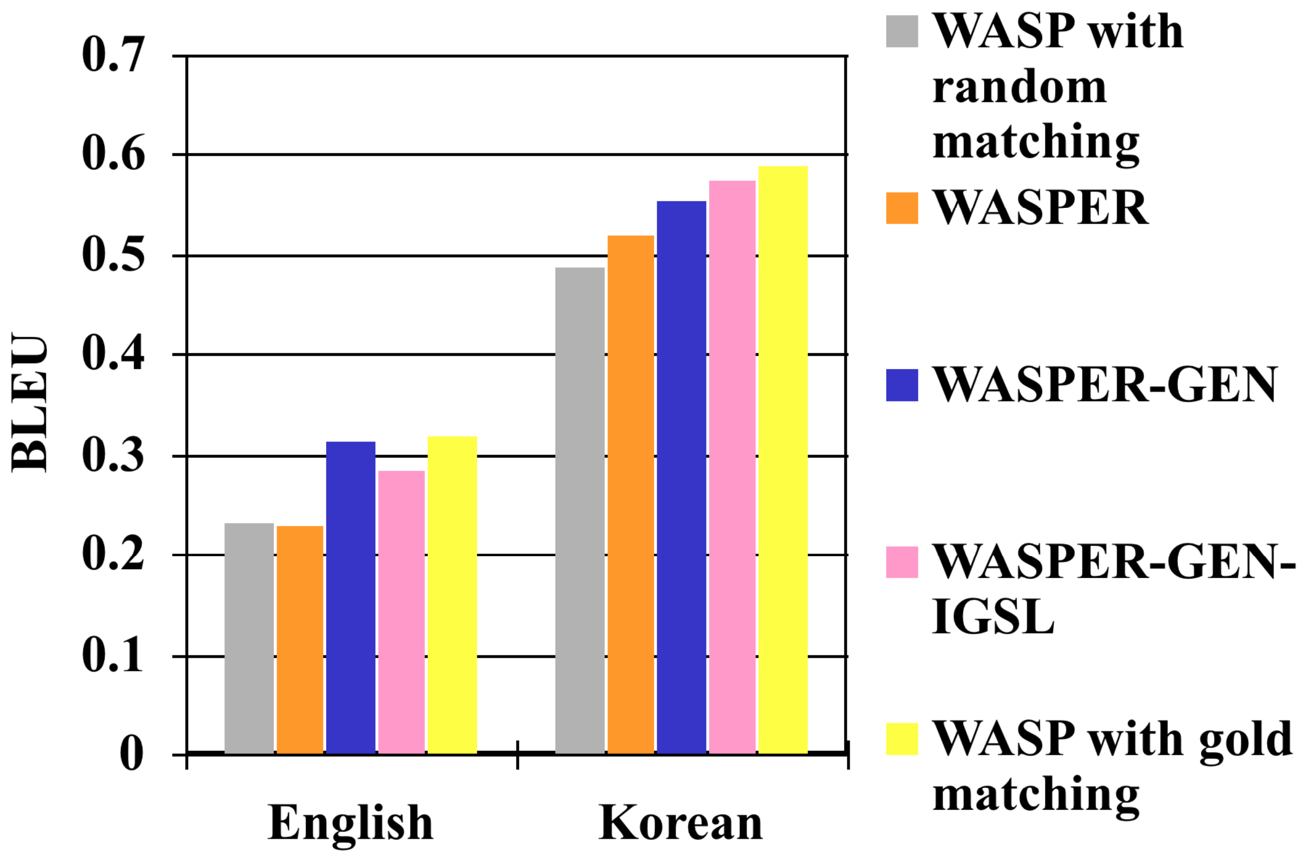
# Matching Results



# Tactical Generation

- Measure how accurately NL generator produces English sentences for chosen MRs in the test games.
- Use gold-standard matches to determine the correct sentence for each MR that has one.
- Leave-one-game-out cross-validation
- Metric:
  - **BLEU score:** [Papineni et al, 2002], N=4

# Tactical Generation Results



# Overview

- Sportscasting task
- Tactical generation
- **Human evaluation**

# Human Evaluation

- Used Amazon's Mechanical Turk to recruit human judges (~40 judges per video)
- 8 commented game clips
  - 4 minute clips randomly selected from each of the 4 games
  - Each clip commented once by a human, and once by the machine
- Presented in random counter-balanced order
- Judges were not told which ones were human or machine generated

# Human Evaluation

| Score | English Fluency | Semantic Correctness | Sportscasting Ability |
|---|---|---|---|
| 5 | Flawless | Always | Excellent |
| 4 | Good | Usually | Good |
| 3 | Non-native | Sometimes | Average |
| 2 | Disfluent | Rarely | Bad |
| 1 | Gibberish | Never | Terrible |

# Human Evaluation

| | Syntax | Semantic | Overall | Human? |
|---|---|---|---|---|
| 2001 Human | 3.735 | 3.588 | 3.147 | 0.206 |
| 2001 Machine | 3.888 | 3.806 | 3.611 | 0.4 |
| 2002 Human | 4.132 | 4.579 | 4.027 | 0.421 |
| 2002 Machine | 3.971 | 3.735 | 3.286 | 0.118 |
| 2003 Human | 3.541 | 3.730 | 2.611 | 0.135 |
| 2003 Machine | 3.893 | 4.263 | 3.368 | 0.193 |
| 2004 Human | 4.029 | 4.171 | 3.543 | 0.2 |
| 2004 Machine | 4.125 | 4.375 | 4.0 | 0.563 |

# Conclusion

- Current language learning work uses expensive, annotated training data.
- We have developed a language learning system that can learn from language paired with an ambiguous perceptual environment.
- We have evaluated it on the task of learning to sportscast simulated Robocup games.
- The system learns to sportscast as well as CS students who don’t watch soccer

# Demo Clip

- Game clip commentated using WASPER-GEN with IGSL, since this gave the best results for generation.
- FreeTTS was used to synthesize speech from textual output.
- YouTube link:

  http://www.youtube.com/watch?v=L_MIRS7NBpU

# Backup Slides

# Overview

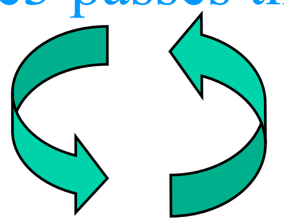
- Sportscasting task
- **Related works**
- Tactical generation
- Strategic generation
- Human evaluation

# Semantic Parser Learners

- Learn a function from NL to MR

NL: "Purple3 passes the ball to Purple5"

Semantic Parsing
(NL → MR)



Tactical Generation
(MR → NL)

MR: Pass ( Purple3, Purple5 )

- We experiment with two semantic parser learners
  - WASP (Wong & Mooney, 2006; 2007)
  - KRISP (Kate & Mooney, 2006)

# KRISP: **K**ernel-based **R**obust **I**nterpretation by **S**emantic **P**arsing

- Productions of MR language are treated like semantic concepts
- SVM classifier is trained for each production with string subsequence kernel
- These classifiers are used to compositionally build MRs of the sentences
- More resistant to noisy supervision but incapable of tactical generation

# Overview

- Sportscasting task
- Related works
- Tactical generation
- **Strategic generation**
- Human evaluation

# Strategic Generation

- Generation requires not only knowing ***how*** to say something (tactical generation) but also ***what*** to say (strategic generation).
- For automated sportscasting, one must be able to effectively choose which events to describe.

# Example of Strategic Generation

pass ( purple7 , purple6 )

ballstopped

kick ( purple6 )

pass ( purple6 , purple2 )

ballstopped

kick ( purple2 )

pass ( purple2 , purple3 )

kick ( purple3 )

badPass ( purple3 , pink9 )

turnover ( purple3 , pink9 )

# Example of Strategic Generation

**pass ( purple7 , purple6 )**

**ballstopped**

**kick ( purple6 )**

**pass ( purple6 , purple2 )**

**ballstopped**

**kick ( purple2 )**

**pass ( purple2 , purple3 )**

**kick ( purple3 )**

**badPass ( purple3 , pink9 )**

**turnover ( purple3 , pink9 )**

# Strategic Generation

- For each event type (e.g. pass, kick) estimate the probability that it is described by the sportscaster.
- Requires correct NL/MR matching
  - Use estimated matching from tactical generation
  - Iterative Generation Strategy Learning

# Iterative Generation Strategy Learning (IGSL)

- Directly estimates the likelihood of an event being commented on
- Self-training iterations to improve estimates
- Uses events not associated with any NL as negative evidence

# Robocup Sportscaster Trace

**Natural Language Commentary**

purple6 passes to purple2

purple3 loses the ball to pink9

purple2 makes a short pass to purple3

**Meaning Representation**

ballstopped

kick ( purple6 )

pass ( purple6 , purple2 )

turnover ( purple3 , pink9 )

kick ( purple2 )

pass ( purple2 , purple3 )

kick ( purple3 )

# Robocup Sportscaster Trace

| Natural Language Commentary | Meaning Representation |
|---|---|
| purple6 passes to purple2 | pass ( purple6 , purple2 ) |
| purple3 loses the ball to pink9 | |
| | pass ( purple2 , purple3 ) |
| purple2 makes a short pass to purple3 | |

# Robocup Sportscaster Trace

| Natural Language Commentary | Meaning Representation |
| --- | --- |
| | kick ( purple6 ) |
| purple6 passes to purple2 | |
| purple3 loses the ball to pink9 | |
| | kick ( purple2 ) |
| | kick ( purple3 ) |
| purple2 makes a short pass to purple3 | |

# Robocup Sportscaster Trace

**Natural Language Commentary**

purple6 passes to purple2

purple3 loses the ball to pink9

purple2 makes a short pass to purple3

**Meaning Representation**

kick ( purple 3 )

ballstopped

kick ( purple6 )

pass ( purple6 , purple2 )

kick ( purple2 )

turnover ( purple3 , pink9 )

kick ( purple2 )

pass ( purple2 , purple3 )

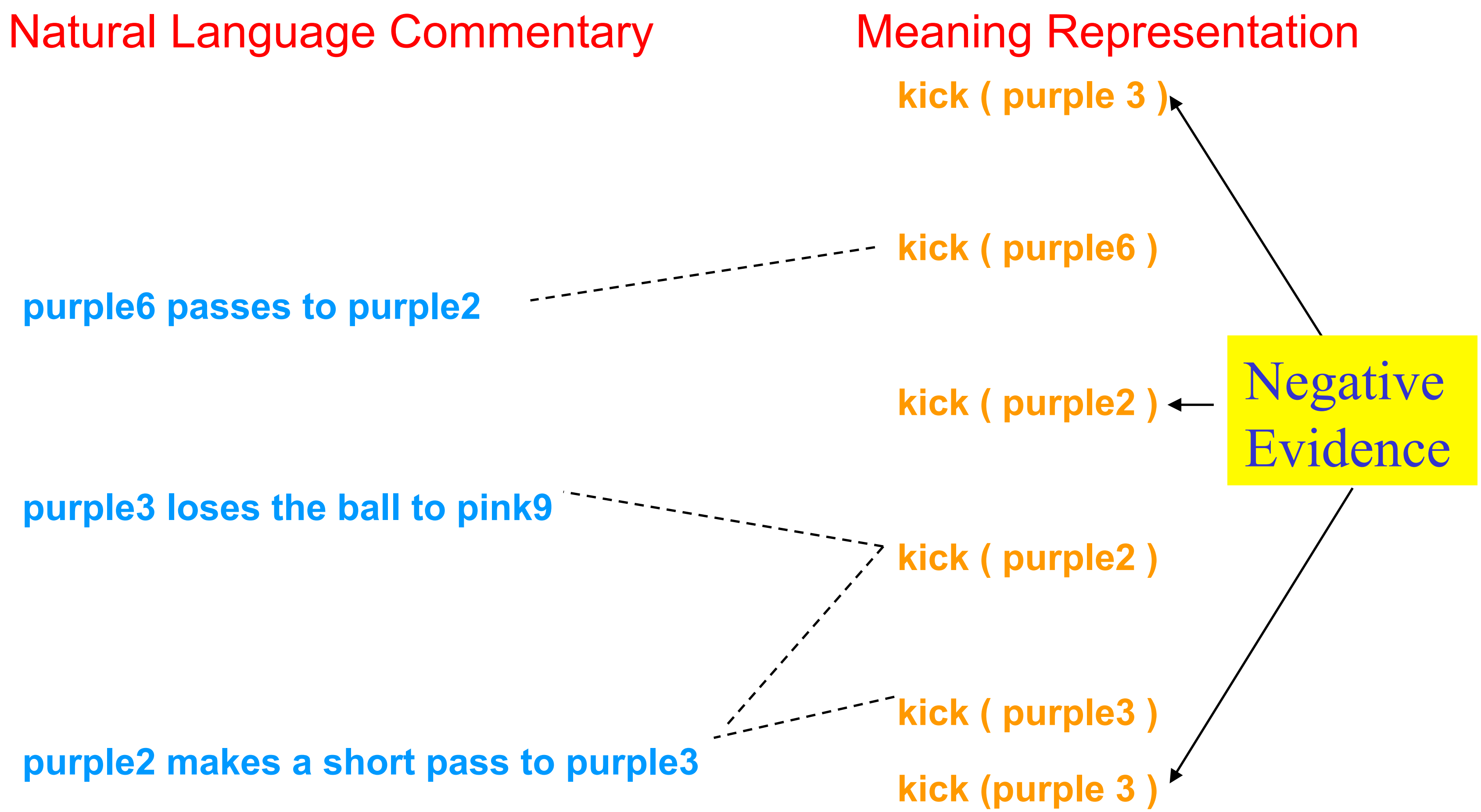
kick ( purple3 )

kick (purple 3

# Robocup Sportscaster Trace

| Natural Language Commentary | Meaning Representation |
| --- | --- |
| | kick ( purple 3 ) |
| | kick ( purple6 ) |
| purple6 passes to purple2 | |
| | kick ( purple2 ) |
| purple3 loses the ball to pink9 | |
| | kick ( purple2 ) |
| | kick ( purple3 ) |
| purple2 makes a short pass to purple3 | |
| | kick (purple 3 |

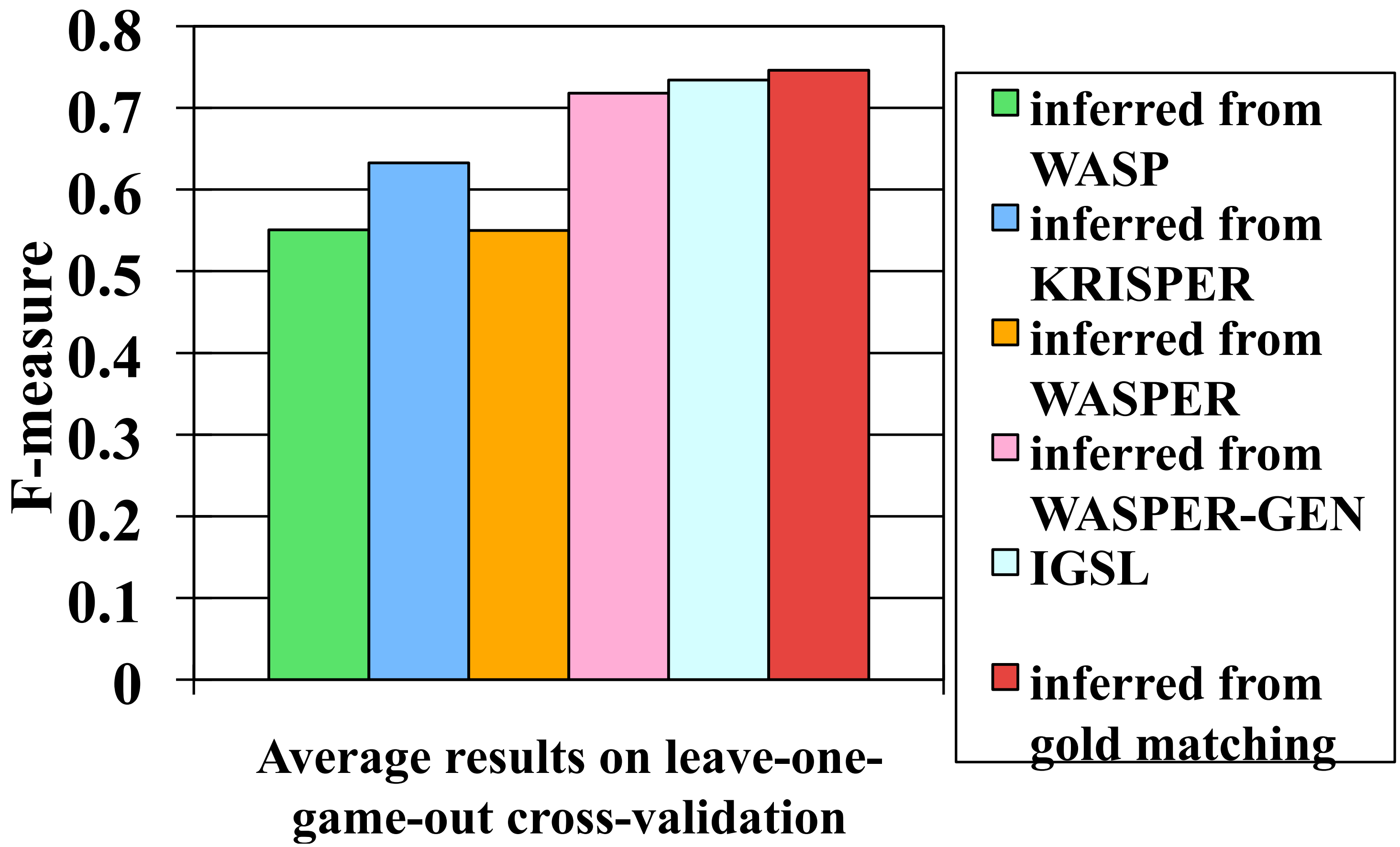
# Robocup Sportscaster Trace



# Strategic Generation Performance

- Evaluate how well the system can predict which events a human comments on
- Metric:
  - **Precision**: % of system's annotations that are correct
  - **Recall**: % of gold-standard annotations correctly produced
  - **F-measure**: Harmonic mean of precision and recall

# Strategic Generation Results



# Demo Clip

- Game clip commentated using WASPER-GEN with IGSL, since this gave the best results for generation.
- FreeTTS was used to synthesize speech from textual output.
- English: http://www.youtube.com/watch?v=L_MIRS7NBpU
- Korean: http://www.youtube.com/watch?v=Dur9K5AiK8Y