Learning to Sportscast: A Test of Grounded Language Acquisition

David Chen & Raymond Mooney

Department of Computer Sciences

University of Texas at Austin

Motivation

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

Goals

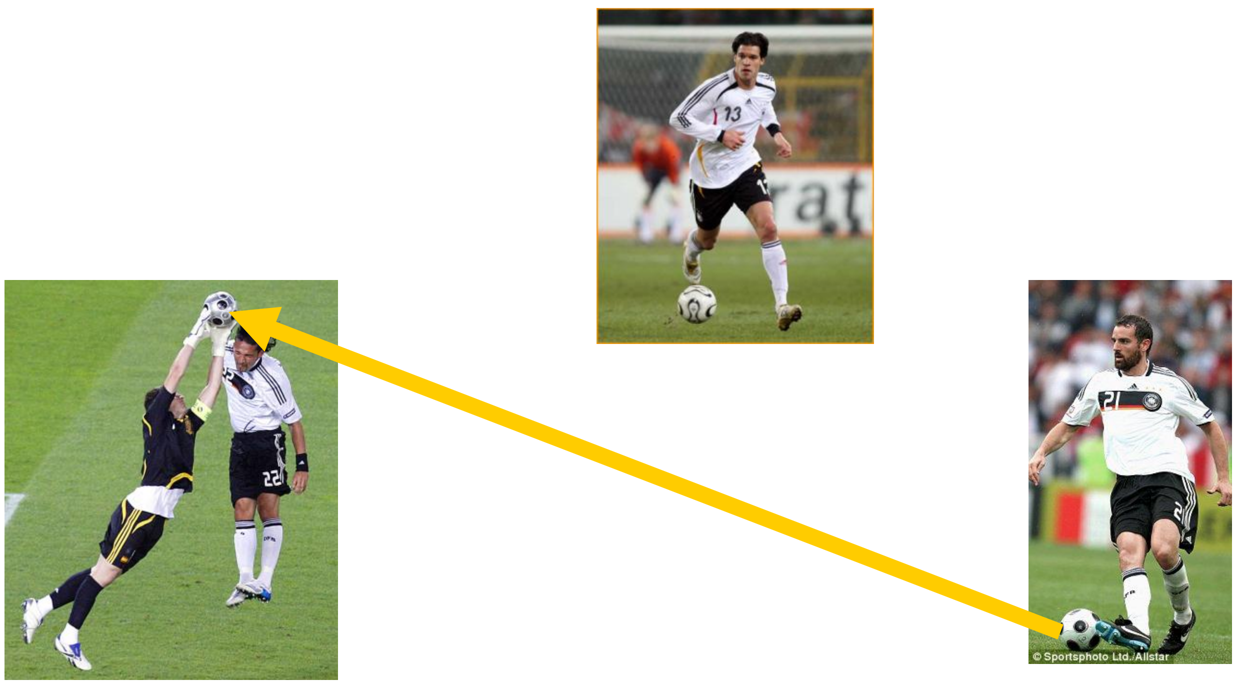
- Learn to ground the semantics of language

Block

- Learn language through correlated linguistic and visual inputs

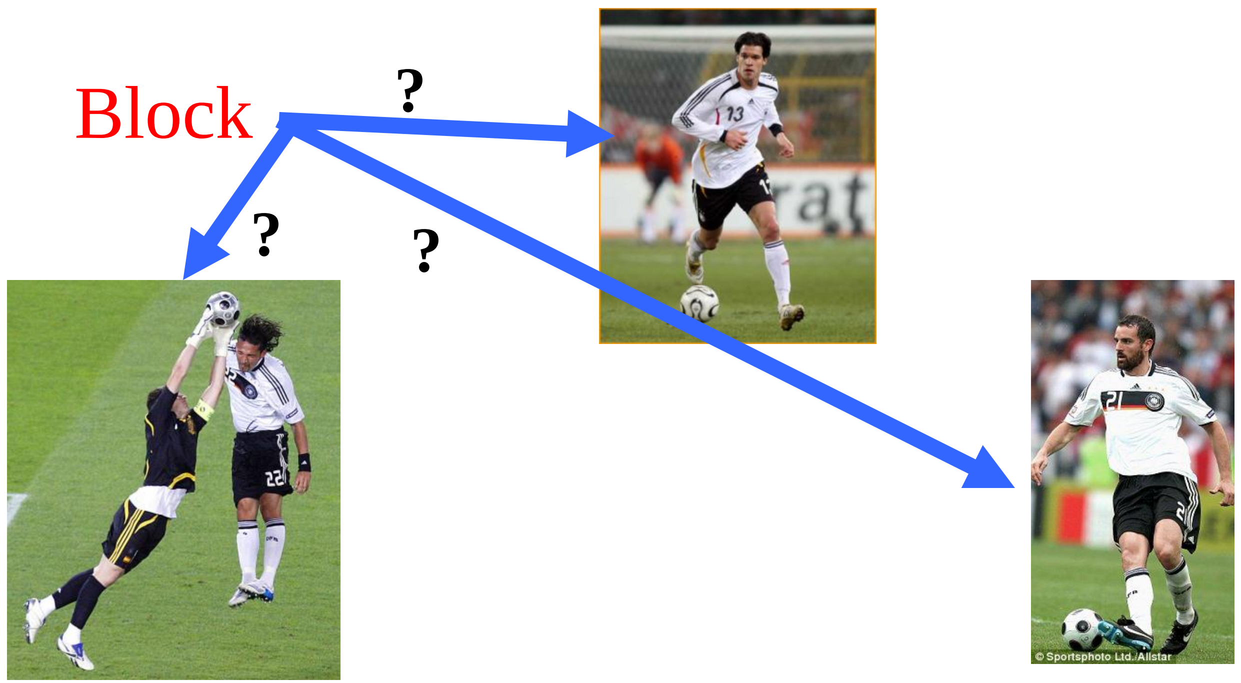
Challenge

Challenge



Challenge

A linguistic input may correspond to many possible events



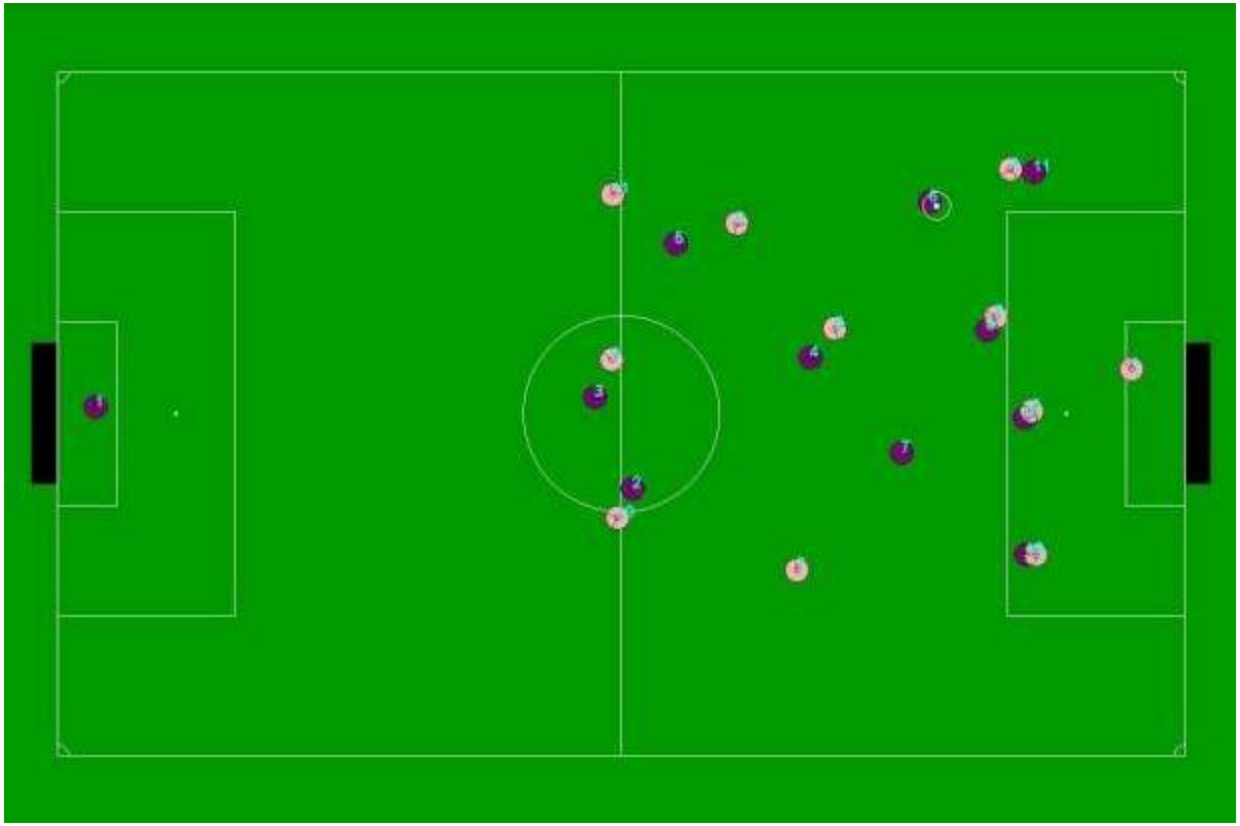
Overview

- Sportscasting task
- Tactical generation
- Strategic generation
- Human evaluation

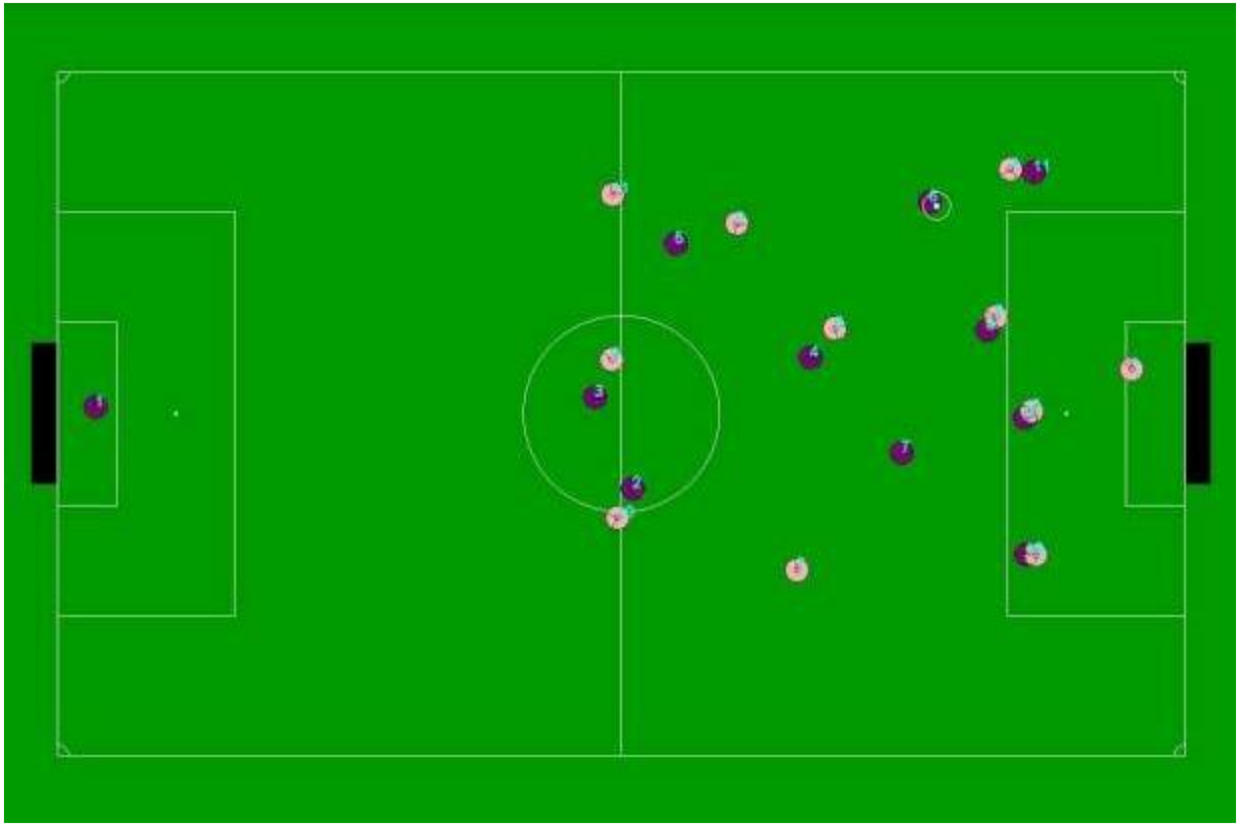
Learning to Sportscast

- Robocup Simulation League games
- No speech recognition
  - Record commentaries in text form
- No computer vision
  - Ruled-based system to automatically extract game events in symbolic form
- Concentrate on linguistic issues

Robocup Simulation League



Robocup Simulation League



Pink4's pass was intercepted by Purple6

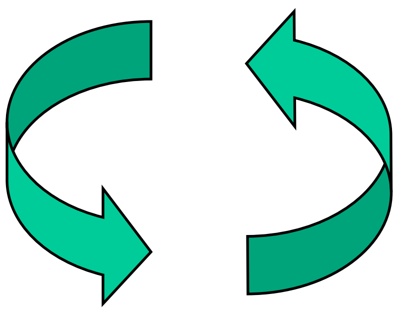
Learning to Sportscast

- 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)
	pass (Pink8, Pink11)
Purple team is very sloppy today	kick (Pink11)
Pink8 passes the ball to Pink11	
Pink11 looks around for a teammate	
	kick (Pink11)
	ballstopped
Pink11 makes a long pass to Pink8	kick (Pink11)
	pass (Pink11, Pink8)
	kick (Pink8)
	pass (Pink8, Pink11)
Pink8 passes back to 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

Meaning Representation

P6 (C1, C19)

Purple goalie turns the ball over to Pink8

P5 (C1, C19)

P1(C19)

Purple team is very sloppy today

P2 (C19, C22)

Pink8 passes the ball to Pink11

P1 (C22)

Pink11 looks around for a teammate

P1 (C22)

P0

Pink11 makes a long pass to Pink8

P1 (C22)

P2 (C22, C19)

P1 (C19)

Pink8 passes back to Pink11

P2 (C19, C22)

Robocup Data

- Collected human textual commentary for the 4 Robocup championship games from 2001-2004.
  - Avg # events/game = 2,613
  - Avg # sentences/game = 509
- 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**
- Strategic 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

System Overview

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

System Overview

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

Semantic Parser Learner

Initial Semantic Parser

System Overview

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)

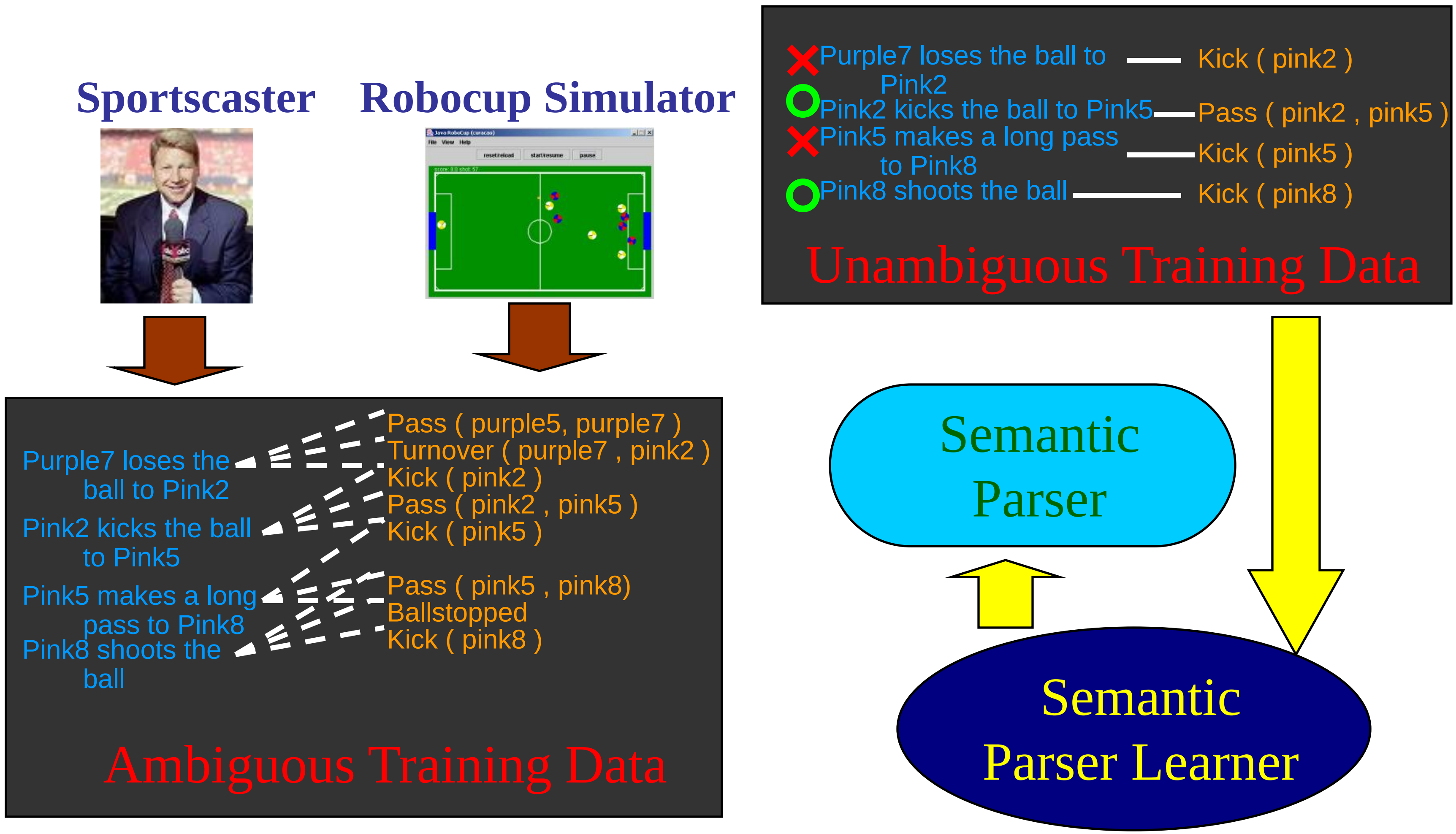
Ambiguous Training Data

Initial Semantic Parser

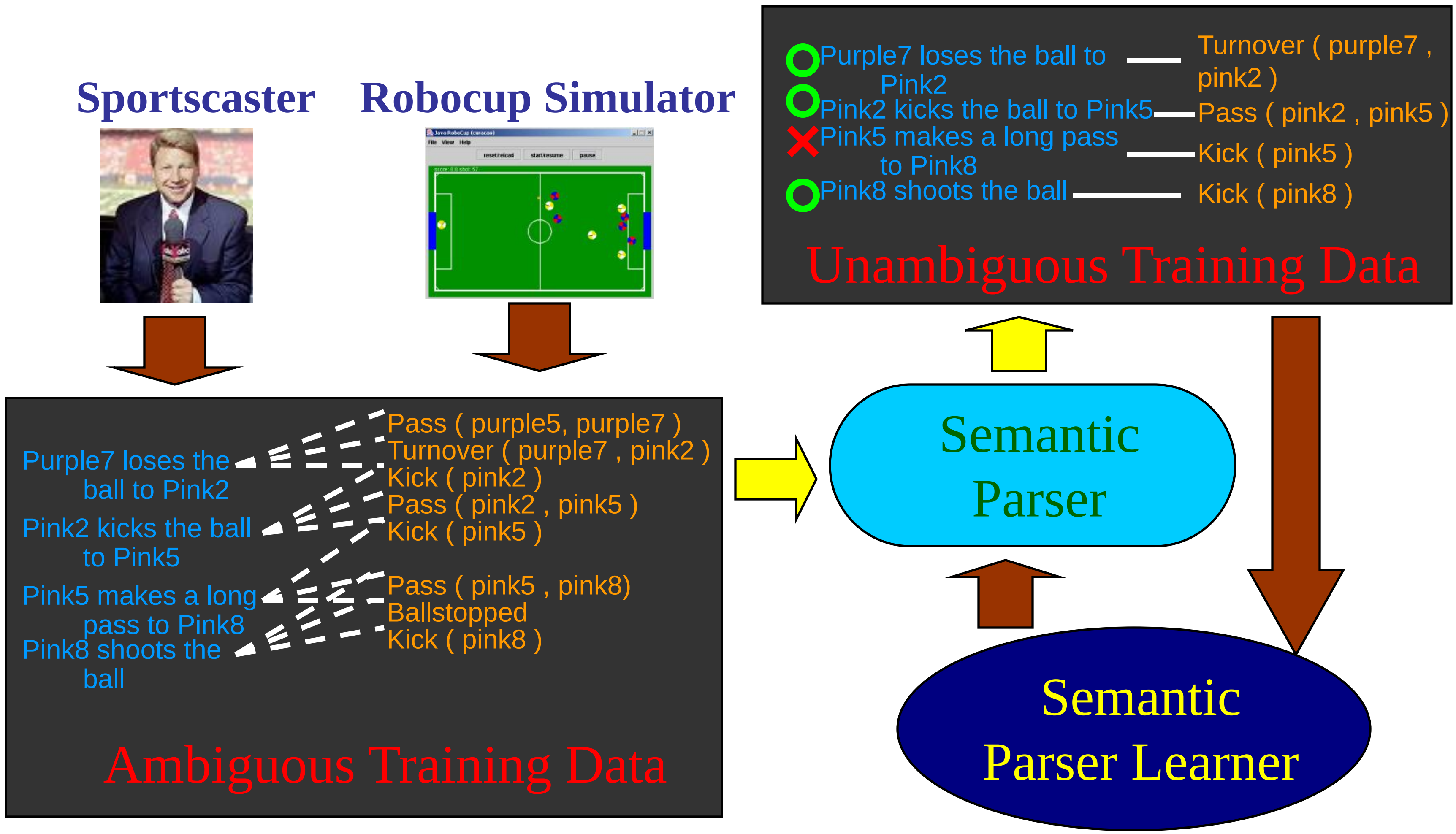
Purple7 loses the ball to Pink2 — Kick (pink2)
Pink2 kicks the ball to Pink5 — Pass (pink2 , pink5)
Pink5 makes a long pass to Pink8 — Kick (pink5)
Pink8 shoots the ball — Kick (pink8)

Unambiguous Training Data

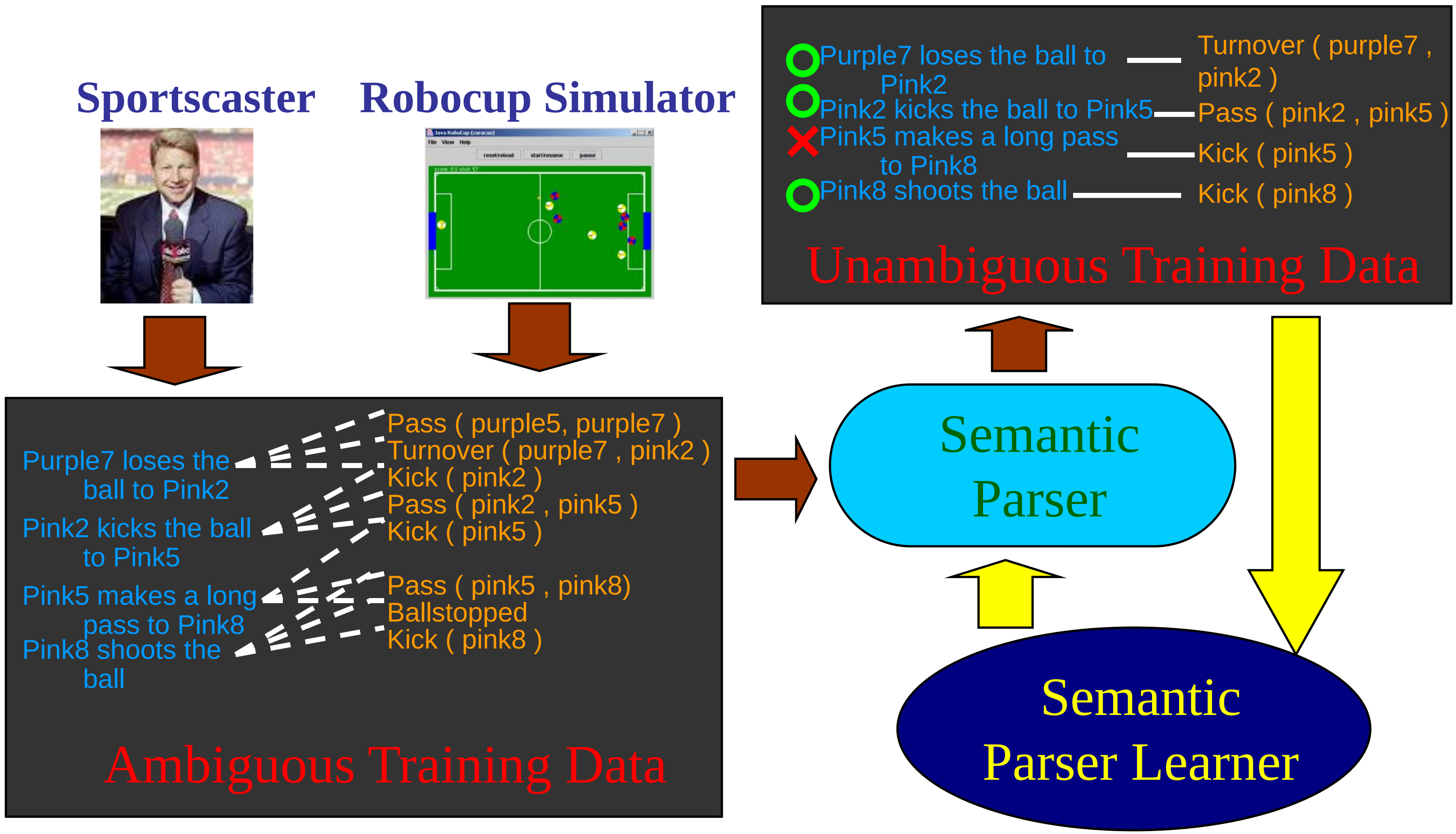
System Overview



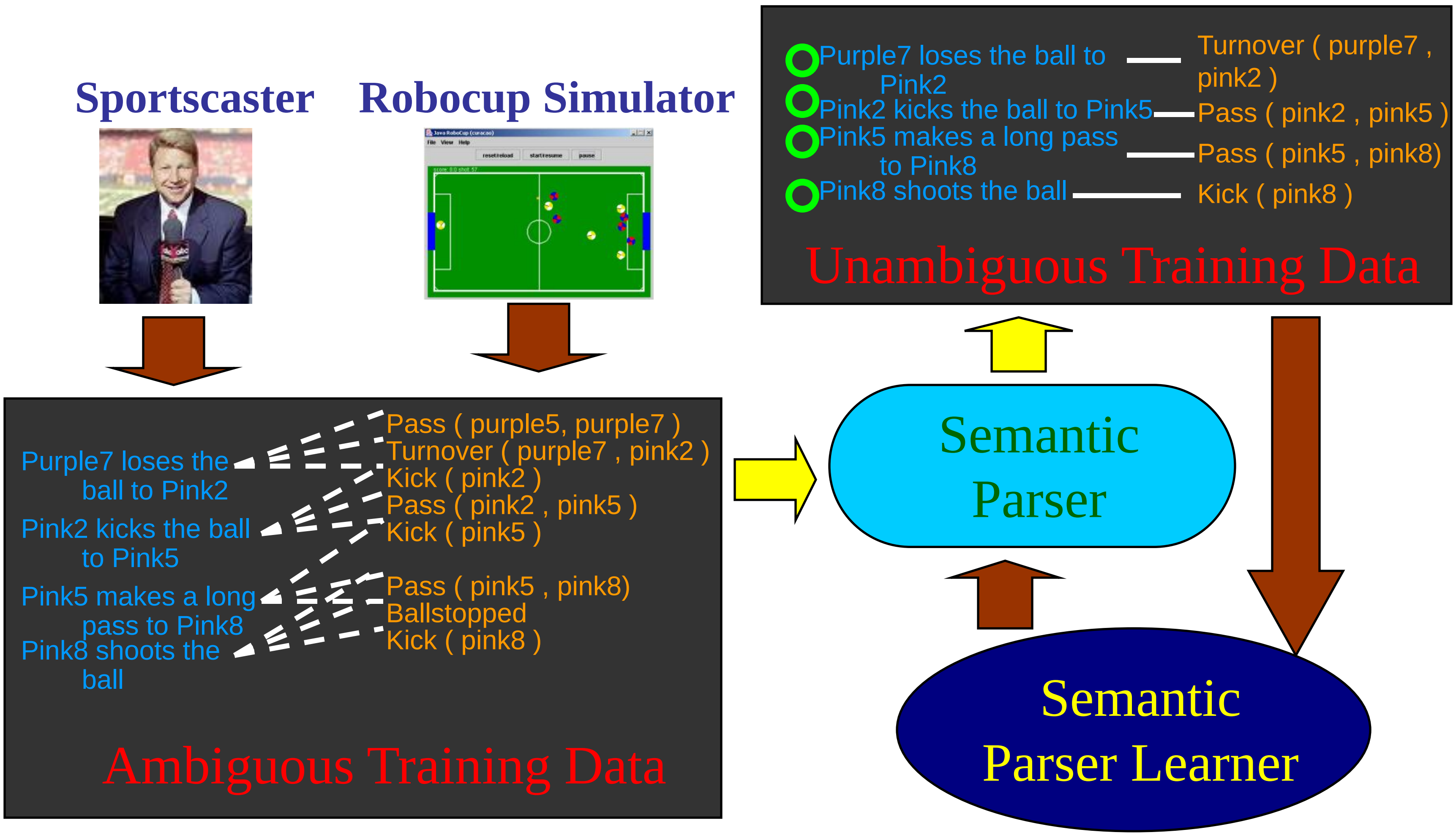
System Overview



System Overview



System Overview

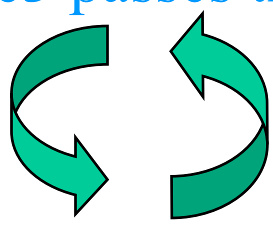


Semantic Parser Learners

- 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)

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)
- Capable of both semantic parsing and tactical generation

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

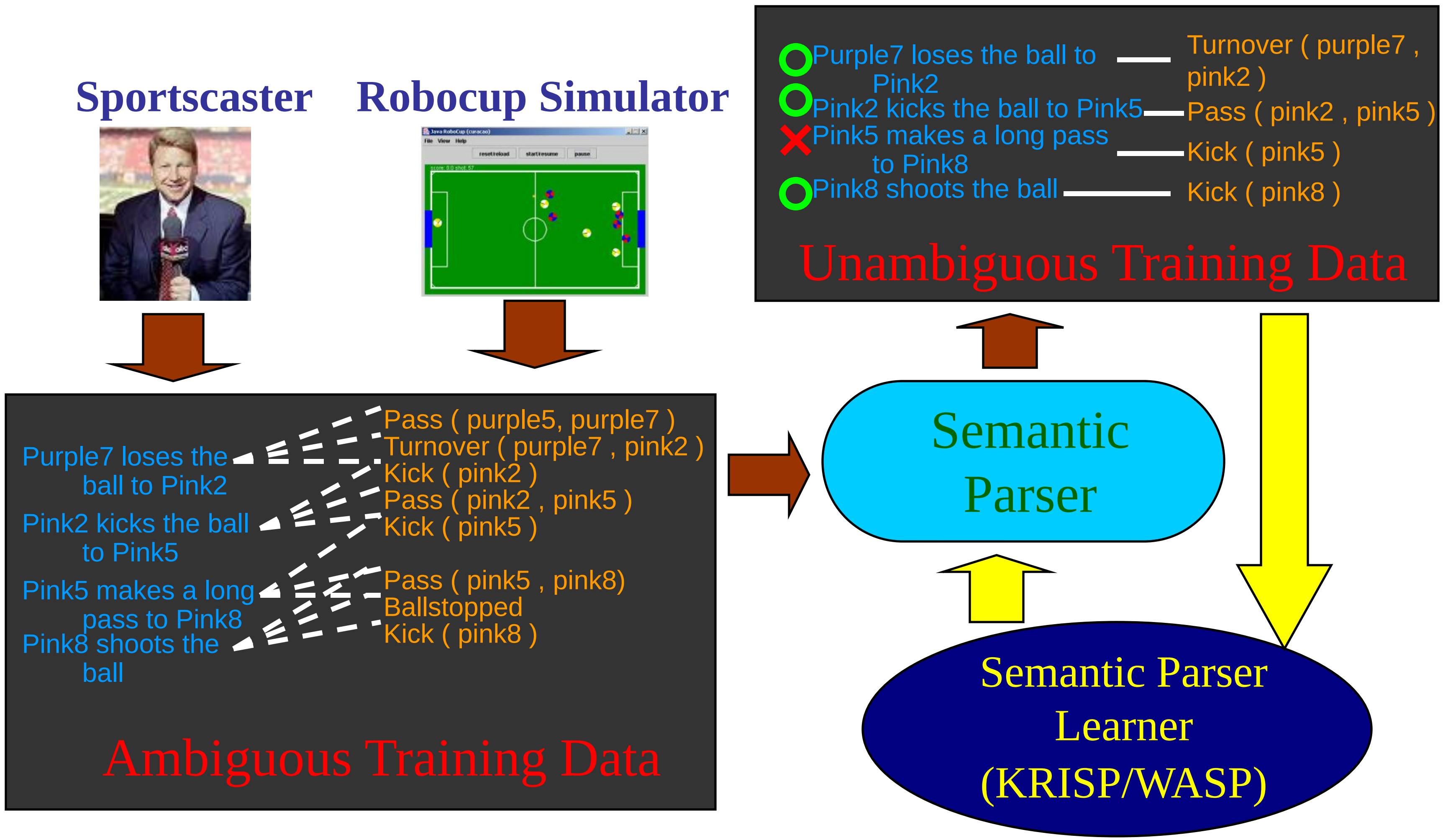
Matching

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

Systems

	Learner
KRISPER (Kate & Mooney, 2007)	KRISP
WASPER	WASP
WASPER-GEN	WASP's language generator

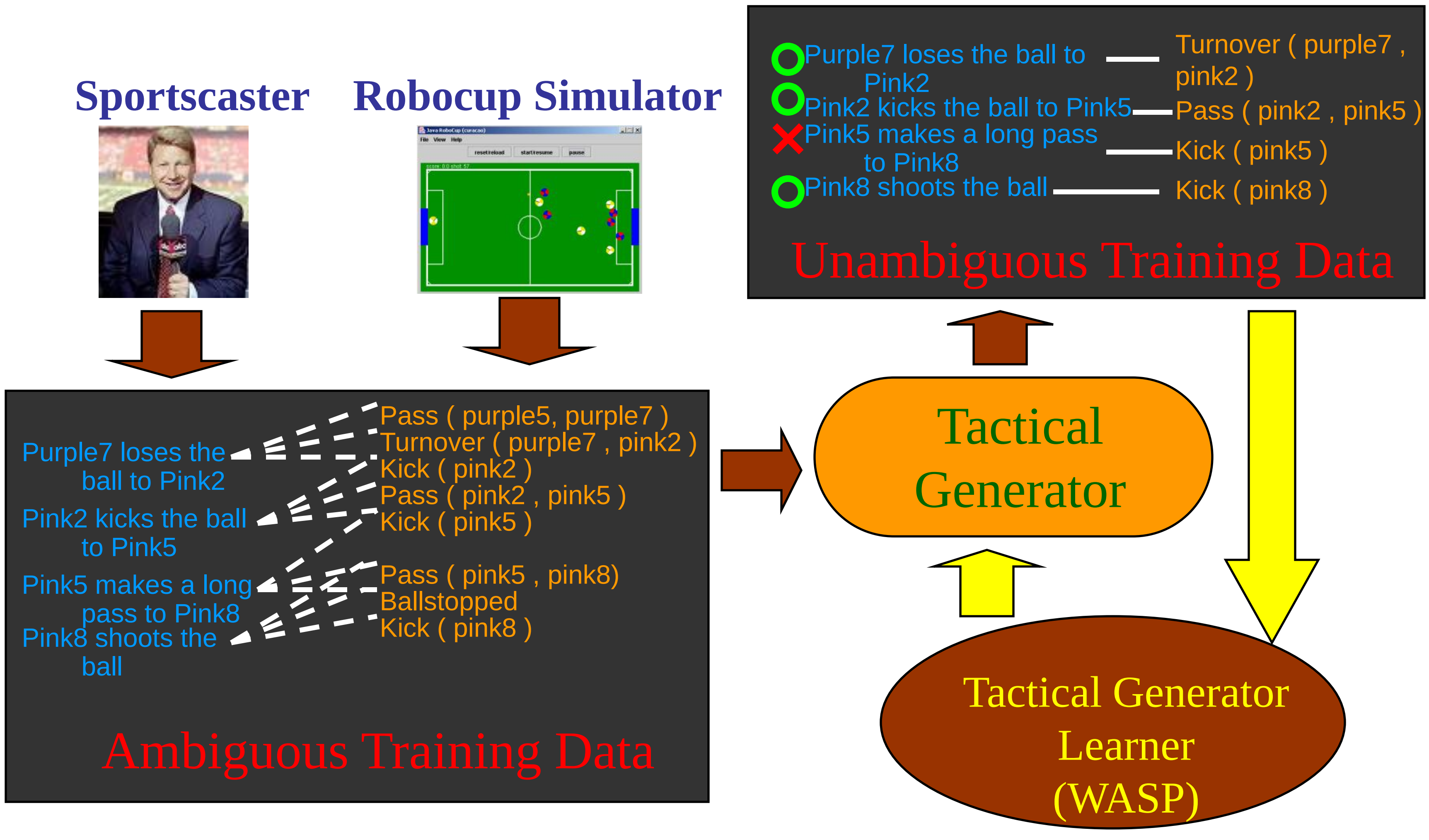
KRISPER and WASPER



Systems

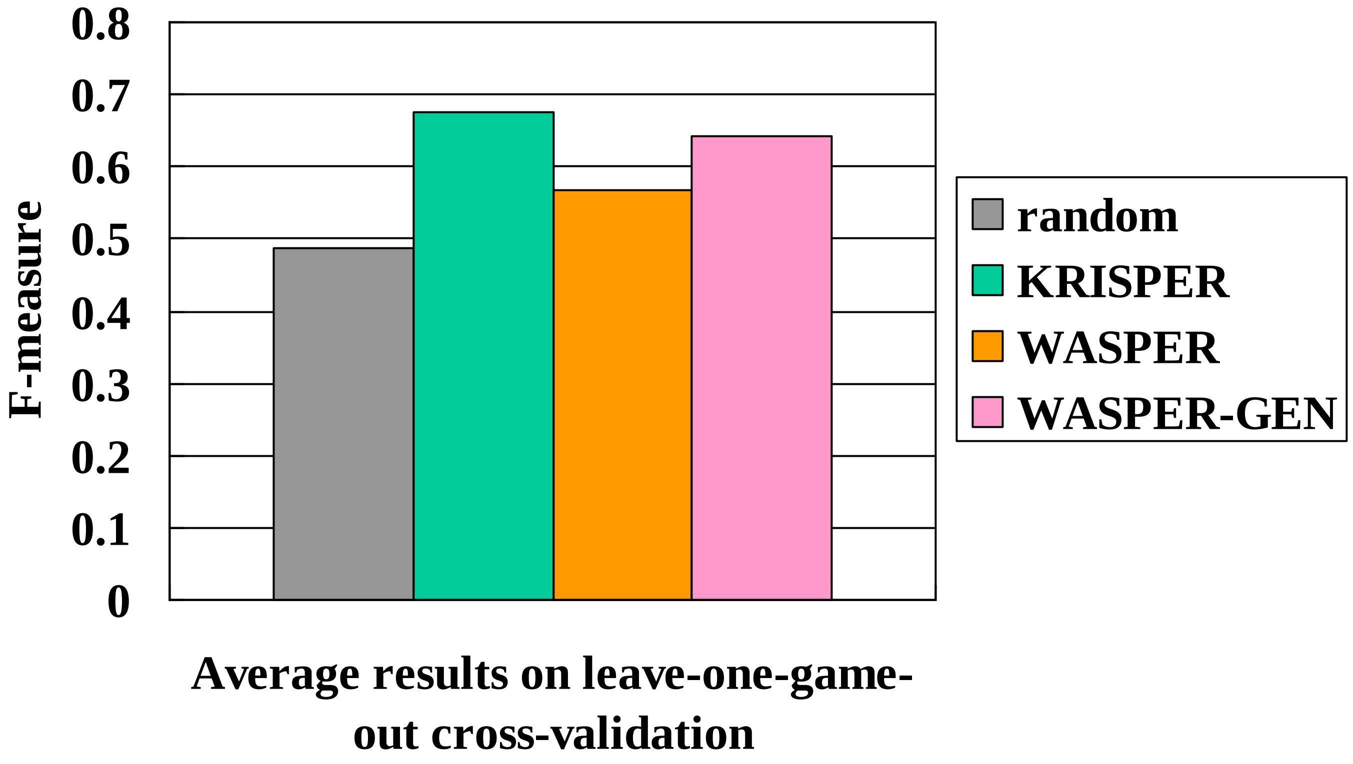
	Learner
KRISPER (Kate & Mooney, 2007)	KRISP
WASPER	WASP
WASPER-GEN	WASP's language generator

WASPER-GEN



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
Tactical Generator
Tactical Generator Learner (WASP)
Purple7 loses the ball to Pink2
Pink2 kicks the ball to Pink5
Pink5 makes a long pass to Pink8
Pink8 shoots the ball
Turnover (purple7 , pink2)
Pass (pink2 , pink5)
Kick (pink5)
Kick (pink8)
Unambiguous Training Data

Matching Results



Overview

- Sportscasting task
- 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

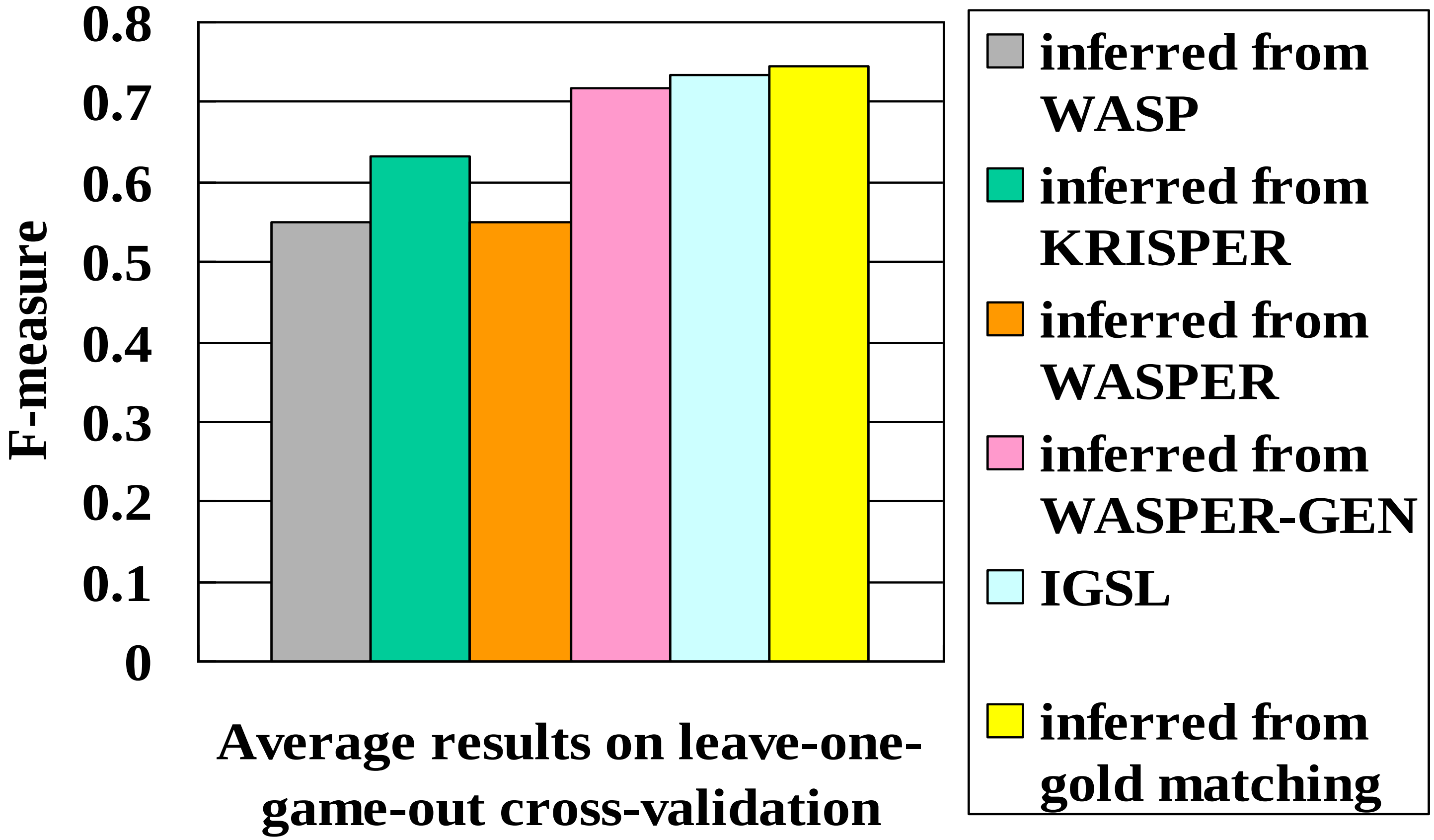
Iterative Generation Strategy Learning (IGSL)

- 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

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



Overview

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

Human Evaluation (Quasi Turing Test)

- 4 fluent English speakers as judges
- 8 commented game clips
  - 2 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

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.

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

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

Commentator	English Fluency	Semantic Correctness	Sportscasting Ability
Human	**3.94**	**4.25**	**3.63**
Machine	**3.44**	**3.56**	**2.94**
Difference	**0.5**	**0.69**	**0.69**

Future Work

- Expand MRs to beyond simple logic formulas
- Apply approach to learning situated language in a computer video-game environment (Gorniak & Roy, 2005)
- Apply approach to captioned images or video using computer vision to extract objects, relations, and events from real perceptual data (Fleischman & Roy, 2007)

Conclusion

- Current language learning work uses expensive, unrealistic 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 almost as well as humans.