Interactive Machines: From Questions **to** Experience

Giuseppe Riccardi

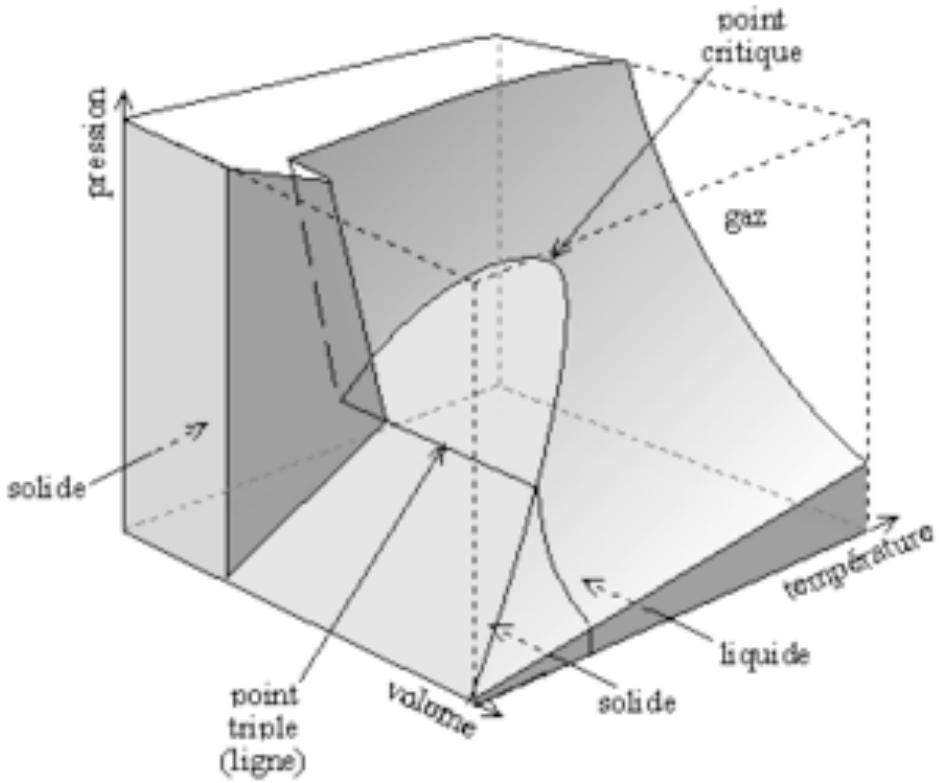
University of Trento, Italy

The Vision

- Why Natural Language Question/Answer ?
- Interactive machines
- Interactive eco-systems

For the sake of asking questions

- Knowledge will do it?
  - Why water increases its volume when it freezes?



For the sake of asking questions

- Knowledge will do it?

  Why water increases its volume when it freezes?

  – GOOGLE.com ➔CHACHA.com➔

  Q : When water freezes its volume increases by how much percentA

  A : When water freezes at 0°C its volume increases by about 9% under STP (Standard Temperature and Pressure).

  Is that correct?

Knowledge-Based Q&A

- Precision of the answers
- Accountability of the resources
- Limitation of Document space (vs WEB)
- Brittleness of the natural language interface

“Computers are useless. They can only give answers.”— *Pablo Picasso, (1881-1973).*

Search Engine Queries

Search vast amount of data using (almost) natural language

The search engine simple "drill"

1. Short Query (" universita' Trento ", "rifugio Passo San Pellegrino ")
2. Human Parse of Retrieved Documents
3. Human Decision whether to
   A. Followup on links OR
   B. GOTO 1

Knowledge-Based Q&A


JEOPARDY!

DEEPQA Project

- IBM Research Funded project
  - Build QA engine (WATSON) capable of challenging Humans in the Jeopardy! Game.
  - Domain is virtually open. NO DB rather unstructured data (light or full supervision).
- Open Collaborative Research Agreement btw UTrento and IBM YorkTown
  - MIT, UTexas, CMU, USC
- Utrento
  - Interactive Machines to resolve Questions into Actions/ Decisions/Tasks
  - Machine Learning models for Parsing Sentences for QA Classification, Reranking Hypotheses (target Answers)

Natural Language Parsing and applications to DEEPQA

w **Alessandro Moschitti**

Outline

- Motivations
- Two important problems in Jeopardy
  - Question Classification
  - Answer Selection
- Results
- Conclusions

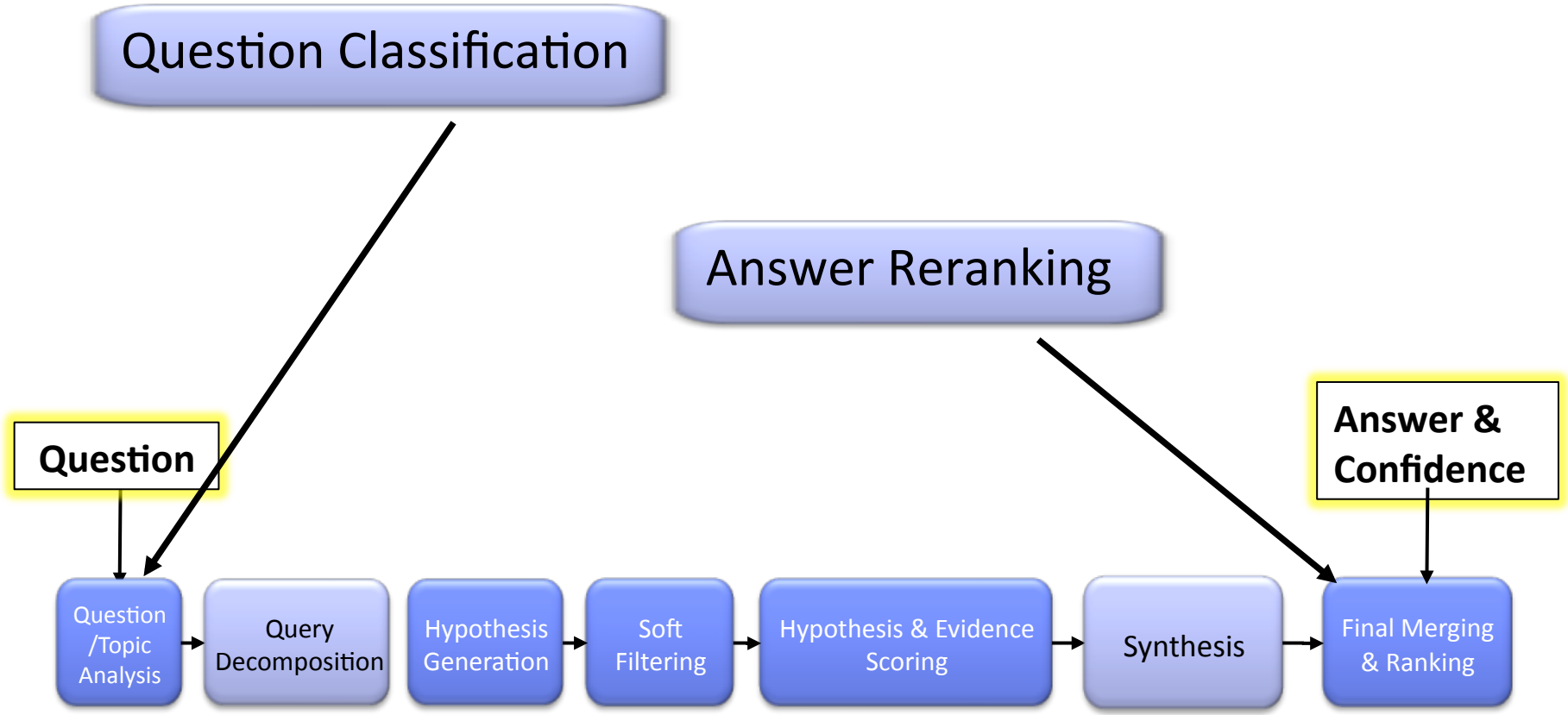
Let us Consider one Jeopardy Cue

- *When hit by electrons, a phosphor gives off electromagnetic energy in this form*
- Solutions: ***photons/light***
- *Electrons*, *phosphor* and *electromagnetic energy* are in a relationship which gives the solution
- How representing and using such relation in a machine?

Issues and solutions

- How exploiting shallow semantic information?
- How dealing with noise and errors of semantic parsing?
- Designing hand-crafted rules to deal with any kind of errors or noise is not realistic
- We need statistical methods as they:
  - "can activate the rules" that maximize the probability of success
  - can automatically learn such probabilities from data

Important Tasks in Jeopardy



Classification in Definition vs not Definition in Jeopardy

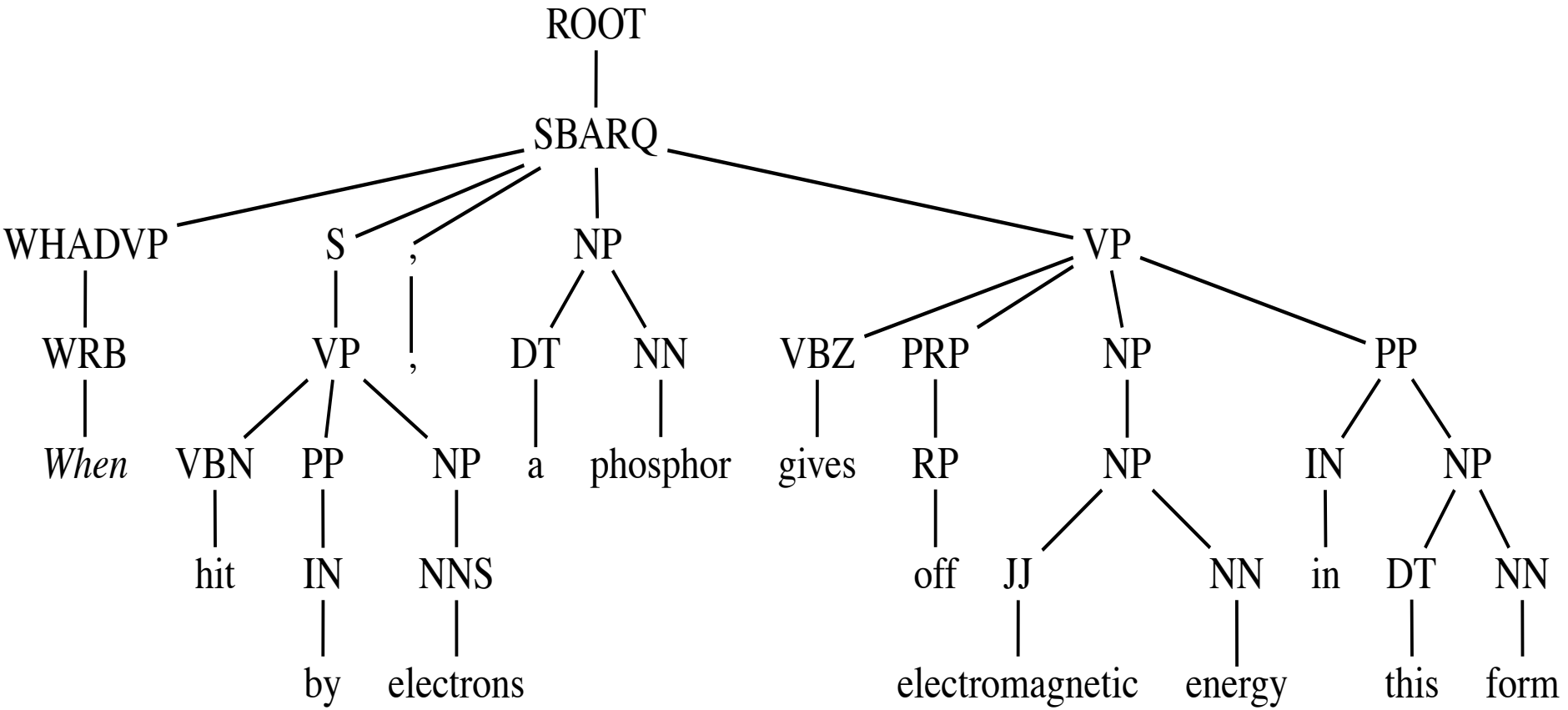
- *Usually, to do this is to lose a game without playing it*

  (solution: *forfeit*)
- *When hit by electrons, a phosphor gives off electromagnetic energy in this form*

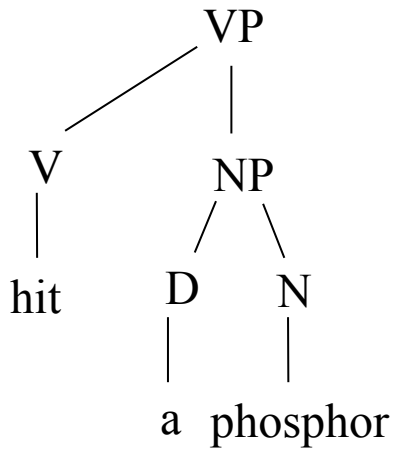
Our Approach

- Supervised Approach
  - Positive and negative examples of definition questions
  - Syntactic information is intuitively important
    - Apply off-the-shelf parsers
  - As this is a new task, to extract features, we exploit tree kernels, e.g. the Partial Tree Kernel (Moschitti, ECML 2006)

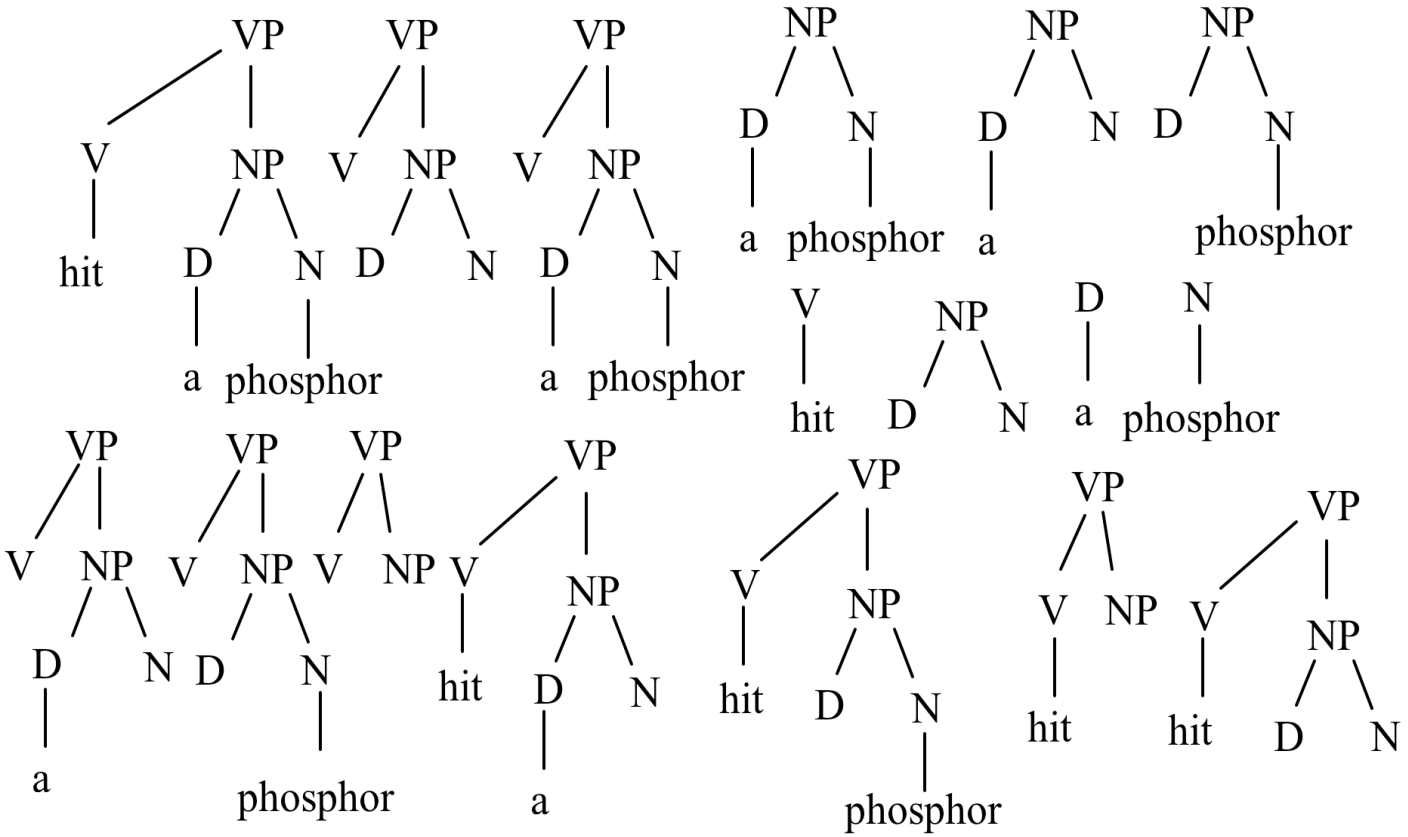
Parse Tree


ROOT
SBARQ
WHADVP
S
,
NP
VP
WRB
VP
,
DT
NN
VBZ
PRP
NP
PP
When
VBN
PP
NP
a
phosphor
gives
RP
NP
IN
NP
hit
IN
NNS
off
JJ
NN
in
DT
NN
by
electrons
electromagnetic
energy
this
form

Similarity based on the number of common substructures



A portion of the substructure set



Experiments on Jeopardy Question Classification

Model	Precision	Recall	F1
RBC	28.27	70.59	**40.38**

66.7% of relative improvement on the rule-based classifier

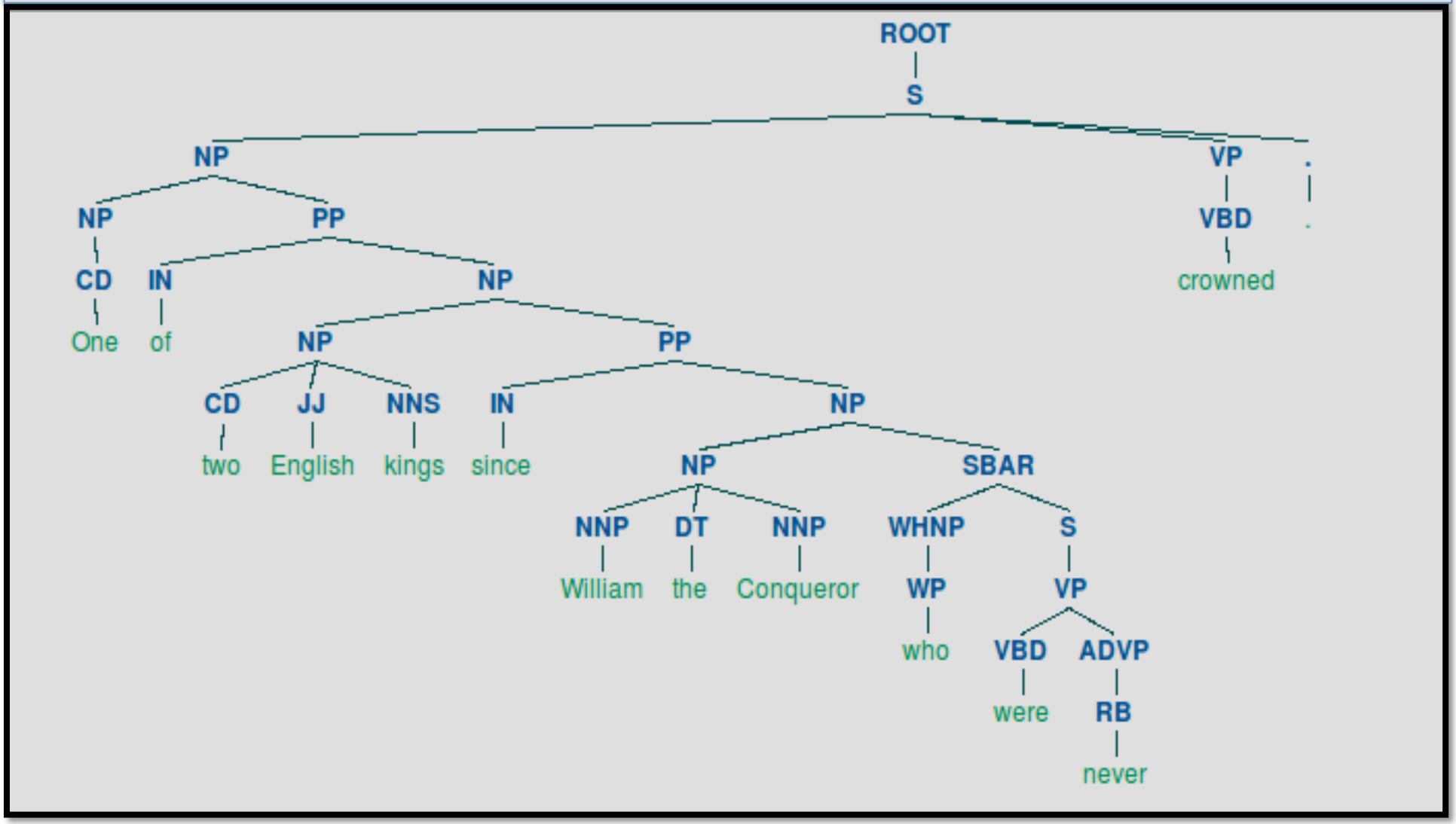
PTK+WSK +CSK+RBC	67.66	66.99	**67.32**

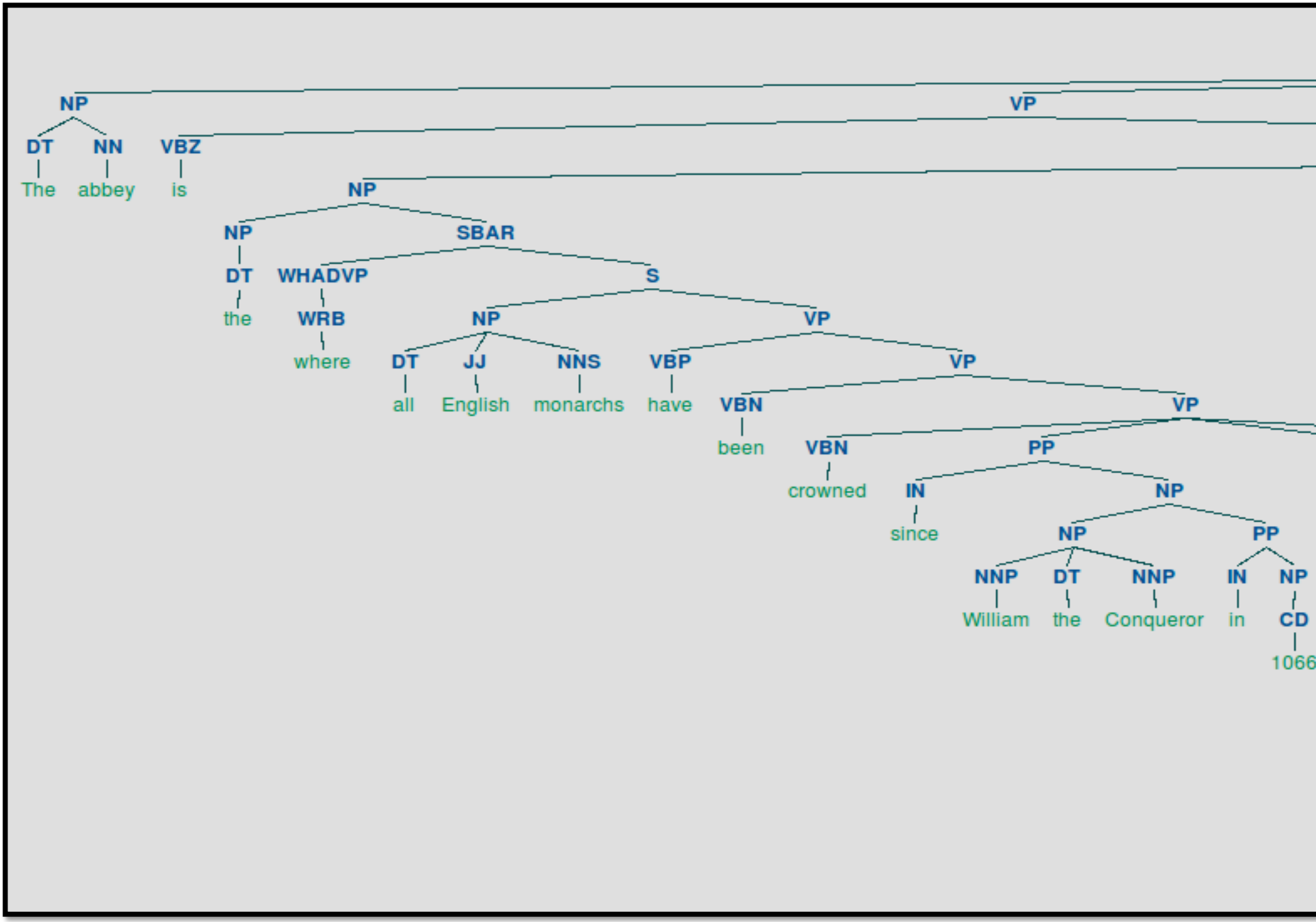
- Rule Based Classifier (RBC)
- Only Word Overlap (BOW)
- Category Subsequences (CSK)
- Partial Tree Kernel (PTK)

Our Approach to Answer Re-ranking

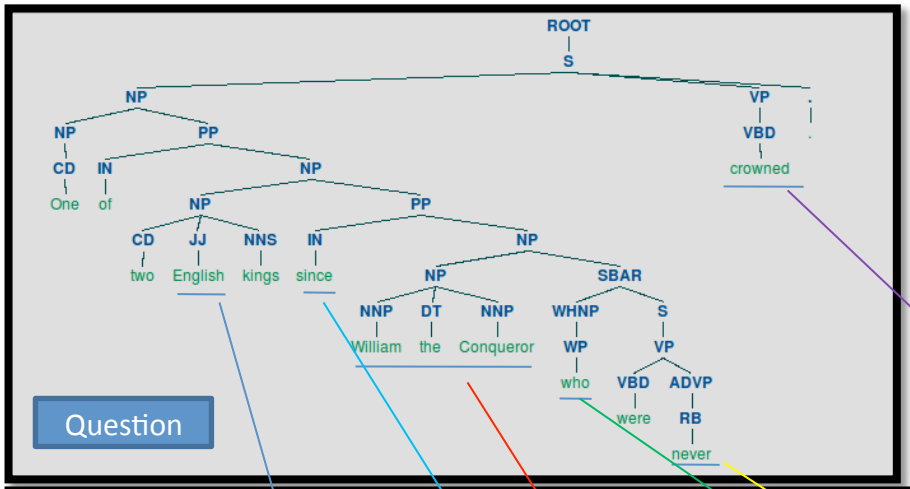
- Learn a classifier of <question,answer> pairs
  - Positive: the answer is correct
  - Negative: otherwise
- Kernel approach
  - Several kernels applied to both questions and answers

An example of Jeopardy Question



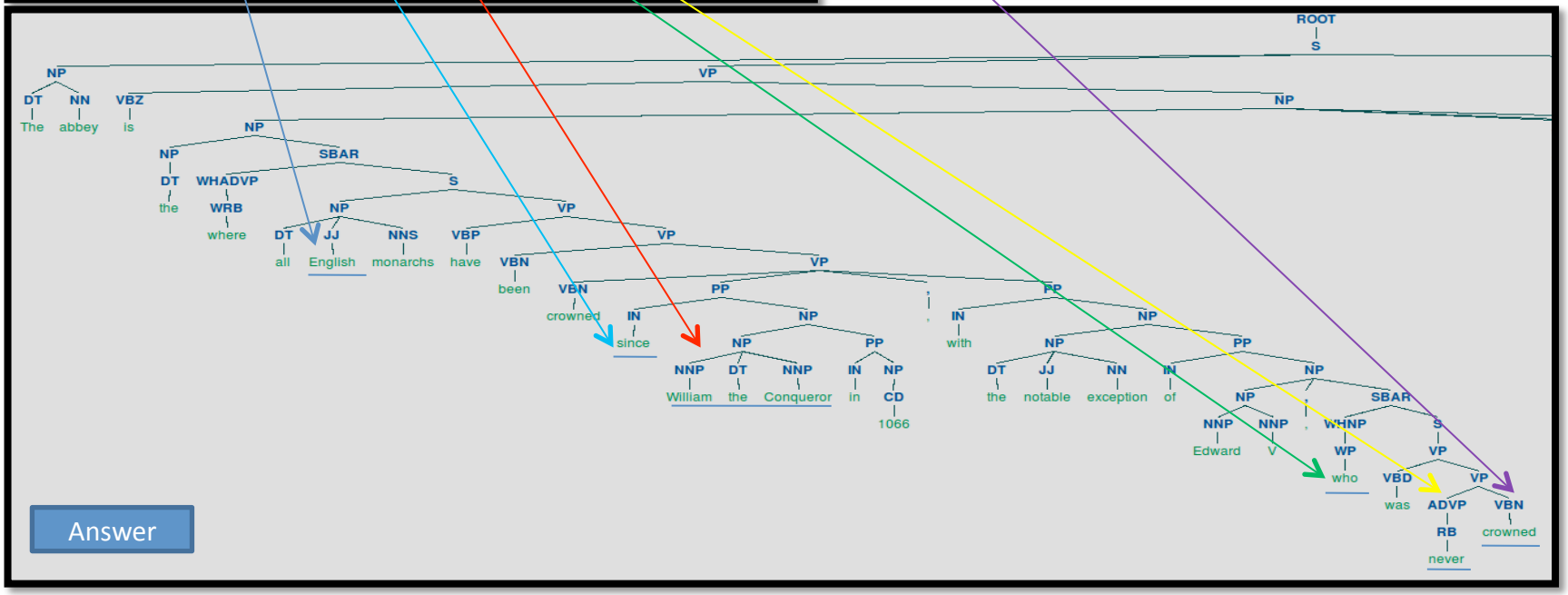

NP
VP
DT
NN
VBZ
The
abbey
is
NP
NP
SBAR
DT
WHADVP
S
the
WRB
NP
VP
where
DT
JJ
NNS
VBP
VP
all
English
monarchs
have
VBN
VP
been
VBN
PP
crowned
IN
NP
since
NP
PP
NNP
DT
NNP
IN
NP
William
the
Conqueror
in
CD
1066

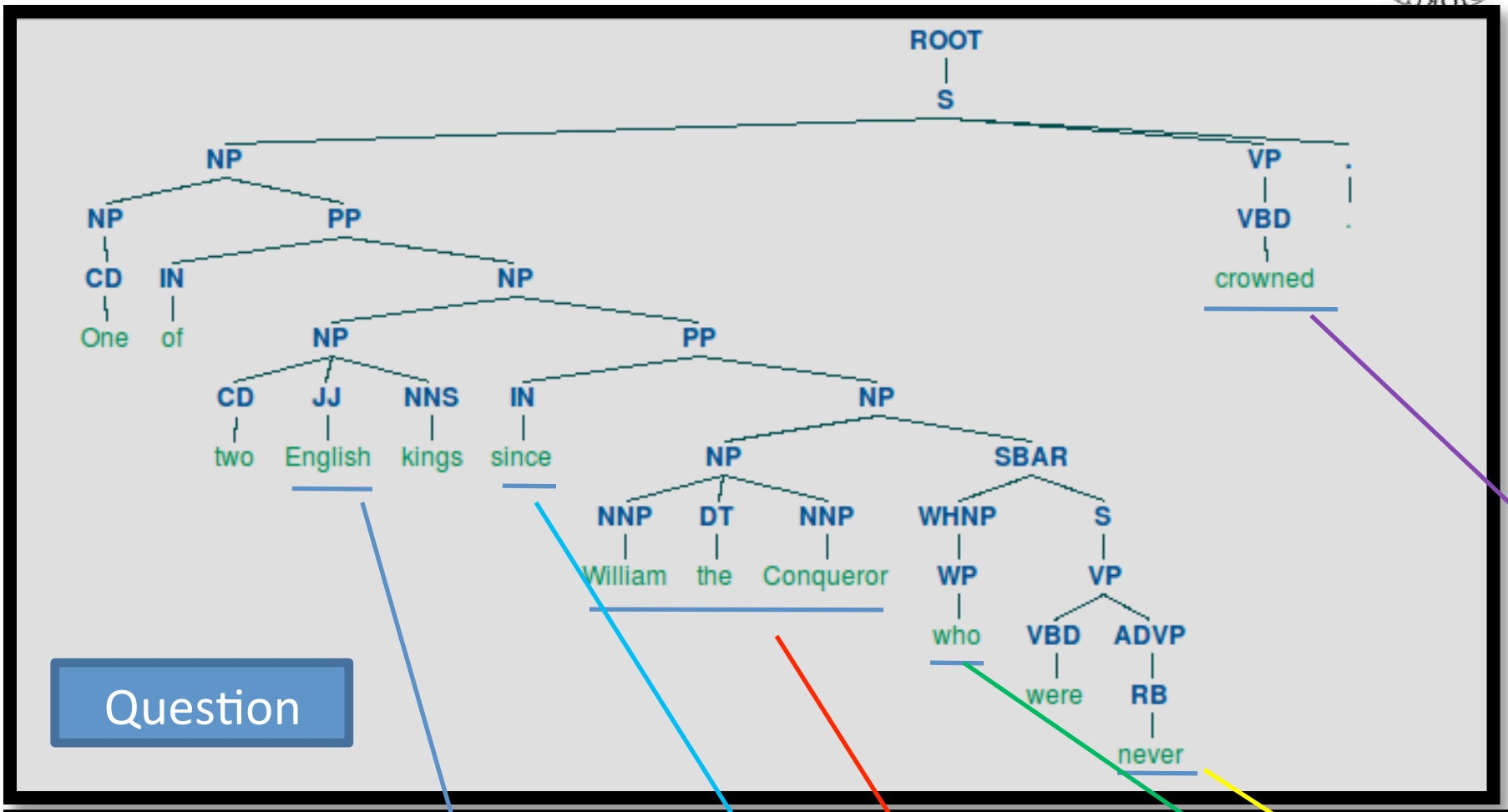
Baseline Model

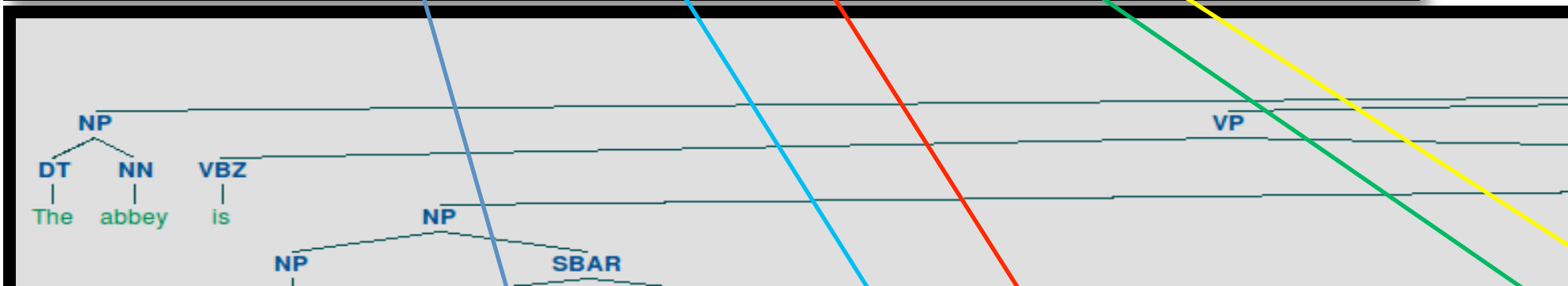


Methodology:

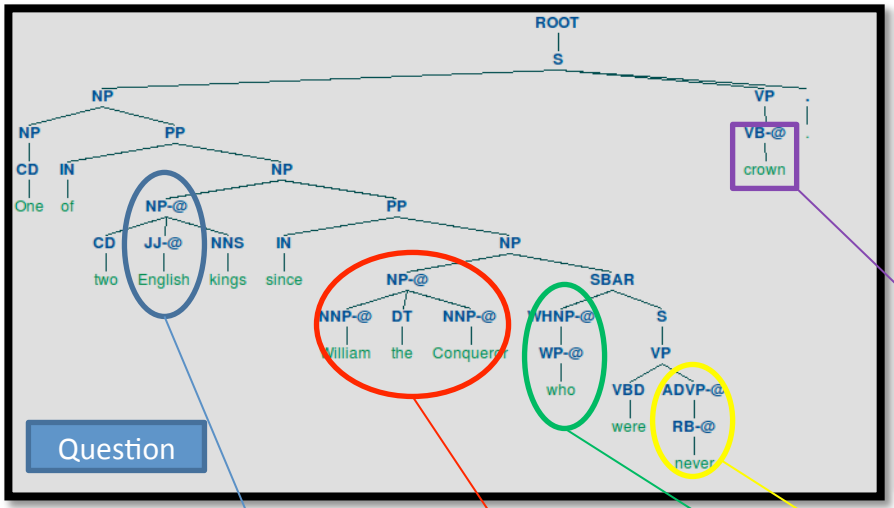
1-Applying PTK without any extra annotation and evaluate the model as baseline.




ROOT
S
NP
VP
.
NP
PP
VBD
.
CD
IN
NP
crowned
One
of
NP
PP
CD
JJ
NNS
IN
NP
two
English
kings
since
NP
SBAR
NNP
DT
NNP
WHNP
S
William
the
Conqueror
WP
VP
who
VBD
ADVP
were
RB
never
Question


NP
VP
DT
NN
VBZ
The
abbey
is
NP
NP
SBAR

Best Model

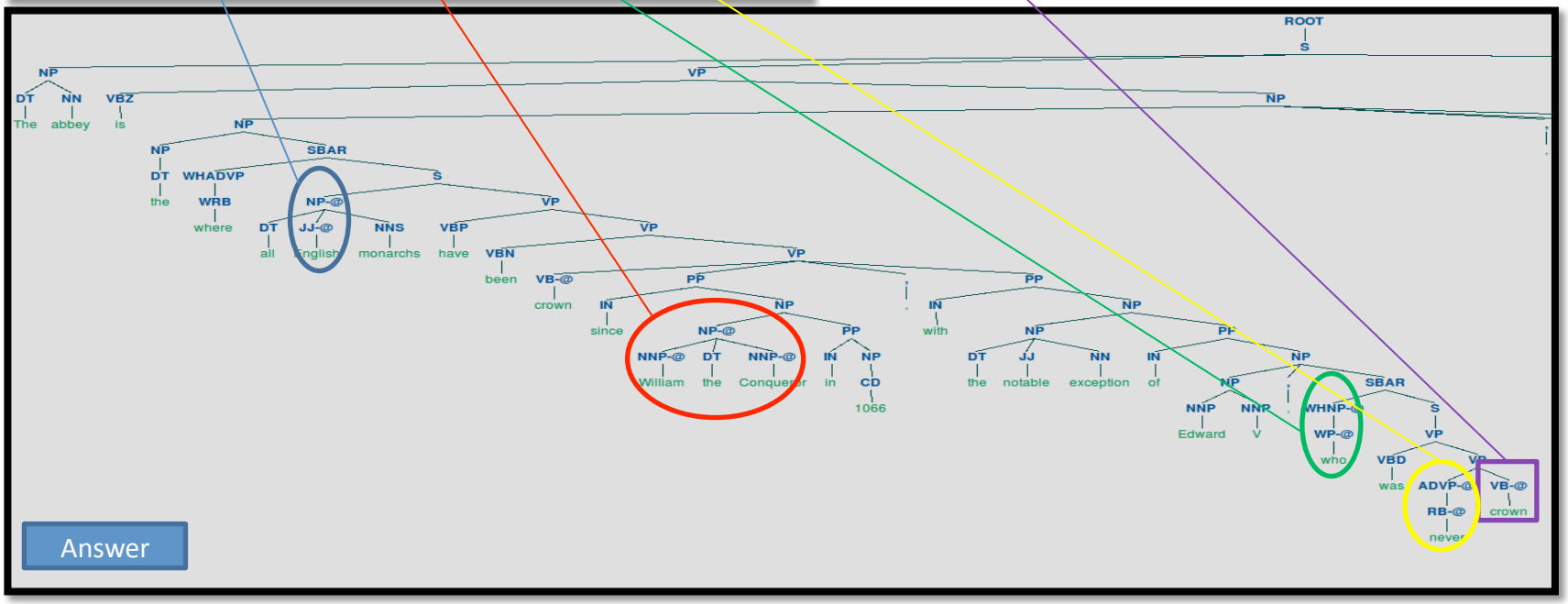


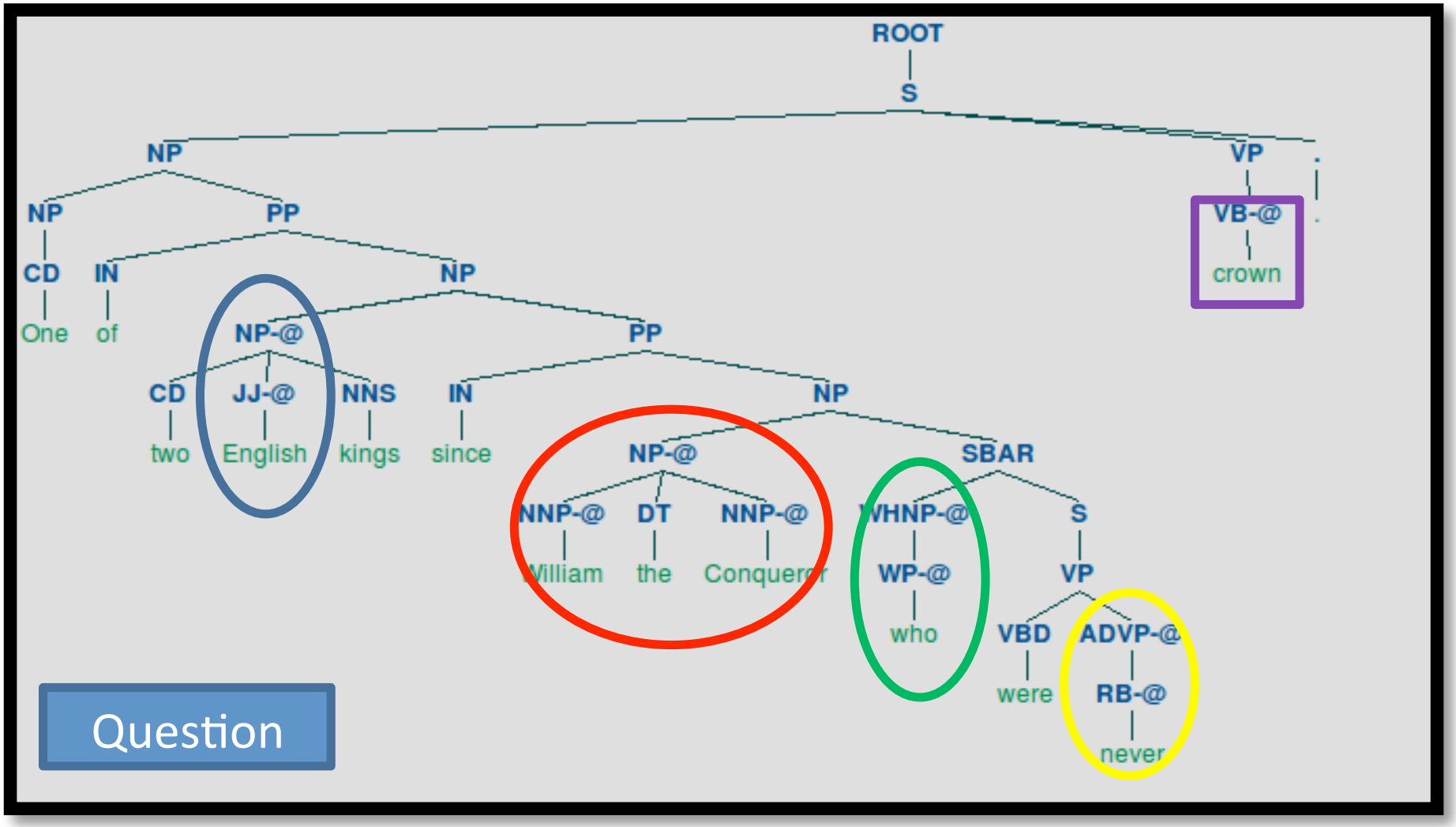
Methodology:

1-Applying lemmatization and stemming in leaves level.

2-Add an anchor to pre-terminal and higher levels if the sub-trees are shared in Q and A.

3-Ignore stop words in matching procedure.




ROOT
S
NP
VP
.
NP
PP
VB-@
CD
IN
NP
crown
One
of
NP-@
PP
CD
JJ-@
NNS
IN
NP
two
English
kings
since
NP-@
SBAR
NNP-@
DT
NNP-@
WHNP-@
S
William
the
Conqueror
WP-@
VP
who
VBD
ADVP-@
were
RB-@
never
Question

Precision/Recall Evaluation

	Precision	Recall	F1
Baseline with PTK (Partial Tree Kernel)	17.1%	59.37%	26.56
Best Model (STK) Syntactic Tree Kernel	17.71%	5.59%	8.50
Best Model with PTK	24.10%	61.12%	34.57

Ranking Precision at 1 and 5 best

	Evaluation on 1-best	Evaluation on 5-best
	Accuracy	Accuracy
Baseline	0.22	0.22
PTK (Baseline)	0.23	0.19
PTK (Best Model)	0.33	0.23

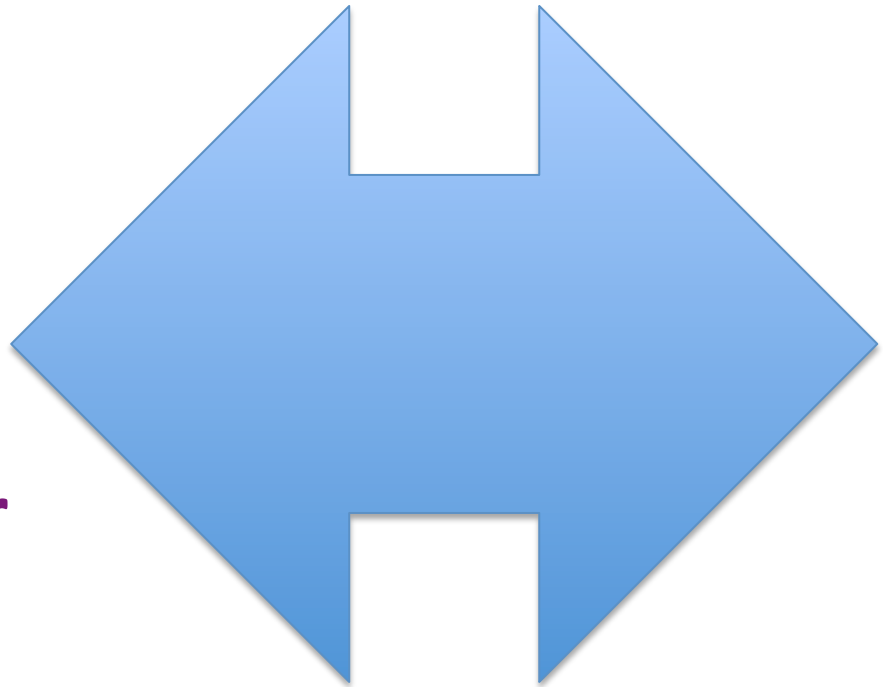
Conclusions

- We use of powerful ML algorithms
  - e.g. Support Vector Machines
  - robust to noise
- Abstract representations of examples
  - Similarity functions (Kernel Methods)
- Modeling Question semantics with advanced syntactic and shallow semantic structures
- Software already integrated in the Jeopardy System

Questions to Actions

Social Networks

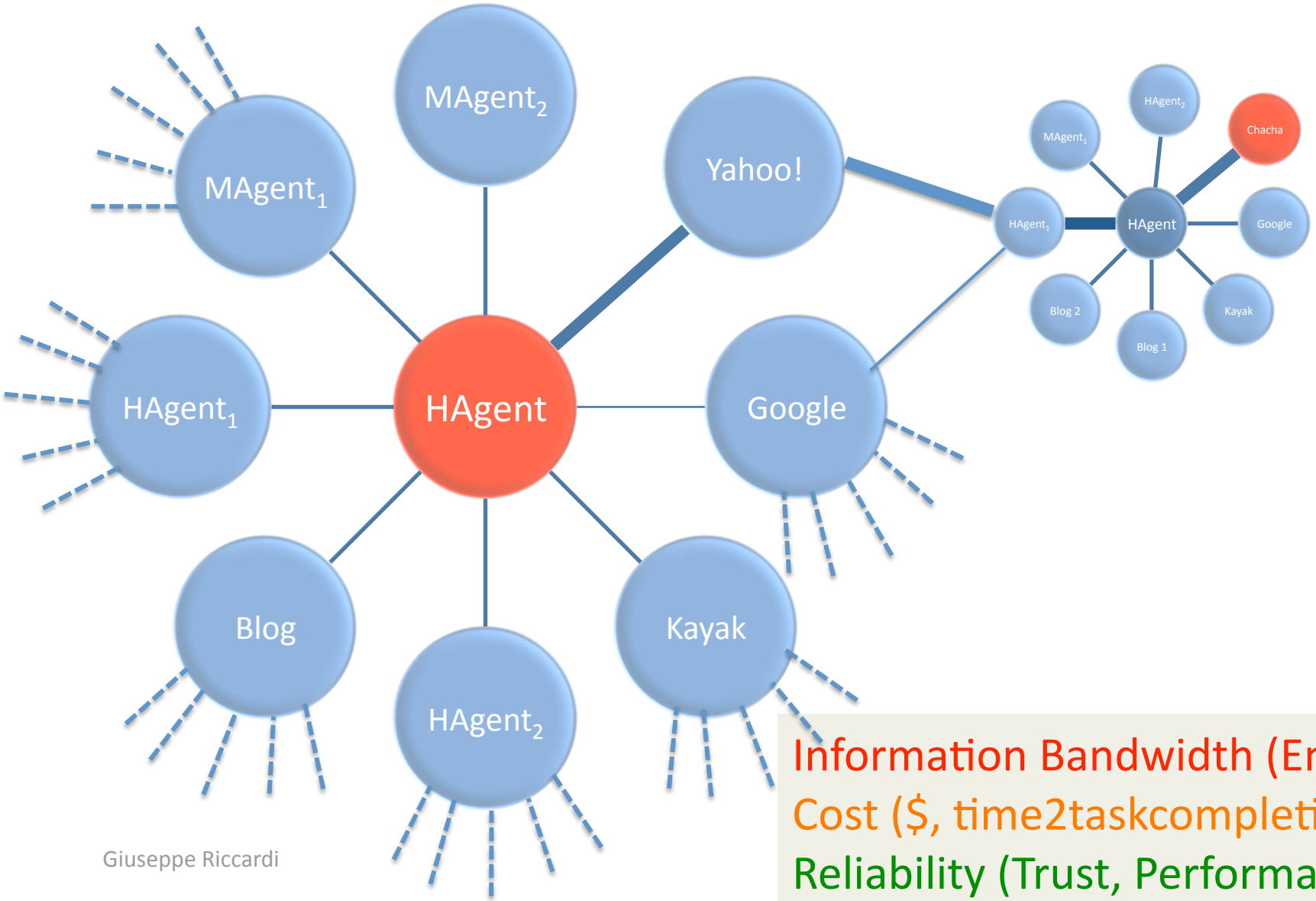
Social Interactions



Customer	Company
Citizen	Public Admin.
Patient	Doctor
Citizen	Political Party
User	User
Worker	Supervisor
Peer	Peer
Relative	Relative
……	……

Can computers be
part of the social network ?

Network of Agents

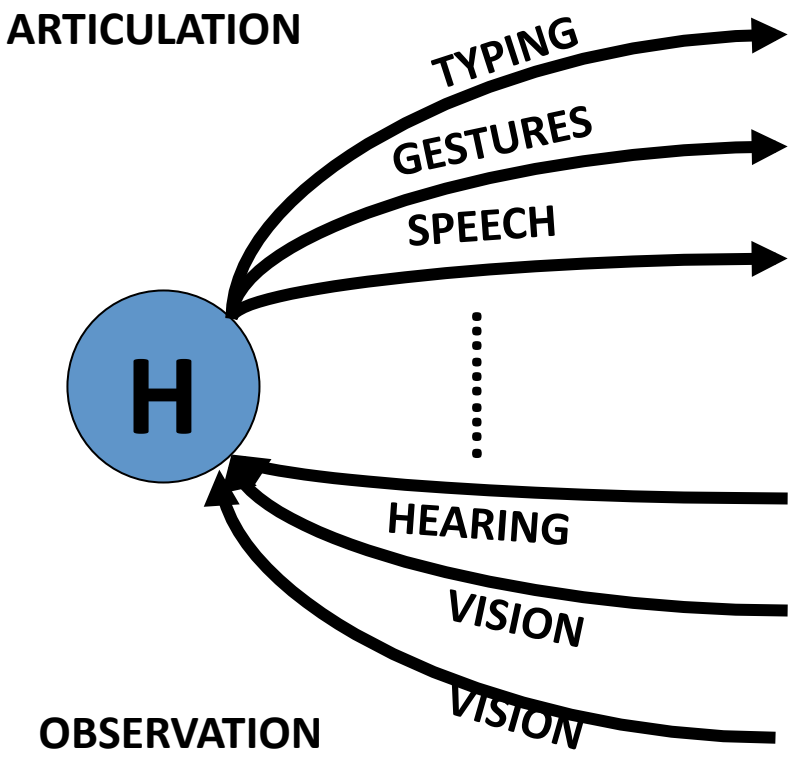


Information Bandwidth (Entropy)

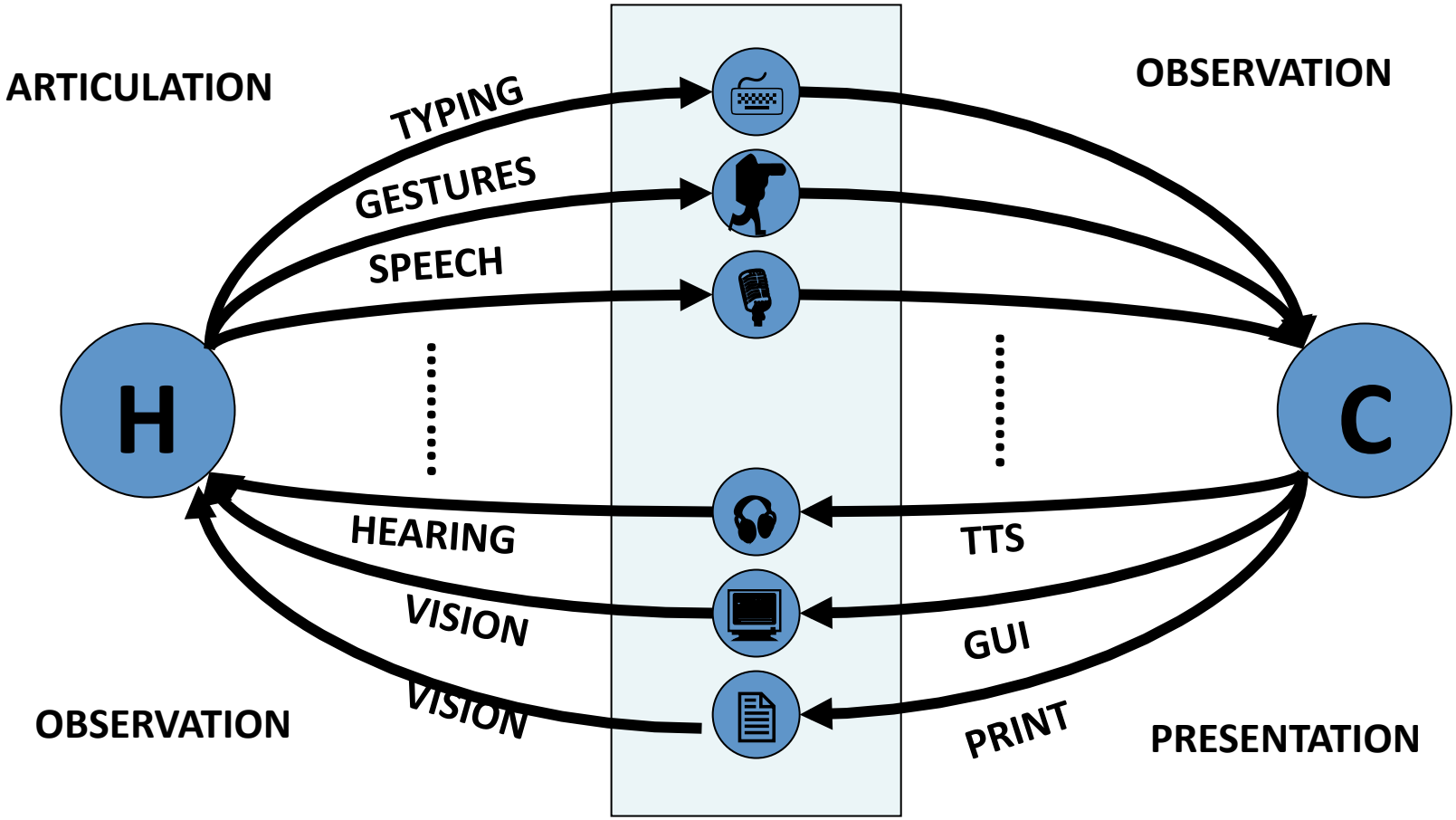
Cost ($, time2taskcompletion)

Reliability (Trust, Performance)

Human – Computer Interaction

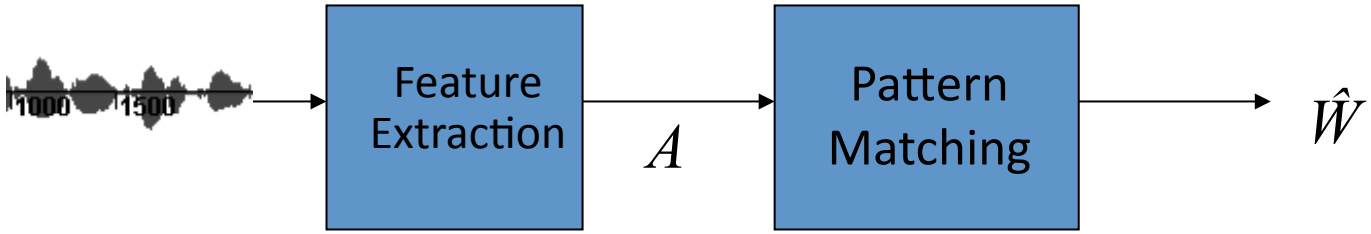


Human – Computer Interaction

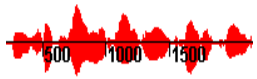


Automatic Speech Understanding

Given the acoustic observation sequence $A=a_1,a_2,\ldots,a_m$,



what is the most likely "word" sequence $W=w_1,w_2,\ldots,w_n$?

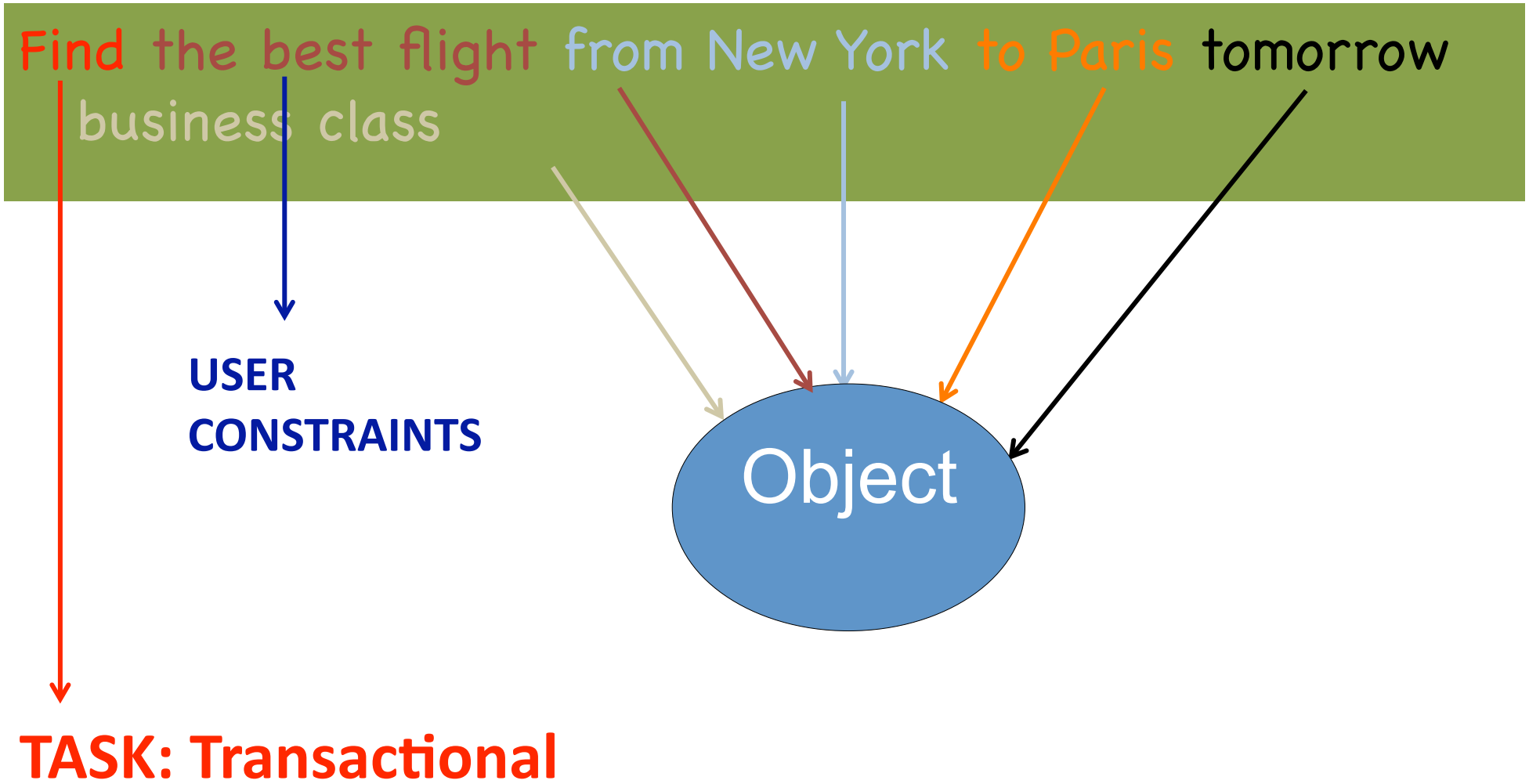


L EH T AH S P R EY

Let us pray

Lettuce spray

Natural Language Query to DB



Machine Understanding

- **User**:

  "Find the best flight from New York to Paris tomorrow business class"

- **Speech Recognition**:

  " Find the bass flight from Newark to Paris tomorrow business class"

- **Modeling Uncertainty**:
  - @action=Request-Reservation (0.9)
  - @origin=Newark (0.5)
  - @time-departure=Tuesday (0.7)
  - @destination=Paris (0.8)

Cooperative Task

SW-HW Help Desk

U Hi Good Morning

O Hi, How May I Help You?

U I am Roberta Sicconi calling from Cultural Affairs at City Hall.

Personal Identification

U I had made a request for a password change yesterday

O Ok do you have the request track id?

U Uhm No I cannot find

O Ok do you have the date of the request?

U Well that was yesterday

O...ok I think I can find it..I got it

O It's for a password reset.

Problem Statement Ticket Record Retrieval

U Right. The problem is that I changed the password when I first logged in..

Problem Resolution (USER)

O You were supposed to change first time you logged in. Now let's try together to log in

O can you tell me you RVS of your computer

U Well let me see. This is a new PC to me. Where can I find it?

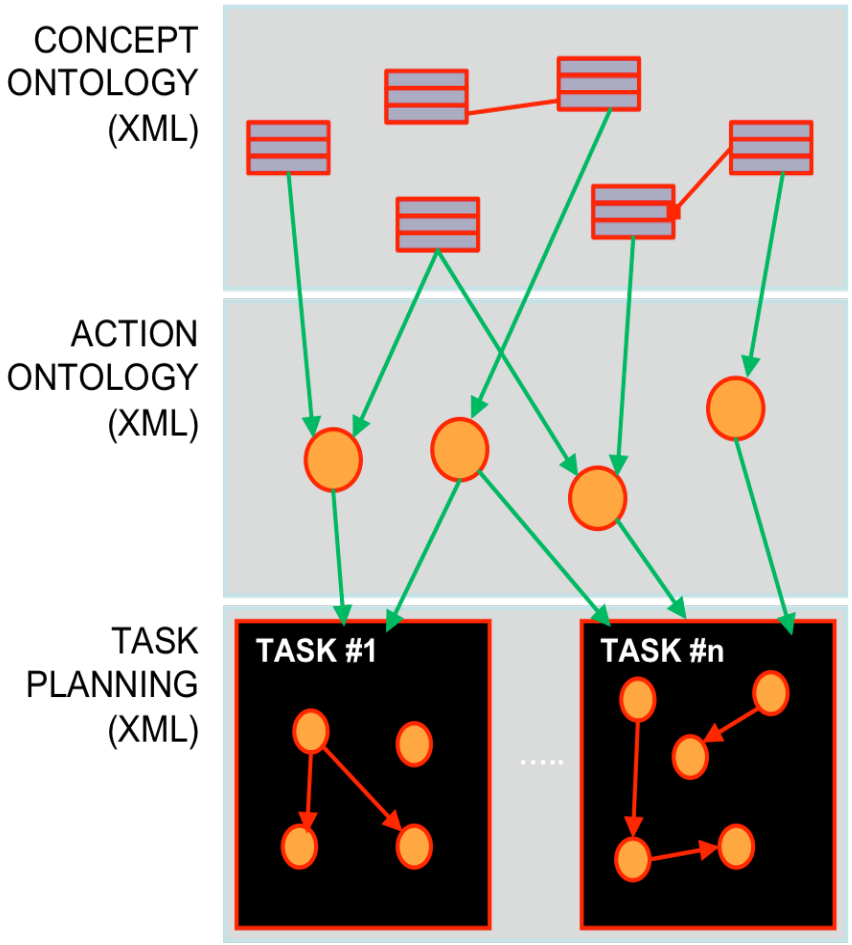
O Usually the tag is right next to the base of the chassy next to the power switch. It reads "inventario settore informatico".

U Inventario Settore

Problem Resolution (PART I)

Dialog Models (DM)

Rule-Based Systems

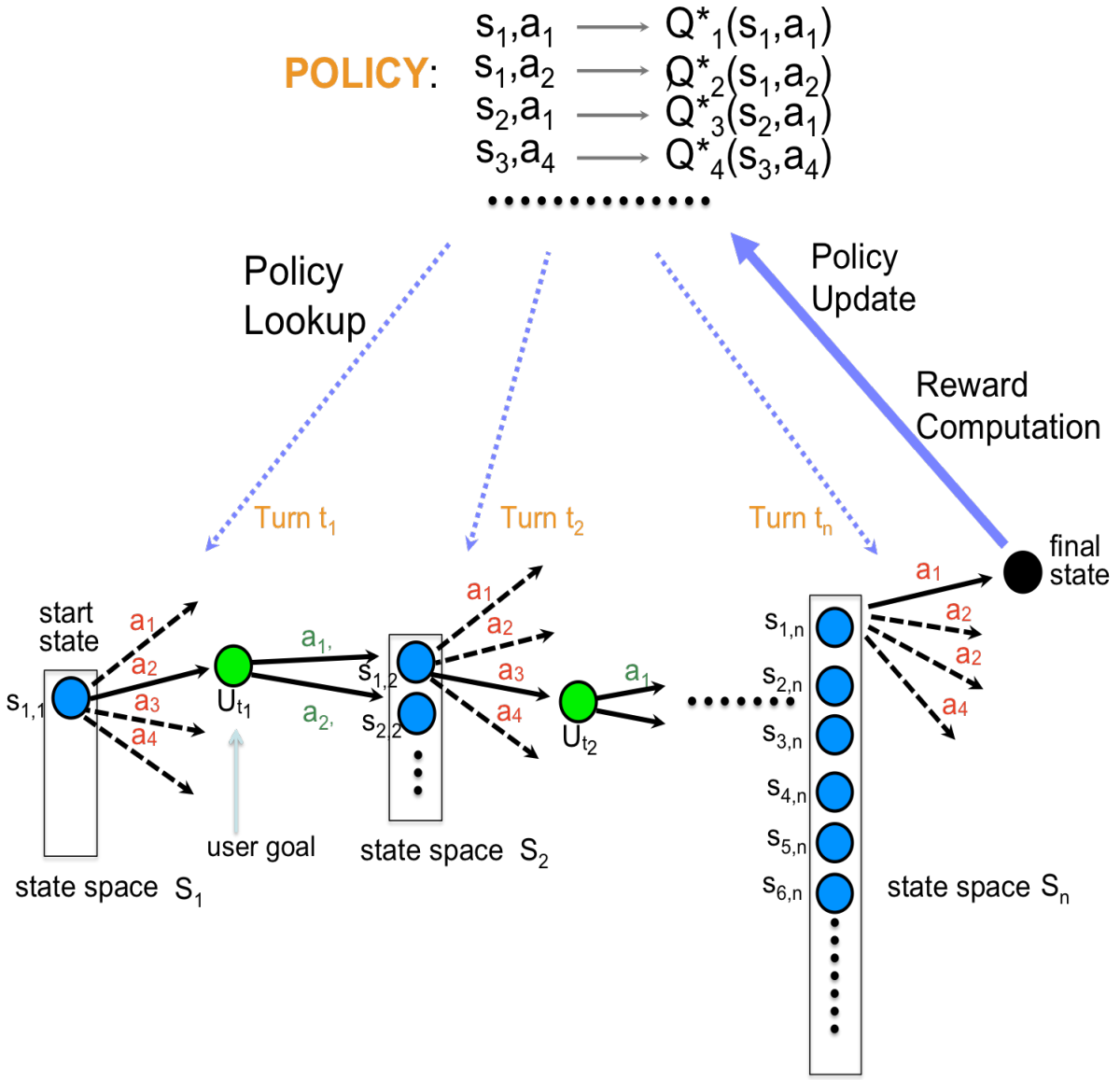


1) Domain Representation

2) Task Representation

3) Task Execution

Markov Decision Processes



Conversational Agents

SYSTEM: Welcome to LUNA. Good day, I am Paola. How may I help you?

USER: Eh, Sorry. I have a problem with the printer.

ASR And I am sorry a problem with the printer

SLU

Concept	Value	Conf
conjugation	*and*	0.725
problem	*a_problem*	0.731
computer_ component Hardware	*with_the _printer*	0.718

CL-R Label: *C1_Printer_Problem*; Confidence: 1

DM
1. Infer *subclass_C1_Printer_Problem*
2. Inferred Class == CL-R Label
3. CLASS LABEL ONLY
4. VERIFY Problem Class

SYSTEM: You have a problem with your printer. Do you confirm?

...

SYSTEM: Thank you, wait in line. An operator will assist you with your Lexmark printer problem!

Figure 1: Example dialogue translated to English

Conversational Agents

Observations

+ Interpretation

+ Uncertainty

+ Decisions

+ Interaction

= **Experience**

Evaluation

- **How good are machines?**
  - Accomplishing tasks
  - Acceptance by the users, by the network
  - Life Span (Evolve over time)
- "Research Systems are difficult to evaluate while Commercial systems are ea$y to evaluate " (Paek, 2007)

DEMOS

Conclusion

- There are Q&A machine tasks which will give (partial) benefits
- Machines may fulfill (social) roles
- **The ultimate goal is to understand the collaborative interaction of machine/human agents**
  - **Language**
  - **Interaction Model**
  - **Social Equation**