DEEP LANGUAGE UNDERSTANDING

JAMES ALLEN

THE UNIVERSITY OF ROCHESTER

- AND -

INSTITUTE FOR HUMAN AND MACHINE COGNITION
WHAT IS DEEP UNDERSTANDING?

Students develop deep understanding when they grasp the relatively complex relationships between the central concepts of a topic or discipline. Instead of being able to recite only fragmented pieces of information, they understand the topic in a relatively systematic, integrated or holistic way. As a result of their deep understanding, they can produce new knowledge by discovering relationships, solving problems, constructing explanations and drawing conclusions.

Students have only shallow understanding when they do not or cannot use knowledge to make clear distinctions, present arguments, solve problems or develop more complex understanding of other related phenomena.

DEPT. OF EDUCATION, QUEENSLAND

IN OTHER WORDS,
CONNECTING LANGUAGE TO OTHER COGNITIVE ABILITIES:
KNOWLEDGE, REASONING, ACTION, LEARNING, ...

SAME WITH MACHINES - DEEP UNDERSTANDING PRODUCES MEANING THAT IS USABLE FOR MULTIPLE TASKS, INCLUDING REASONING & EXPLANATION
SHALLOW NLP HAS HAD GREAT SUCCESS

- **"TAGGING" TECHNIQUES**
  - E.G., PART OF SPEECH TAGGING, SENTIMENT ANALYSIS, EVENT DETECTION, DISCOURSE ACTS, WORD SENSES, SEMANTIC ROLES, ....
  - PREDOMINANTLY TRAINED OFF LABELLED DATA

- **"STRUCTURE" TECHNIQUES**
  - E.G., PARSING, CFGS, CCGS, DEPENDENCIES, ...
  - PREDOMINANTLY TRAINED OFF LABELLED DATA

- **"CORRESPONDENCE" TECHNIQUES**
  - E.G., MACHINE TRANSLATION
  - PREDOMINANTLY TRAINED OFF PARALLEL CORPORA

- **"HAND-ENGINEERED" TECHNIQUES**
  - E.G., MOST DIALOGUE SYSTEMS, CONVERSATIONAL ASSISTANTS
  - DRIVEN BY PATTERN MATCHING, WITH SOME SHALLOW PROCESSING
EXAMPLES OF NEED FOR DEEP(ER) METHODS

- CONVERSATIONAL SYSTEMS - SOLVING PROBLEMS THAT REQUIRE EXTENDED INTERACTION
  - E.G., INTERACTIVE PLANNING AND SCHEDULING, PERSONAL ASSISTANTS, HEALTH ADVISORS, TECH SUPPORT, ...

- BETTER MACHINE TRANSLATION
  - ABSTRACT MEANING REPRESENTATION, USING SEMANTIC ROLE MODELS, ...

- LEARNING BY READING
  - ACQUIRING COMMONSENSE KNOWLEDGE TO SUPPORT NEXT GENERATION OF INTELLIGENT SYSTEMS
LEARNING SEMANTIC PARSERS

- **ANNOTATE DESIRED “LOGICAL FORM” FOR A SET OF SENTENCES**
  - E.G., ROBOT NAVIGATION COMMANDS (TURN LEFT, GO FORWARD, ...), DATABASE QUERIES
  - BY HAND, OR REVERSE ENGINEERING USING AMAZON TURK

- **USE AN EXISTING SYNTACTIC PARSER (E.G., CFG, CCG)**
  - CREATES A SET OF PARSE TREE/LOGICAL FORM PAIRS

- **LEARN STRUCTURAL CORRESPONDENCES BETWEEN SYNTAX AND LOGICAL FORM**
  - E.G., WITHIN CCG FRAMEWORK, THEY LEARN LEXICAL ENTRIES FOR WORDS USED

E.G. MATUSZEK(2012), ZETTLEMOYER & COLLINS(2007), TELLEX (2012), ...
STATUS OF SEMANTIC PARSERS

- PROMISING RESULTS IN SIMPLE DOMAINS
  - E.G., A SMALL SET OF ROBOT NAVIGATION COMMANDS
  - LANGUAGE IS RESTRICTED TO SIMPLE COMMANDS

- BUT NO PORTABILITY
  - LEXICON LEARNED IN ONE ROBOT DOMAIN IS NOT USABLE IN ANOTHER, LET ALONE IN A DATABASE QUERY DOMAIN

- UNCLEAR IF GENERALIZES TO MORE COMPLEX DOMAINS
  - CREATING THE LOGICAL FORMS IS INCREASINGLY DIFFICULT
  - LANGUAGE LOSES ITS ONE-TO-ONE CORRESPONDENCE WITH MEANING
WHY NATURAL DOMAINS ARE DIFFICULT

At a grocery store ...
Customer: *black beans?*
clerk: *aisle 3.*

**BUT IN A HOME ENVIRONMENT...**
When arriving home ...
Spouse: *black beans?*
You: *Oh, sorry, I forget to get them.*

When exploring nutrition options ...
Spouse: *black beans?*
You: *227 calories in a cup*

A SEMANTIC PARSER COULD LEARN THAT
“BLACK BEANS” PARSES TO DATABASE QUERY
[LOCATION “BLACK BEANS” ?X]

DEEP UNDERSTANDING REQUIRES INTENTION RECOGNITION IN CONTEXT
When cooking ...
Spouse: *black beans?*
You: *in the cupboard.*

When cooking (adding black beans to a pot) ...
Spouse: *black beans?*
You: *don’t you like them.*
DEEP(ER) UNDERSTANDING AND INTENTION RECOGNITION

☐ THE PROBLEM

☐ LANGUAGE LOSES ITS ONE-TO-ONE CORRESPONDENCE TO INTENDED MEANING

☐ INTENDED MEANING IS INTRICATELY TIED TO THE GOALS AND INTENTIONS OF THE SPEAKER/WRITER

☐ HOW TO MAKE PROGRESS ON SUCH A COMPLEX ISSUE?

LANGUAGE \[\rightarrow\] SEMANTIC PARSING \[\rightarrow\] CONTEXT INDEPENDENT LOGICAL FORM \[\rightarrow\] INTENTION RECOGNITION \[\rightarrow\] INTENDED MEANING IN CONTEXT

A PRACTICAL MIDDLE GROUND?
PRACTICAL BROAD-COVERAGE SEMANTIC PARSING
THE TRIPS LOGICAL FORM

TRIPS Parser Web Interface

TRIPS ONTOLOGY TYPE
LEXICAL ITEM/WORD SENSE
SEMANTIC ROLES
STRUCTURAL/SCOPING LINKS

WWW.CS.ROCHESTER.EDU/RESEARCH/CISD/PROJECTS/TRIPS/PARSER/CGI/WEB-PARSER-XML.CGI
TRIPS Semantic Parser

“Let me show you how to find a restaurant”

Hand-built Lexicalized Grammar of English

Compiled Lexicon: patterns with selectional restrictions/preferences

Hand-built Core Deep Semantic Lexicon ~7000 lemmas

Chart Parser

Generic ~3000 concept ontology

LF graph
"Advanced Medical paid $106 million in cash for its share in a unit of Henley’s Fisher Scientific subsidiary."
Uses the same intuition as and VerbNet:

**the subcategorization structures a verb takes reflect its semantic meaning**

**Ontology Mapping**

- change%2:30:01 ➔ ONT::MODIFY
- decrease%2:30:00 ➔ ONT::DECREASE
- weaken%2:30:01 ➔ ONT::DECREASE
- weaken%2:30:00 ➔ ONT::MODIFY

**Lexicon Lookup**

- “alter” ➔ AGENT-AFFECTED-RESULT(to)-optional
- “modify” ➔ AGENT-AFFECTED
- “modulate” ➔ AGENT-AFFECTED
- “compress” ➔ AGENT-AFFECTED
- “constrict” ➔ AFFECTED
- “decrease” ➔ AGENT-AFFECTED-RESULT(to)-optional
- “decrease” ➔ AFFECTED-FORMAL(in)

**New Lexical Entries**

- ONT::MODIFY AGENT-AFFECTED-RESULT(PP to)
- ONT::MODIFY AGENT-AFFECTED
- ONT::DECREASE AGENT-AFFECTED-RESULT(PP to)
- ONT::DECREASE AGENT-AFFECTED
- ONT::DECREASE AFFECTED-FORMAL(PP in with)
New Lexical Entries

- Y attenuates
- Y attenuates in weight
- X attenuates Y to Z
- X attenuates Y

New Lexical Entries

- "attenuate"
- "attenuate" in weight
- "attenuate" X to Y
- "attenuate" X Y

New Lexical Entries

- ONT::DECREASE AGENT-AFFECTED-RESULT(PP to)
- ONT::DECREASE AGENT-AFFECTED
- ONT::DECREASE AGENT-AFFECTED-RESULT(PP to)
- ONT::DECREASE AGENT-AFFECTED
- ONT::DECREASE AFFECTED-FORMAL(PP in with)
Robustness: Heuristic Extraction

- Find least-cost path from start to end
  - where cost depends on type of constituent, score assigned by the parsing, ...
- Assemble logical forms from fragments
- Try to connect using discourse processing

<table>
<thead>
<tr>
<th>CP</th>
<th>ADVBL</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>PP</td>
</tr>
<tr>
<td>NP</td>
<td>VP</td>
</tr>
<tr>
<td>VP</td>
<td>NP</td>
</tr>
</tbody>
</table>

He ran and to the store

(SPEECHACT SA_TELL) (F (:* CONJUNCT AND)) (F (:* TO-LOC TO))

:CONTENT (F (:* SELF-LOCOMOTE RUN)) :OF V106896 (IMPRO (<OR> REFERENTIAL-SEM))

:AGENT (PRO (:* PERSON HE))
:TENSE W::PAST

:VAL (THE (:* COMMERCIAL-FACTORY STORE))

Coref
PRELIMINARY EVALUATION
(STEP 2008)

- Six paragraphs submitted by six different research groups
- Gold standard built by hand and verified, including word senses (relative to TRIPS ontology), and semantic roles
- But not “deeper” content, including coreference, temporal relations, etc

<table>
<thead>
<tr>
<th></th>
<th>Baseline TRIPS System</th>
<th>plus NER, POS, and WordNet lookup</th>
<th>plus advice from Stanford Parser</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prec.</td>
<td>61.40%</td>
<td>78.2%</td>
<td>79.0%</td>
</tr>
<tr>
<td>Recall</td>
<td>67.20%</td>
<td>82.6%</td>
<td>82.8%</td>
</tr>
<tr>
<td>F-score</td>
<td>64.2%</td>
<td>80.3%</td>
<td>80.9%</td>
</tr>
</tbody>
</table>

Amid the tightly packed row houses of North Philadelphia, a pioneering urban farm is providing fresh local food for a community that often lacks it, and making money in the process. Greensgrow, a one-acre plot of raised beds and greenhouses on the site of a former steel-galvanizing factory, is turning a profit by selling its own vegetables and herbs as well as a range of produce from local growers, and by running a nursery selling plants and seedlings. The farm earned about $10,000 on revenue of $450,000 in 2007, and hopes to make a profit of 5 percent on $650,000 in revenue in this, its 10th year, so it can open another operation elsewhere in Philadelphia.
## Connecting Language, Reasoning and Action

<table>
<thead>
<tr>
<th>Planning</th>
<th>Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sam’s Goal: Eat a peach</td>
<td>Sam’s action: “Do you know where I can find a peach?”</td>
</tr>
<tr>
<td>Sam’s subgoal: Know where a peach is</td>
<td>Sam’s plan: Ask Sue where a peach is</td>
</tr>
</tbody>
</table>

### Goal Formation

<table>
<thead>
<tr>
<th>Planning</th>
<th>Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sue infers: Sam wants to eat a peach</td>
<td>Sue hears: “Do you know where I can find a peach?”</td>
</tr>
<tr>
<td>Sue infers: Sam wants to know where a peach is</td>
<td>Sue hears: “Do you know where I can find a peach?”</td>
</tr>
</tbody>
</table>

### Plan/Intention Recognition

<table>
<thead>
<tr>
<th>Goal Recognition</th>
<th>Goal Recognition</th>
<th>Surface Intention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sue infers: Sam wants to eat a peach</td>
<td>Sue infers: Sam wants to know where a peach is</td>
<td>Sue hears: “Do you know where I can find a peach?”</td>
</tr>
</tbody>
</table>
BALANCING DEEP AND BROAD: THREE EXAMPLES

- DEEP AND NARROW
  - TRIPS: INTERACTIVE PLANNING

- NOT-SO-DEEP BUT BROAD
  - PLOW: LEARNING EXECUTABLE TASKS IN WEB BROWSERS FROM DEMONSTRATION

- DEEP AND BROAD
  - FIRST STEPS: A BROAD-COVERAGE DEEP SEMANTIC LEXICON
THE DEEP DIALOGUE AS COLLABORATIVE PROBLEM SOLVING

COMMUNICATION ACTS
E.G., SUGGEST AN ACTION,
ASK A QUESTION

INTERACTION LEVEL
E.G., INITIATE ADOPT-GOAL,
REJECT ADOPT-RESOURCE

COLLABORATIVE PROBLEM SOLVING
E.G., ADOPT A GOAL,
EVALUATE A SOLUTION

TASK LEVEL
E.G., EVACUATE AN ISLAND,
DRIVE BUS, ...

Figure 1: Collaborative problem solving model
DIALOGUE AS COLLABORATIVE PROBLEM SOLVING

Human-Machine Communication driven by models of Collaboration

**Collaborative Planning**

**TRIPS** engaged in spoken dialogue to collaboratively develop an evacuation plan in an island world

**Execution & Coordination**

**LOU** engaged in spoken dialogue to coordinate activities between human and robot team members

**Advising**

**CHESTER** engaged in spoken dialogue to help advise a patient in medication management

**Task Learning**

**PLOW** learned executable tasks in a web browser by watching demonstrations with commentary
TRIPS

- TASK: CREATE AN EVACUATION PLAN
  - REMOVE ALL PERSONNEL FROM THE ISLAND OF PACIFICA WITHIN TEN HOURS

- MINIMALLY TRAINED HUMAN USERS
  - BRIEF INTRODUCTION TO THE TASK, IMPORTANCE OF ROUTING TO MEET DEADLINES
  - NO INSTRUCTION ON WHAT TO SAY

- DEEP BUT NARROW
  - REQUIRES COMPLEX PLANNING & TEMPORAL REASONING
  - VOCABULARY LIMITED TO MOTION, LOADING, VEHICLES, ROUTES, PLANNING, ETC (~1000 WORDS)
THE NEED FOR INTENTION RECOGNITION

- THE CONTEXT
- GOAL: EVACUATE-ISLAND
- SUBGOAL ESTABLISHED:
  - ABYSS->DELTA
- SOLUTION DEVELOPED:
  - USING A TRUCK TO MOVE
  - PEOPLE FROM ABYSS TO DELTA.

- UTTERANCE: “CAN WE USE A HELICOPTER TO GET THE PEOPLE AT ABYSS?”

A REQUEST TO MODIFY THE SOLUTION FOR ABYSS--->DELTA (USE THE TRUCK INSTEAD OF THE HELICOPTER)
A PROPOSAL OF A NEW SUBGOAL TO EVACUATE BARNACLE BY HELICOPTER
PLOW*

- Teach your computer to perform tasks on the web by demonstrating them once
- Describing your actions as you do them
- Minimally trained human users
- PLOW can learn an executable task in just one structured demonstration

- Somewhat deep & broad
  - Collaborative problem solving model
  - But does not require complex reasoning, just ontological reasoning
  - Unlimited vocabulary
  - Can talk about whatever is on the web!

*Best Paper, AAAI 2007
Plow as an Intelligent Assistant

User Instruction
e.g., go to Amazon dot com

User Actions
e.g., opens URI

Identification of Intent

Query about collab. problem solving state

Intended Actions

Collaborative Agent
(decides what CALO should do next)

Query about tasks
what can happen next?

Knowledge Base

New steps, examples, procedures, corrections, ...

PLOW Execution Capabilities

Task Invocation

Task Learning

Collaborative Execution and Learning
### PLOW: Integrating Language and Learning

#### THE NAME OF THE RESTAURANT

<table>
<thead>
<tr>
<th>Restaurant</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philly's Cheesesteaks &amp; Hoagie</td>
<td>0.56 miles</td>
</tr>
<tr>
<td>Pizza Hut</td>
<td>0.56 miles</td>
</tr>
<tr>
<td>Culinary Production</td>
<td>0.72 miles</td>
</tr>
</tbody>
</table>

#### Additional Information

- Map
- Directions To
- Directions From

### Let's Practice

**USR:** <sil> LET'S PRACTICE <sil>  
**SYS:** What is the state the address and the city  
**SYS:** Ok  
**SYS:** I am finished
Let me show you how to find hotels near an address.

**Output parameter:** List of hotels

**Input parameter:** Address

Put the zip code here.

Here is the list of hotels.

The *zipcode role of *ADDRESS

The output parameter
User says “put the name here”
- Parses to an action of type *FILL with :object role of type *NAME
User types “Harold Pinter” in a text field
- Observe action FILL-FIELD with :object role INPUT22 and :value role “Harold Pinter”
How Well Does PLOW Work?

**TRAINING PHASE**

PLOW

Training
- system instruction
- multiple practice examples

16 subjects

**TEACHING**

Teaching
- Users have a day to define the new task models using whichever system they want

10 new task specifications

**EVALUATION PHASE**

System #1

System #2

System #3

Testing
- Learned task models run on test inputs
- Performance scored based on grading scheme

3 test inputs per task

Learned Task Models

Grading scheme per task
PLOW used to define 55% of the tasks (30 out of 55)
PLOW was used successfully by most number of subjects (12 out of 16)
PLOW received highest average grade (2.2/4)
SELECTED REFERENCES