KNOWLEDGE REPRESENTATION AND PROCESS IN NLP

Theme

Background on Knowledge Representation, as relates to NLP: formalism and framework. Language closely mirrors representation (the right representation helps). There are several layers of representation of a text, including syntax, semantics, discourse, information structure, pragmatics. Here we focus on semantics. First: KR is two separate areas: content and system. Types of content theories: naive semantic theories (Conceptual Dependency Theory, Conceptual Graphs, Dahlgren’s naive semantics, Hobbs’s ontological promiscuity) and types of systems and logics. Following this, the nature of NLP processes: Parsing is essentially the process of inference to the best explanation, and generation is essentially planning under constraints of different types.

Summary of Contents

1. Levels of Language Representations

Each text consists of sentences, and each sentence of words, and many words of various parts. A full representation of the contents and structure of a text has to include many different aspects, including but not limited to syntax (grammatical structure) and semantics (meaning). Here are the names of NLP processes to identify some of the more important levels:

- Lemmatizing → demorphed, individuated word tokens “didn’t” → “did” + “not”
- Tagging → part-of-speech (POS) tags (“noun”, “verb”, etc.)
- Syntactic parsing → parse trees (many theories of grammar)
- Semantic analysis → semantic readings (many theories of semantics)
- Discourse analysis → discourse structures (trees?) → cohesiveness graphs (lexical chains) → Information Structure (theme/rheme, given/new, topic/focus)
- Pragmatic analysis → Speech Acts; Locution/Illocution/Perlocution → Interpersonal and Textual metafunctions (Systemic grammar)

In this lecture we focus on one of the most central—semantics.

2. Syntax

Object of analysis: sentence.

Form: tree structure. The leaves are words of the sentence; internal nodes are syntactic classes such as Noun Phrase and sentence (the top node).

Important issues: ID/LP (immediate dominance / linear precedence): Which nodes are dominated by a node (their mother)? What is the left-to-right order of sister nodes?

Theories in NL computation: HPSG, GPSG, and many others.

Uses in CL: a step toward semantics. Useful in cases where semantics is not unique (“the boy sees the woman” is not the same as “the woman sees the boy”; however syntax is not particularly useful when semantics is clear: “the car sees the boy” in normal cases will mean “the car is what the boy sees”).
3. Information Structure

Object of analysis: sentence.

Form: segmentation of the sentence.

Important issues: at least three different and often ill-defined notions studied here:

- **Theme-rheme**: theme indicates what the sentence is ‘about’; it usually points backward into the discourse to a previously mentioned item that is now being further discussed. Usually in English theme is the first NP. Rheme is everything that is not theme.

- **Topic-focus**: focus is the most interesting and important new material said about the theme. Usually in English it comes toward the end of the sentence, with the importance building up (in well-written sentences) and the most important thing (the focus) last. Topic is everything that is not focus.

- **Given-new**: given is all the material in the sentence that has already been said, and new is all the rest. What ‘been said’ means is not clearly defined: it could mean literally or it could mean implicitly implied.

Theories in NL computation: essentially none. The Prague School view is probably the closest.

Uses in CL: possibly useful in dialogue, but usually a rudimentary approximation of focus is enough.

Example:

**Temporal perspective**: Last week I traveled around Europe. On Tuesday, I visited Rome, and saw the Coliseum. On Thursday, I was in Paris and went up the Eiffel Tower. On Saturday in London I saw Buckingham Palace. (theme: time PP; focus: monument)

**Spatial perspective**: I traveled around Europe last week. In Rome, I saw the Coliseum on Tuesday. In Paris on Thursday I went up the Eiffel Tower. In London (on Saturday) I saw Buckingham Palace. (theme: location PP; focus: monument)

**Monument perspective**: I saw wonderful sights in Europe last week. The Coliseum, in Rome on Tuesday, was grand. The Eiffel Tower, in Paris on Thursday, had me gasping for breath. But Buckingham Palace, in London (on Saturday) was the most impressive thing I saw. (theme: monument; focus: level of impressiveness)

4. Discourse Structure

Object of analysis: whole text.

Form: tree(-like) structure over the text, usually with sentences or clauses as leaves.

Important issues: What are the names of the internal nodes (i.e., the relations between discourse segments)? How many families of relations are there? Are they simply semantic relations or are they something else? How large are the leaf nodes—sentences or clauses—and what happens when a clause gets demoted to a single word (the book that is green → the green book)?

Theories in NL computation: RST (Mann and Thompson), DRT (Kamp and Reyle; Asher and Lascarides), and some others (Grosz and Sidner; Polanyi).

Uses in CL: important for planning coherent multisentence text. Useful for text summarization.
5. Pragmatics

Object of analysis: sentence or whole text.
Form: annotations on sentences or on whole text.

Important issues: There seem to be two disjoint sets of concern. The first, which has been more thoroughly studied, focuses on the connotations of extra-sentence meaning that a sentence may carry, such as “do you know the time?” really meaning “please tell me the time”. The Speech Acts of Austin and Searle include
- locutions (normal sentences)
- illocutions (“do you know the time?” really means “please tell me the time”)
- perlocutions (when the sentence itself makes the act, as in “I hereby marry you” or “I promise that…”)

The second is especially important in multisentence text generation and dialogue processing, and involves interpersonal and situational factors.

Interpersonal factors include speaker-hearer friendship level, level of acquaintance, relative social status, speaker friendliness, etc. Situational factors involve the surroundings of the communication (medium, noise level, level of formality, etc.). These factors are reflected in the style of the language produced and send signals about the speaker, hearer and situation.

Theories in NL computation: no real theories yet. Speech Acts and formal pragmatics.
Uses in CL: Style issues important for sophisticated text generation.

6. Semantics: Two Aspects of Representing Knowledge

The content and organization of knowledge plays a central role in language processing. Two classes of issues studied: representation content (symbol systems, ontologies) and representation systems and reasoning (inference frameworks and axioms). The AI subfield of Knowledge Representation (KR) has worked with semantics, focusing much attention on systems and reasoning.

In both aspects there are widely different approaches. One can represent them as follows:

<table>
<thead>
<tr>
<th>Content</th>
<th>Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>parsimonious</td>
<td>promiscuous</td>
</tr>
<tr>
<td>primitives</td>
<td>plenty</td>
</tr>
<tr>
<td>CD Dahlgren Sowa Hobbs</td>
<td>FRL KEE KL-ONE, Loom</td>
</tr>
</tbody>
</table>

6.1 Content

There are two principal styles: Parsimonious (choosing just a few primitive concepts and composing them into many meanings) and Profligate (choosing many different concepts, each represented independently). Both approaches have strengths and weaknesses.

The Parsimonious Approach

Represent a set of example sentences, first in Conceptual Dependency (CD) (Schank and Abelson, 77) and then using open-ended standard representation predicates:
Mary gives John a book
Mary gave John the book
Mary sells John a book
John buys a book from Mary
Mary read John a story
Mary gave John a kiss
Mary gave John a black eye
Mary saw a plane flying to New York
Mary saw the Grand Canyon flying to New York

John got a book
John got better
John was healthy
to illustrate PTRANS, MTRANS, STATEs, STATE-CHANGEs, ENABLE, RESULT, etc. The idea is to try to represent the meanings of these sentences, by capturing the meanings of their words and composing them appropriately. Notice that you need for this composition both the meaning symbols and the relations that link them.

Some typical case relations for processes:
- agent, patient (= object), instrument, beneficiary,
- location, source, destination,
- manner,
- cause/result, temporal-sequence, enable…

and for objects:
- name, size, age, location, nationality, color, weight…

and for states:
- experiencer, state, degree

The Parsimonious approach is nice, because your system can know that “buy” and “sell”, or “come” and “go”, are more or less the same thing. But you pay a price: you have to decompose complex concepts into these primitives! How do “Democracy” or “Love” decompose? You generally end up unable to build a real-world domain model.

The Profligate Approach
On the other hand, the Promiscuous (profligate) approach is nice too, because your work in representing is much easier. But you pay a price here too: your system cannot know the relatedness of closely related concepts: “come” and “go”, or “price” and “buy” and “buyer”. Also you have no guidance on what primitives to create, or why. If you build a domain model of more than 1500 concepts you generally end up with a bit of a mess, and have trouble enforcing consistency.

To help with consistency and concept organization, people often build Ontologies—taxonomies of the symbols, arranged in increasing specificity. This is the subject of a whole course, and is a fascinating mixture of philosophy, linguistics, and psychology. Near the top level you need to differentiate (at least): Actions/Events, States, State-Changes, Objects, Relations, Qualities. And it goes on from there.

A problem in the Profligate approach is to determine how many concepts there are — that is, how to ‘chop up’ the continuum of meaning fields into discrete concepts. The recently developed OntoNotes procedure that uses word senses as a ‘bridge’ is instructive. Exercise: how many word
senses do the following sentences contain? How are the senses taxonomized, from least to most similar?

1. Drive the demons out of her and teach her to stay away from my husband!!
2. Shortly before nine I drove my jalopy to the street facing the Lake and parked the car in shadows.
3. He drove carefully in the direction of the brief tour they had taken earlier.
4. Her scream split up the silence of the car, accompanied by the rattling of the freight, and then Cappy came off the floor, his legs driving him hard.
5. With an untrained local labor pool, many experts believe, that policy could drive businesses from the city.
6. Treasury Undersecretary David Mulford defended the Treasury’s efforts this fall to drive down the value of the dollar.
7. Even today range riders will come upon mummified bodies of men who attempted nothing more difficult than a twenty-mile hike and slowly lost direction, were tortured by the heat, driven mad by the constant and unfulfilled promise of the landscape, and who finally died.
8. Cows were kept in backyard barns, and boys were hired to drive them to and from the pasture on the edge of town.
9. He had to drive the hammer really hard to get the nail into that plank!
10. She learned to drive a bulldozer from her uncle, who was a road maker.
11. I used to drive a taxi (for work) before I went to night school.
12. Beware—Ralph drives a hard bargain; you will probably lose all your money.
6.2 KR Formalisms and Representation Mechanisms

What functionalities do you want your KR system to have?

1. Representation of information
   Associate bits of info together; build frames/assertion packets/etc.

2. Consistency maintenance
   Do you want this? Automatic checking that all is well. Humans aren’t consistent; why should a system be? The problems of inconsistency

3. Inheritance
   Why do you want this?—for saving space against redundancy. Tweety the Bird and Oliver Ostrich. Strong vs. weak property inheritance. Classification only possible in strong inheritance

4. Defaults
   The Nixon diamond. What to do?—disallow/allow anything/ask user

5. Inference
   Rules defined in T-Box hold on A-Box entities; automatic inference (classification an example). Forward and backward-chaining

As discussed above, there’s no “best” solution.

The range here goes from simple free property-inheritance schemes like KRL (Bobrow & Winograd 77) or FRL (Roberts & Goldstein 77) to highly structured ones like KL-ONE (Brachman ~67) and Loom (MacGregor 89). Examples in FRL:

E1 := (make-new-frame
   :isa EAT
   :actor JOHN1
   :object (make-new-frame :isa EGG :number 2 :prep SCAMBLED)
   :time Wednesday15))

Here inheritance fills in what is not given (:loc = ?) and what is given simply overrides the ISA’s value. Freedom but no structure or enforcement of coherence. In contrast, KL-ONE family enforces agreement to role-filler constraints, which makes the system much harder to deal represent things in but guarantees consistency and enables classification. Instantial and conceptual knowledge. A Loom example:

(defconcept Phys-Obj
   :constraints (and (:exactly 1 weight) (:exactly 1 location)))

(defconcept Fragile-Thing :is (:and Phys-Obj :primitive))

(tellm Fragile-Thing FT1)

Redundancy in doing all this. Property-inheritance. Frameworks in which to represent information. Terminological and Instantial knowledge bases. KL-ONE, NIKL, Loom, and Brachman vs FRL and KEE: strong and weak inheritance. Defaults and overruling, and strong property inheritance requirements and the classifier.

Tying the two KR parts together: an example (the Upper Model).

Predicate calculus. formula := (pred -args-); arg := const l var | fn applic (which is evaled); connective := and | or | not | if | iff | forall | exists; preds; funcs. Denotations: terms → individuals or
classes; formulas -> ? Not T/F because depends on interp of denotation. Hard question. Soudness (cannot prove non-T thing) and consistency (all statements eventually T).

Tweety is a bird. (bird Tw)
Tweety eats all good foods. (∀X) ((food X) and (good X) ⇒ (eat Tw X))
Sunflower seeds are good food. (food SS) and (good SS)
⇒ Some bird eats sunflower seeds. (∃y) (eat Y SS) and (bird Y) via (eat Tw SS)


**The most important lesson** here: for both representation content and representation system, there’s no “best” choice; every point has advantages and disadvantages. If you need inference, then choosing a more parsimonious content set is better—you have to select what’s appropriate for your task. If you’re doing just parsing or generation then choosing a more profligate representation is easier. In both cases, as soon as you start working with more than 1500 or so items that are very different in nature, you run into trouble.

7. **NLP Processes in Terms of KR Processes**

This part discusses the nature of the parsing and generation processes. Parsing is seen to be essentially the process of inference to the best explanation, and generation is seen to be a process of planning with increasingly grammatical operators.

**Parsing** and inference in memory. Deduction, induction, and abduction. Deduction example Socrates mortal; induction example; abduction example from Charniak “John got a rope. He wanted to commit suicide”. Inadequacy of deduction, problems with induction and abduction. Charniak and Hobbs for different strategies of abductive parsing, very briefly.

**Generation**: text planning and realization, as a planning problem. Appelt’s TELEGRAM idea.

**(Optional) Readings**

Chapter 14 in Jurafsky and Martin

Optional: Some CD (Schank and Abelson 77, or Schank 82) or Jackendoff (85)

Some formal semantics

Predicate Calculus: Appendix A from Allen, or chapter 6 from Charniak & McDermott

**Assignment**

None
Example 1. Newspaper Article

Partiality: unbiased
Formality: formal
Force: medium

In early April, a shantytown -- named Winnie Mandela City -- was constructed on Beinecke Plaza by several students, so that Yale University would divest from companies doing business in South Africa. At 5:30 am on April 14, it was destroyed by officials; also, at that time, the police arrested 76 students. The students requested that Yale give them permission to reassemble the shantytown while several local politicians and faculty members expressed criticism of Yale's action. Finally, the university permitted the students to reconstruct it and, concurrently, Yale University announced that a commission would go to South Africa in July to examine the system of apartheid.

Example 2. Protester Literature

Partiality: biased
Formality: formal
Force: medium

As a reminder to Yale University to divest from companies doing business in South Africa, a large number of concerned students erected a shantytown -- named Winnie Mandela City -- on Beinecke Plaza in early April. The local community expressed support for the students' action. The university told the students to erect the shantytown elsewhere. Later, at 5:30 am on April 14, the shantytown was destroyed by officials; also, at that time, the police arrested 76 students. The students requested that Yale University give them permission to reassemble it on Beinecke Plaza; also, at that time, several local politicians and faculty members expressed criticism of Yale's actions. Finally, Yale permitted them to reconstruct the shantytown and, concurrently, the university announced that a commission would go to South Africa to examine the system of apartheid in July.

Example 3. Official Yale Literature

Partiality: biased
Formality: formal
Force: medium

In early April, a small number of students were involved in a confrontation with Yale University over Yale's investment in companies doing business in South Africa. The students constructed a shantytown --- named Winnie Mandela City --- on Beinecke Plaza in order to force the university to divest from those companies. Yale requested that the students erect it elsewhere, but they refused to leave. The university intended to be reasonable. The university gave it permission to exist until the meeting of the Yale corporation, but even after that the students still refused to move. At 5:30 am on April 14, officials had to disassemble the shantytown. Finally, Yale, being conciliatory toward the students, not only permitted them to reconstruct it, but also announced that a commission would go to South Africa in July to examine the system of apartheid.

Example 4. Protester's Description on Street

Partiality: biased, medium
Formality: informal
Force: medium

I am angry about Yale's actions. The university had officials destroy a shantytown called Winnie Mandela City on Beinecke Plaza at 5:30 am on April 14. A lot of concerned students built it in early April. Not only did Yale have officials destroy it, but the police arrested 76 students. After the local community's huge outcry, the university allowed the students to put the shantytown up there again.