Natural Language Processing

Lecture 17

Today's Outline

Announcements (AI Topics, ISP)

Homework 3

Finish Word Sense Disambiguation

Begin Discourse Analysis

Reference Resolution

ISP Seminar

Nov 30 (note date change!): Antal van den Bosch (WhizBang! Labs-Research and Tilburg University, The Netherlands)

Current issues in machine learning of natural language: memory and representation

In machine learning, natural language processing tasks have attracted a markedly large amount of interest as applications, with mutual benefits. For machine learning, language data has been used for symbolic feature representation and classification research. For computational linguistics, machine learning has offered a way out of its knowledge acquisition bottleneck. This talk centers around two current issues relevant to both communities: memory and representation. As for memory, empirical evidence exists showing that retaining full memory of experienced language data in machine-learned language processing systems outperforms abstraction-based (minimal-description-length driven) learning, As for representation, there is a growing feeling that discovering good feature representations yields much shorter pathways to developing good machine-learned NLP systems than comparing machine learning algorithms. The talk will summarize empirical findings and open up discussion about their meaning to both scientific communities.

HW3

Due: December 3, 2001

- (5 pts) Simplify the following formulas using lambda reduction:
  - ((lambda x (P x)) A)
  - ((lambda x ((lambda y (P y)) x)) A)
- (15 pts) Exercise 15.4, Jurafsky and Martin (p. 586). Be sure to show a derivation to validate your answer.
- (20 pts) Exercise 16.6, Jurafsky and Martin (p. 629), but just for the verb "drink".
- (20 pts) Exercise 16.7, Jurafsky and Martin (p. 629). For your examples, use the American portion of the CoBuildDirect Corpus at http://titania.cobuild.collins.co.uk/form.html. Try queries such as drink@/VERB.
- (20 pts) Exercise 17.2, Jurafsky and Martin (p. 664). Use as your corpus sentences 16.1, 16.2, 16.3, 16.4, 16.7, and 17.11 from Chapters 16 and 17.
- (20 pts) Exercise 17.3, Jurafsky and Martin (p. 664).

Cue Phrase Sense Disambiguation

Sentential sense

- **Harry:** You know I see more coupons *now* than I've ever seen before and I'll bet you have too.

Discourse sense

- **Doris:** I have a couple quick questions about the income tax. The first one is my husband is retired and on social security and in '81 he ...few odd jobs for a friend uh around the property and uh he was reimbursed for that to the tune of about 640 dollars. *Now* where would he where would we put that on the form?

Both senses

- *Now now* that we have all been welcomed here it's time to get on with the business of the conference.

ML, using prosodic and orthographic features, can disambiguate (Litman 96)

The Problem of Discourse

Discourse

Monologue and dialogue (dialog)

Human-computer interaction

Examples

- *John* went to <u>Bill</u>'s <u>car dealership</u> to check out *an Acura Integra*. *He* looked at *it* for about half an hour.
- I'd like to get from Boston to San Francisco, on either December 5th or December 6th. It's okay if *it* stops in another city along the way.

Another Example

Gracie: Oh yeah ... and then Mr. and *Mrs. Jones* were having matrimonial trouble, and my brother was hired to watch *Mrs. Jones*.

George: Well, I imagine *she* was a very attractive woman.

Gracie: *She* was, and my brother watched *her* day and night for six months.

George: Well, what happened?

Gracie: *She* finally got a divorce.

George: *Mrs. Jones*?

Gracie: No, <u>my brother's wife</u>.

Reference Resolution

John went to Bill's car dealership to check out an Acura Integra. *He* looked at it for about half an hour.

Reference

- the process of associating *John* and *he* with a particular person

Referring expressions

- *John*, *he*

Referents

- the person named John

Coreferring referring expressions

- *John*, *he*

Antecedent

- *John*

Anaphor

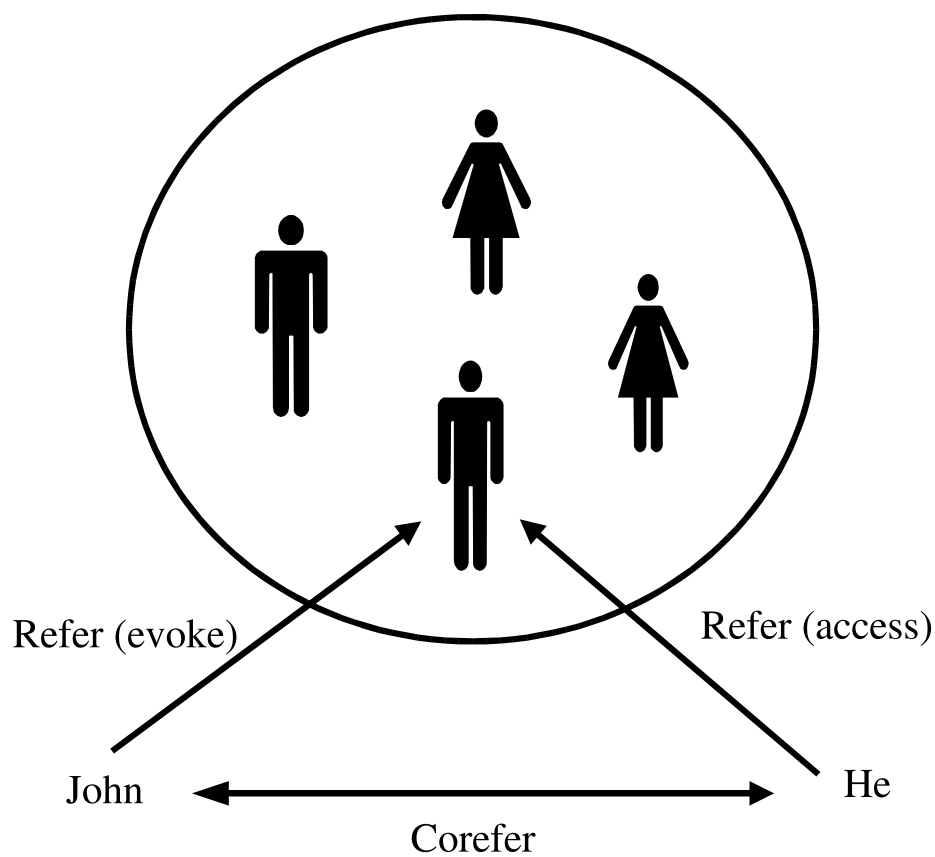
- *he*

Discourse Model

Needed because referring expressions (e.g. *it, this, that, this car, that car, the car, the Acura, the Integra, ...*) encode information about beliefs about the referent.

When a referent is first mentioned in a discourse, a representation for it is <u>evoked</u> into the model.

Upon subsequent mention, it is <u>accessed</u> from the model.



Many Types of Reference

Non-entity examples (from Webber 91)

According to John, Bob bought Sue an Integra, and Sue bought Fred a Legend.

- But *that* turned out to be a lie. (a speech act)
- But *that* was false. (proposition)
- *That* struck me as a funny way to describe the situation. (manner of description)
- *That* caused Sue to become rather poor. (event)
- *That* caused them both to become rather poor. (combination of several events)

Reference Phenomena

Indefinite noun phrases

- I saw *an Acura Integra* today.
- *some, this*

Definite noun phrases

- I saw an Acura Integra today. *The Integra* was white.
- *The Indianapolis 500* (kb)
- *The fastest car in ...* (unique)

Demonstratives

- *This Acura* is white.

One-anaphora

- I saw 6 Acuras today. Now I want *one*.

Reference Phenomena (cont.)

Pronouns

- I saw an Acura Integra today. *It* was white.
- compared to definite noun phrases, pronouns require more referent salience in the discourse model
- salience via recency
  - John went to Bob's party, and parked next to *a beautiful Acura Integra.*
  - He went inside and talked to Bob for more than an hour.
  - Bob told him that he recently got engaged.
  - ?? He also said that he bought *it* yesterday.
  - He also said that he bought *the Acura* yesterday.
- salience via structural recency
  - E: So you have the engine assembly finished. Now attach *the rope.* By the way, did you buy <u>the gas can</u> today?
  - A: Yes.
  - E: Did <u>it</u> cost much?
  - A: No.
  - E: OK, good. Have you got *it* attached yet?

Reference Phenomena (cont.)

Inferrables

- I almost bought an Acura Integra today, but *a door* had a dent and *the engine* seemed noisy.
- Mix the flour, butter, and water. Knead *the dough* until smooth and shiny.

Discontinuous sets

- *John* has an Acura and *Mary* has a Subaru. *They* drive them all the time.

Generics

- I saw 6 Acuras today. *They* are the coolest cars.

Constraints on Coreference

Number agreement

- John has *an Acura. It* is red.

Singular	Plural	Unspecified
she, her, he, him, his, it	we, us, they, them	you

Person and case agreement

- (⋆) *John and Mary* have <u>Acuras</u>. *We* love <u>them</u>. (where We=John and Mary)

	First	Second	Third
Nominative	I, we	you	he, she, they
Accusative	me, us	you	him, her, them
Genitive	my, our	your	his, her, their

Gender agreement

- John has an Acura. He/it/she is attractive.

masculine	feminine	nonpersonal
he, him, his	she, her	it

Constraints (continued)

Syntactic constraints

- John bought himself a new Acura. (himself = John)
- John bought him a new Acura (him = not John).

Selectional restrictions

- John parked *his Acura* in <u>the garage</u>. He had driven *it* around for hours.

Pronoun Interpretation Preferences

Recency

- John has an Integra. Bill has *a Legend*. Mary likes to drive *it*.

Grammatical role

- *John* went to the Acura dealership with Bill. *He* bought an Integra.
- *Bill* went to the Acura dealership with John. *He* bought an Integra.
- (?) John and **Bill** went to the Acura dealership. *He* bought an Integra.

Repeated mention

- *John* needed a car to go to his new job. *He* decided that *he* wanted something sporty. Bill went to the Acura dealership with *him*. *He* bought an Integra.
- Note: overrides grammatical role preference

Preferences (continued)

Parallelism

- Mary went with *Sue* to the Acura dealership. `Sally` went with *her* to the Mazda dealership.
- Mary went with *Sue* to the Acura dealership. `Sally` told her not to buy anything.

Verb semantics, thematic roles

- *John* telephoned Bill. *He* lost the pamphlet on Acuras.
- *John* criticized Bill. He lost the pamphlet on Acuras.

Pronoun Resolution Algorithm

Two steps

- discourse model update
- pronoun resolution

Salience values are introduced when a noun phrase that evokes a new entity is encountered.

These values are computed as a sum of weights assigned by salience factors (set empirically).

Salience Factors and Weights

From Lappin and Leass's system:

Sentence recency	100
Subject emphasis	80
Existential emphasis	70
Accusative (direct object) emphasis	50
Indirect object and oblique complement emphasis	40
Non-adverbial emphasis	50
Head noun emphasis	80

Recency

- weights are cut in half after each sentence is processed
- this, and a sentence recency weight, captures the recency preferences discussed earlier

Lappin and Leass (continued)

Grammatical role preference

- subject $>$ existential predicate nominal $>$ object $>$ indirect object or oblique $>$ demarcated adverbial PP

Examples

- *An Acura Integra* is parked in the lot. (subject)
- There is *an Acura Integra* parked in the lot. (existential predicate nominal)
- John parked *an Acura Integra* in the lot. (object)
- John gave *his Acura Integra* a bath. (indirect object)
- In *his Acura Integra*, John showed Susan his new CD player. (demarcated adverbial PP)

Head noun emphasis gives above examples 80 points, but following embedded NP nothing:

- The owner's manual for *an Acura Integra* is on John's desk.

Algorithm

1. Collect the potential referents (up to four sentences back).

2. Remove potential referents that do not agree in number or gender with the pronoun.

3. Remove potential referents that do not pass intrasentential syntactic coreference constraints.

4. Compute the total salience value of the referent by adding any applicable values for role parallelism (+35) or cataphora (-175) to the existing salience value.

5. Select the referent with the highest salience value. In case of a tie, select the closest referent in terms of string position.

Example

John saw a beautiful Acura Integra at the dealership last week. He showed it to Bob. He bought it.

See tables on pages 687-689 in Jurafsky and Martin.

Test Results

Lappin and Leass applied their algorithm to computer manuals and got 86% accuracy on unseen data.

Centering Algorithm (BFP)

Centering(Grosz, Joshi, and Weinstein): additional concept of a *center*.

Transitions

	$C_b(U_{n+1}) = C_b(U_n)$ or undefined $C_b(U_n)$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	Continue	Smooth-Shift
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	Retain	Rough-Shift

Rules

1. If any element of $C_f(U_n)$ is a pronoun in utterance U_{n+1}, then $C_b(U_{n+1})$ must be a pronoun.
2. Continue is preferred to Retain is preferred to Smooth-Shift is preferred to Rough-Shift.

Algorithm

1. Generate $C_b - C_f$ combinations for all reference assignments.
2. Filter by constraints.
3. Rank by transition orderings.

Trace

John saw a beautiful Acura Integra at the dealership. *He* showed `it` to Bob. He bought it.

Grammatical role hierarchy

- $C_f(U_1)$: { John, Integra, dealership }
- $C_p(U_1)$: John
- $C_b(U_1)$: undefined

Assume *it* refers to the Integra

- $C_f(U_2)$: { John, Integra, Bob }
- $C_p(U_2)$: John
- $C_b(U_2)$: John
- Continue ($C_p(U_2) = C_b(U_2); C_b(U_1)$ undefined)

Assume *it* refers to the dealership

- $C_f(U_2)$: { John, dealership, Bob }
- $C_p(U_2)$: John
- $C_b(U_2)$: John
- Continue ($C_p(U_2) = C_b(U_2); C_b(U_1)$ undefined)

Trace (continued)

Going with first option and assuming he refers to $John$ ...

- $C_f(U_3)$: { John, Acura }
- $C_p(U_3)$: John
- $C_b(U_3)$: John
- Continue $(C_p(U_3) = C_b(U_3) = C_b(U_2))$

Going with first option and assuming he refers to Bob ...

- $C_f(U_3)$: { Bob, Acura }
- $C_p(U_3)$: Bob
- $C_b(U_3)$: Bob
- Smooth-Shift ($C_p(U_3) = C_b(U_3)$ does NOT $= C_b(U_2)$)

Since Continue preferred over Smooth-Shift, algorithm picks John as as the referent.

Centering (continued)

Centering sometimes resolves a pronoun to something Lappin and Leass (or Hobbs, see text) would consider low salience.

Continue preference in centering mistakenly gives

- Bob opened up a new dealership last week. John took a look at the Acuras in his lot. He ended up buying one.

Centering is just now being specified enough to be automatically tested on actual data.

- The Role of Centering Theory's Rough-Shift in the Teaching and Evaluation of Writing Skills (Miltsakaki and Kukich, ACL 2000)
- Specifying the Parameters of Centering Theory: A Corpus-Based Evaluation using Text from Application-Oriented Domains (Poesio et al., ACL 2000)

For Next Time

Finish Chapter 18

Acknowledgements

Some of this material was taken from the course web pages of Drago Radev.