

# Computational Discourse

11-711 Algorithms for NLP

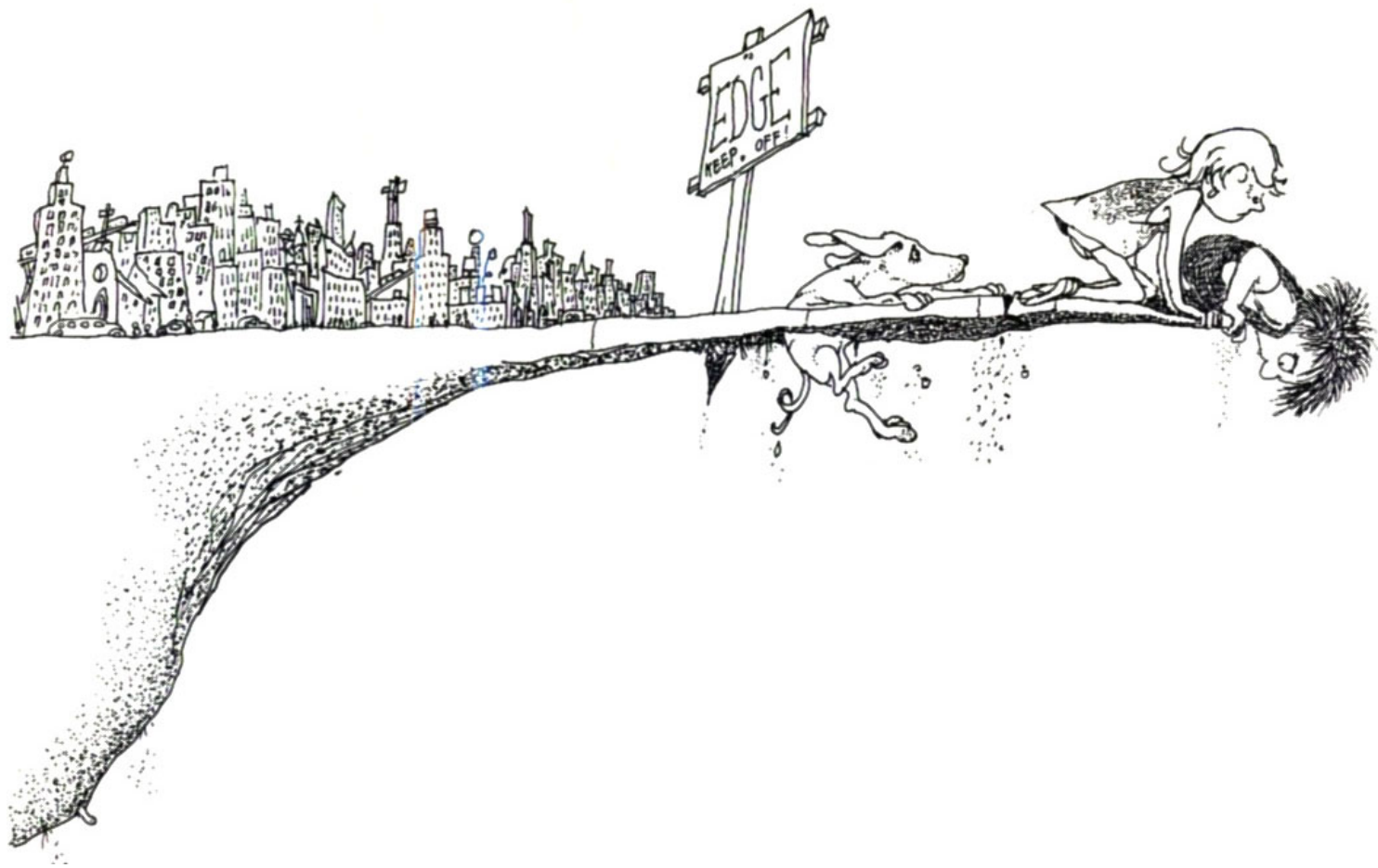
7 December 2017

# What Is Discourse?

**Discourse** is the coherent structure of language above the level of sentences or clauses. A **discourse** is a coherent structured group of sentences.

What makes a passage coherent?

A practical answer: It has meaningful connections between its utterances.



Cover of Shel Silverstein's *Where the Sidewalk Ends* (1974)

# Applications of Computational Discourse

- Automatic essay grading
- Automatic summarization
- Meeting understanding
- Dialogue systems

# Kinds of discourse analysis

- Discourse, monologue, dialogue, (conversation)
- Discourse (*SLP* Ch. 21) vs. (Spoken) Dialogue Systems (*SLP* Ch. 24)

# Discourse mechanisms vs. Coherence of thought

- “Longer-range” analysis (discourse) vs. “deeper” analysis (real semantics):
  - *John bought a car from Bill*
  - *Bill sold a car to John*
  - *They were both happy with the transaction*

# Coherence, Cohesion

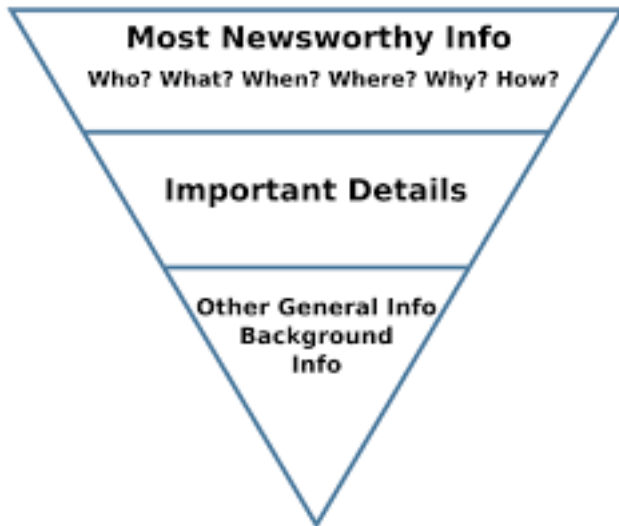
- Coherence relations:
  - *John hid Bill's car keys. He was drunk.*
  - *John hid Bill's car keys. He likes spinach.*
- Entity-based coherence (Centering) and lexical cohesion:
  - *John went to the store to buy a piano*
  - *He had gone to the store for many years*
  - *He was excited that he could finally afford a piano*
  - *He arrived just as the store was closing for the day*versus
  - *John went to the store to buy a piano*
  - *It was a store he had gone to for many years*
  - *He was excited that he could finally afford a piano*
  - *It was closing for the day just as John arrived*

# Cohesion in NLP



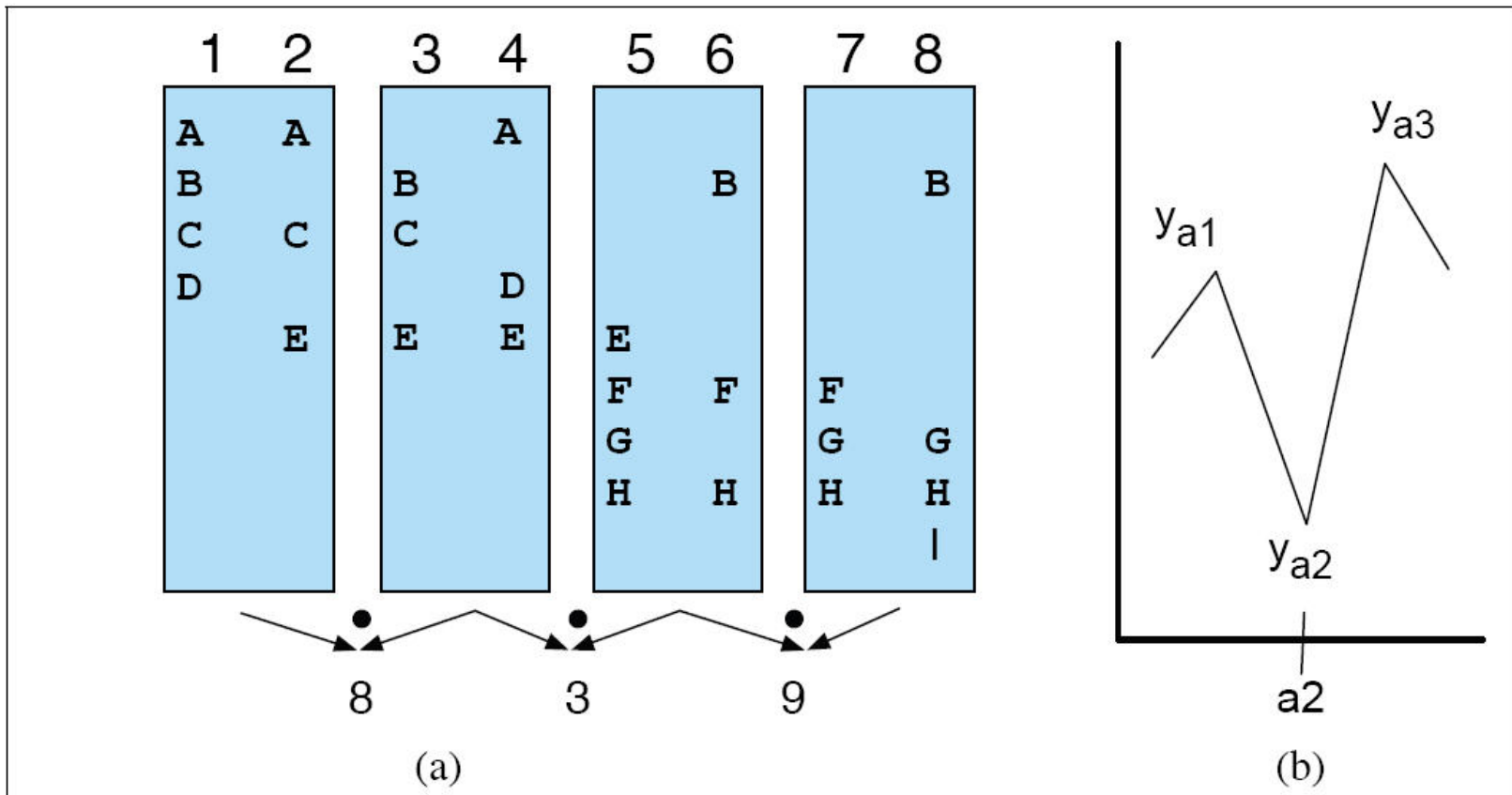
# Discourse Segmentation

Goal: Given raw text, separate a document into a linear sequence of subtopics.



- 1-3 Intro - the search for life in space
- 4-5 The moon's chemical composition
- 6-8 How early earth-moon proximity shaped the moon
- 9-12 How the moon helped life evolve on earth
- 13 Improbability of the earth-moon system
- 14-16 Binary/trinary star systems make life unlikely
- 17-18 The low probability of nonbinary/trinary systems
- 19-20 Properties of earth's sun that facilitate life
- 21 Summary

# Discourse segmentation: TextTiling



- Using dips in **cohesion** to segment text.

# Supervised Discourse Segmentation

Our **instances**: place markers between sentences (or paragraphs or clauses)

Our **labels**: yes (marker is a discourse boundary) or no (marker is not a discourse boundary)

What features should we use?

- Discourse markers or cue words
- Word overlap before/after boundary
- Number of coreference chains that cross boundary
- Others?

# Coherence in NLP

# Coherence Relations

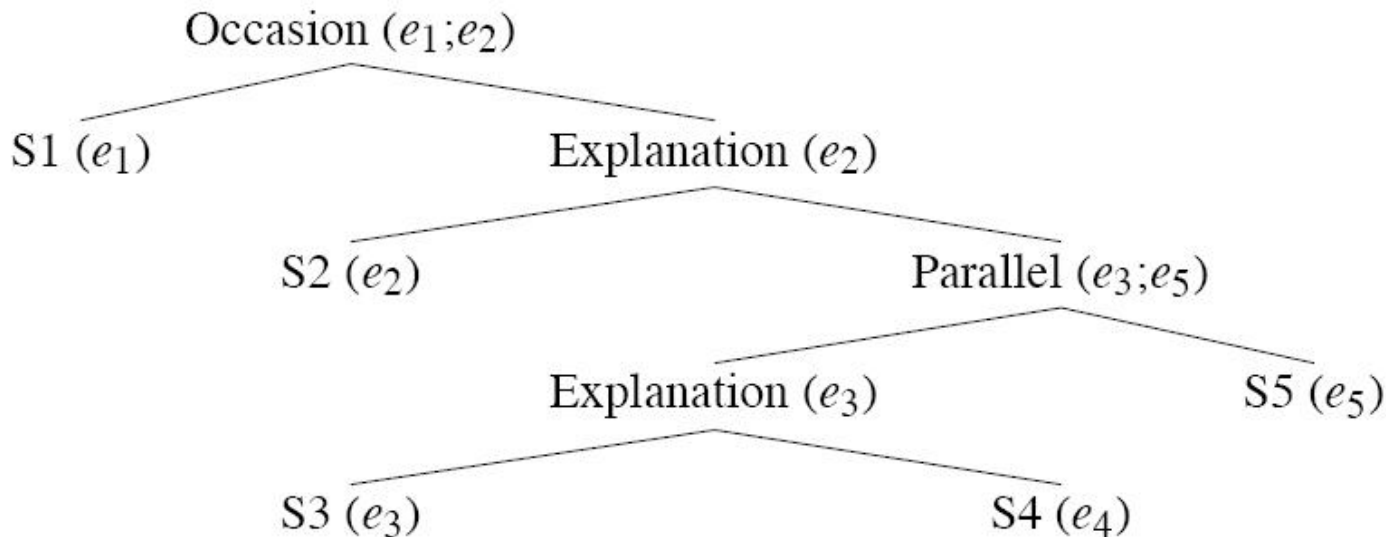
*S1: John went to the bank to deposit his paycheck*

*S2: He then took a bus to Bill's car dealership*

*S3: He needed to buy a car*

*S4: The company he works for now isn't near a bus line*

*S5: He also wanted to talk with Bill about their soccer league*

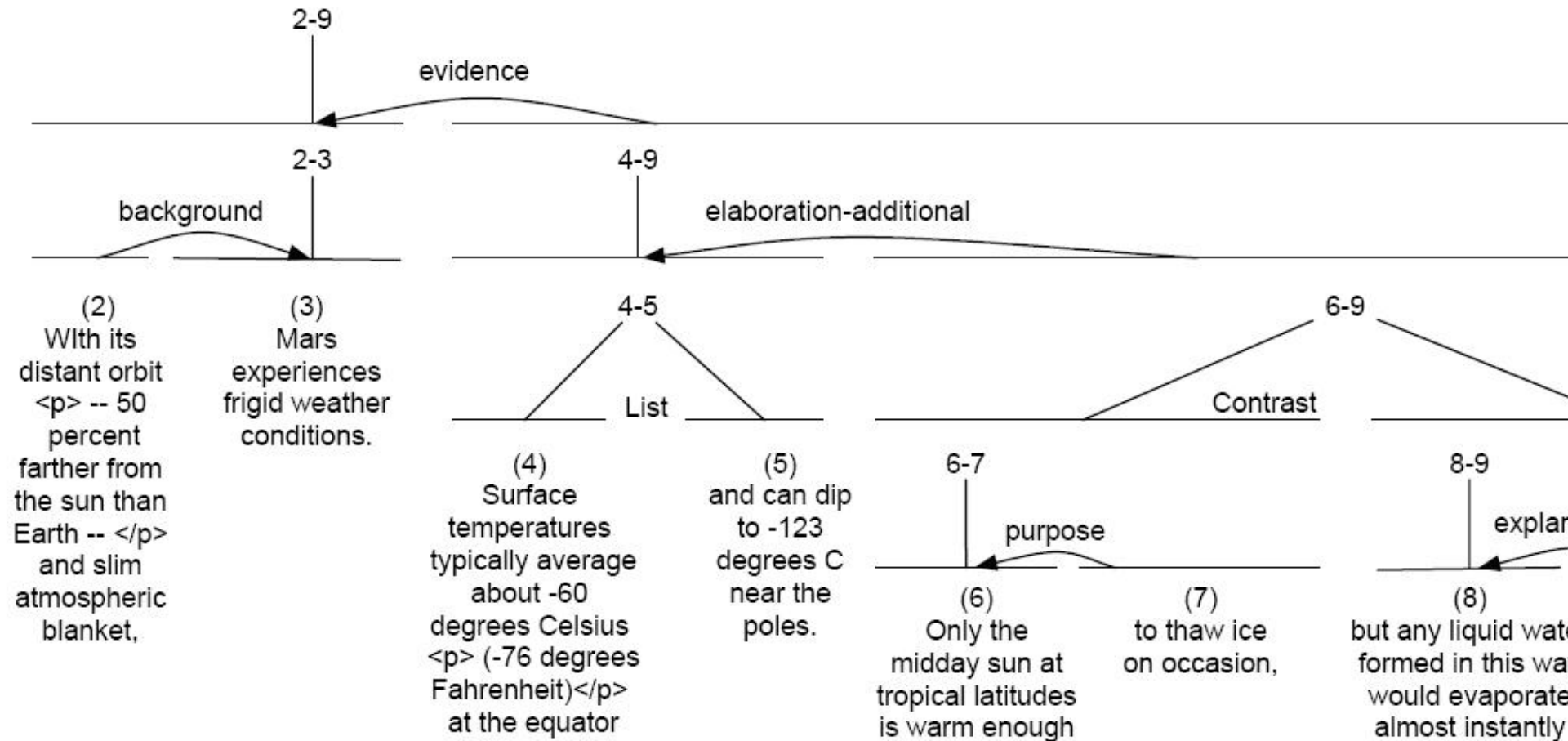


# Some Coherence Relations

How can we label the relationships between utterances in a discourse? A few examples:

- **Explanation:** Infer that the state or event asserted by  $S_1$  causes or could cause the state or event asserted by  $S_0$ .
- **Occasion:** A change of state can be inferred from the assertion of  $S_0$ , whose final state can be inferred from  $S_1$ , or vice versa.
- **Parallel:** Infer  $p(a_1, a_2, \dots)$  from the assertion of  $S_0$  and  $p(b_1, b_2, \dots)$  from the assertion of  $S_1$ , where  $a_i$  and  $b_i$  are similar for all  $i$ .

# RST Coherence Relations



# RST formal relation definition

- Relation name: **Evidence**
- Constr on N: R not believing N enough for W
- Constr on S: R believes S, or would
- Constr on N+S: R's believing S would increase R's believing N
- Effects: R's belief of N is increased



# Automatic Coherence Assignment

Given a sequence of sentences or clauses , we want to automatically:

- determine coherence relations between them (coherence relation assignment)
- extract a tree or graph representing an entire discourse (discourse parsing)

# Automatic Coherence Assignment

**Very difficult.** One existing approach is to use cue phrases.

John hid Bill's car keys because he was drunk.

The scarecrow came to ask for a brain. Similarly, the tin man wants a heart.

- 1) Identify cue phrases in the text.
- 2) Segment the text into discourse segments.
- 3) Classify the relationship between each consecutive discourse segment.

# Automatic Coherence Assignment

- “Discourse parsing”?
- Use **cue phrases**/discourse markers
  - *although, but, because, yet, with, ...*
  - but often implicit, as in car key example
- Use **abduction**, defeasible inference
  - All men are mortal
  - Max was mortal
  - **Maybe** Max was a man
- *The city denied the demonstrators a permit because they (feared/advocated) violence*

# Pragmatics

# Pragmatics

Pragmatics is a branch of linguistics dealing with language use in context.

When a diplomat says yes, he means 'perhaps';

When he says perhaps, he means 'no';

When he says no, he is not a diplomat.

(Variously attributed to Voltaire, H. L. Mencken, and Carl Jung)

# In Context?

- Social context
  - Social identities, relationships, and setting
- Physical context
  - Where? What objects are present? What actions?
- Linguistic context
  - Conversation history
- Other forms of context
  - Shared knowledge, etc.

# Speech Acts

# (Direct) Speech Acts

---

- **Mood** of a sentence indicates relation between speaker and the concept (proposition) defined by the LF
- There can be operators that represent these relations:
  - ASSERT: the proposition is proposed as a fact
  - YN-QUERY: the truth of the proposition is queried
  - COMMAND: the proposition describes a requested action
  - WH-QUERY: the proposition describes an object to be identified



# Indirect Speech Acts

---

- Can you pass the salt?
- It's warm in here.

# Austin, How to do things with words

- In addition to just saying things, sentences perform actions.
- When these sentences are uttered, the important thing is not their truth value, but the *felicitousness* of the action (e.g., do you have the *authority* to do it):
  - I name this ship the Queen Elizabeth.
  - I take this man to be my husband.
  - I bequeath this watch to my brother.
  - I declare war.
    - [http://en.wikipedia.org/wiki/J. L. Austin](http://en.wikipedia.org/wiki/J._L._Austin)

# Performative sentences

- You can tell whether sentences are performative by adding “**hereby**”:
  - I hereby name this ship the Queen Elizabeth.
  - I hereby take this man to be my husband.
  - I hereby bequeath this watch to my brother.
  - I hereby declare war.
- Non-performative sentences do not sound good with hereby:
  - Birds hereby sing.
  - There is hereby fighting in Syria.

# Austin continued

- *Locution*: say some words
- *Illocution*: an action performed *in* saying words
  - Ask, promise, command
- *Perlocution*: an action performed *by* saying words, probably the effect that an illocution has on the listener.
  - Persuade, convince, scare, elicit an answer, etc.

# Searle's speech acts

Searle (1975) has set up the following classification of illocutionary speech acts:

- **assertives** = speech acts that commit a speaker to the truth of the expressed proposition, e.g. reciting a creed
- **directives** = speech acts that are to cause the hearer to take a particular action, e.g. requests, commands and advice
- **commissives** = speech acts that commit a speaker to some future action, e.g. promises and oaths
- **expressives** = speech acts that express the speaker's attitudes and emotions towards the proposition, e.g. congratulations, excuses and thanks
- **declarations** = speech acts that change the reality in accord with the proposition of the declaration, e.g. baptisms, pronouncing someone guilty or pronouncing someone husband and wife

- [http://en.wikipedia.org/wiki/Speech\\_act](http://en.wikipedia.org/wiki/Speech_act)

# Searle example

- Indirect speech acts:
  - Can you pass the salt?
    - Has the form of a *question*, but the effect of a *directive*.

# Speech Acts in NLP

# Task-Oriented Dialogue

- Making travel reservations (flight, hotel room, etc.)
- Scheduling a meeting.
- Task oriented dialogues that are frequently done with computers:
  - Finding out when the next bus is.
  - Making a payment over the phone.



# Ways to ask for a room

- I'd like to make a reservation
- I'm calling to make a reservation
- Do you have a vacancy on ...
- Can I reserve a room
- Is it possible to reserve a room

Domain-specific speech acts:  
travel scheduling (NESPOLE! Project)  
(a primitive version of the speech translation)

- 61.2.3 olang ITA lang ITA Prv IRST “Telefono per prenotare delle stanze per quattro colleghi”
- 61.2.3 olang ITA lang ENG Prv IRST “I am calling to book some rooms for four colleagues”
- 61.2.3 IF Prv IRST c:request-action+reservation+room (room-spec=(room, quantity=some), for-whom=(colleague, quantity=4))
- comments: dial-oo5-spkB-roca0-02-3

# Task-oriented dialogue acts related to negotiation

- Suggest
  - I recommend this hotel.
- Offer
  - I can send some brochures.
  - How about if I send some brochures.
- Accept
  - Sure. That sounds fine.
- Reject
  - No. I don't like that one.



*"No, Thursday's out. How about never—is never good for you?"*

# Examples of Speech Act inventories used in language technologies

- These inventories are actually *annotation schemes*.
- They are used for corpus annotation.
- The corpus annotation is used for automated learning.
- They are highly developed and checked for intercoder agreement.
  - But still take a long time to learn.

# Examples of task-oriented speech acts

- Identify self:
  - This is Lori
  - My name is Lori
  - I'm Lori
  - Lori here
- Sound check: Can you hear me?
- Meta dialogue act: There is a problem.
- Greet: Hello.
- Request-information:
  - Where are you going.
  - Tell me where you are going.

# Examples of task-oriented speech acts

- Backchannel:
  - Sounds you make to indicate that you are still listening
  - ok, m-hm
- Apologize/reply to apology
- Thank/reply to thanks
- Request verification/Verify
  - So that's 2:00? Yes. 2:00.
- Resume topic
  - Back to the accommodations....
- Answer a yes/no question: yes, no.

# DAMSL

## Dialogue Act Markup in Several Layers

- For task-oriented or non-task-oriented dialogue.
- However, much of the development was related to task-oriented dialogues:
  - Trains corpus
  - Maptask corpus
  - Meeting scheduling corpus
- Although it has been used for non-task-oriented dialogue:
  - Switchboard corpus (JHU workshop 1997)
  - Spanish CallHome corpus (Clarity Project, Waibel, Levin, Lavie)
  - Text message corpus (Proprietary project, Levin, Rudnicky, Tenny)
- What are the layers?
  - Forward function: offer, ask
  - Backward function: backchannel, accept, reject



# Forward looking functions

- Statement
  - Assert
  - Reassert
  - Other-statement
- Influencing-addressee-future-action
  - Open-option
  - Action-directive
- Info-request
- Committing-speaker-future-action
  - [Offer](#)
  - Commit
- Conventional Opening Closing
- Explicit-performative
- Exclamation
- Other-forward-function

# Backward looking functions

- Agreement
  - Accept
  - Accept part
  - Maybe
  - Reject part
  - Reject
  - Hold
- Understanding
  - Signal non-understanding
  - Signal understanding
    - Acknowledge
    - Repeat
    - Complete
  - Correct misspeaking
- Answer

Now, a famous bad idea  
(linked to a good idea)

# Grice's Maxims

- Why do these make sense?
  - *Are you 21?*
  - *Yes. I'm 25.*
  
  - *I'm hungry.*
  - *I'll get my keys.*
  
  - *Where can I get cigarettes?*
  - *There is a gas station across the street.*

# Grice's Maxims

- Why are these strange?
  - (The students are all girls.)
  - *Some students are girls.*
  
  - (There are seven non-stop flights.)
  - *There are three non-stop flights.*
    - Jurafsky and Martin, page 820
  
  - (In a letter of recommendation for a job)
  - *I strongly praise the applicant's impeccable handwriting.*

# Grice's *Cooperative Principle*

- “Make your contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.”
- The Cooperative Principle is good and right.
- On the other hand, we have the Maxims:

# Grice's actual Maxims

- Maxim of Quality
  - Try to say something true; do not say something false or for which you lack evidence.
- Maxim of Quantity
  - Say as much as is required to be informative
  - Do not make your contribution more informative than required
- Maxim of Relevance
  - Be Relevant
- Maxim of Manner
  - Be perspicuous
  - Avoid ambiguity
  - Be brief
  - Be orderly

# *Flouting* the Cooperative Principle

- “Nice throw.” (*said after terrible throw*)
- “If you run a little slower, you’ll never catch up to the ball.” (*during mediocre pursuit of ball*)
- You *can* indeed imply something by clearly violating the principle.
  - The Maxims *still* suck.



# *Flout ≠ Flaunt*

- *Flout*: openly disregard (a rule, law or convention).
- *Flaunt*: display (something) ostentatiously, especially in order to provoke envy or admiration or to show defiance.
  - Source: Google

# My paper on the Maxims

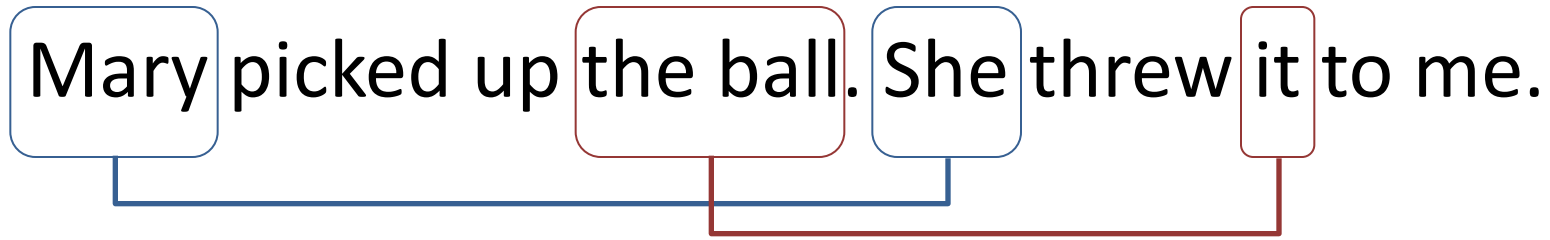
- [Grice's Maxims: "Do the Right Thing"](#) by Robert E. Frederking. Argues that the Gricean maxims are too vague to be useful for natural language processing. [from Wikipedia article]
- “I used to think you were a nice guy.”
  - Actual quote from a grad student, after reading the paper

Reference resolution

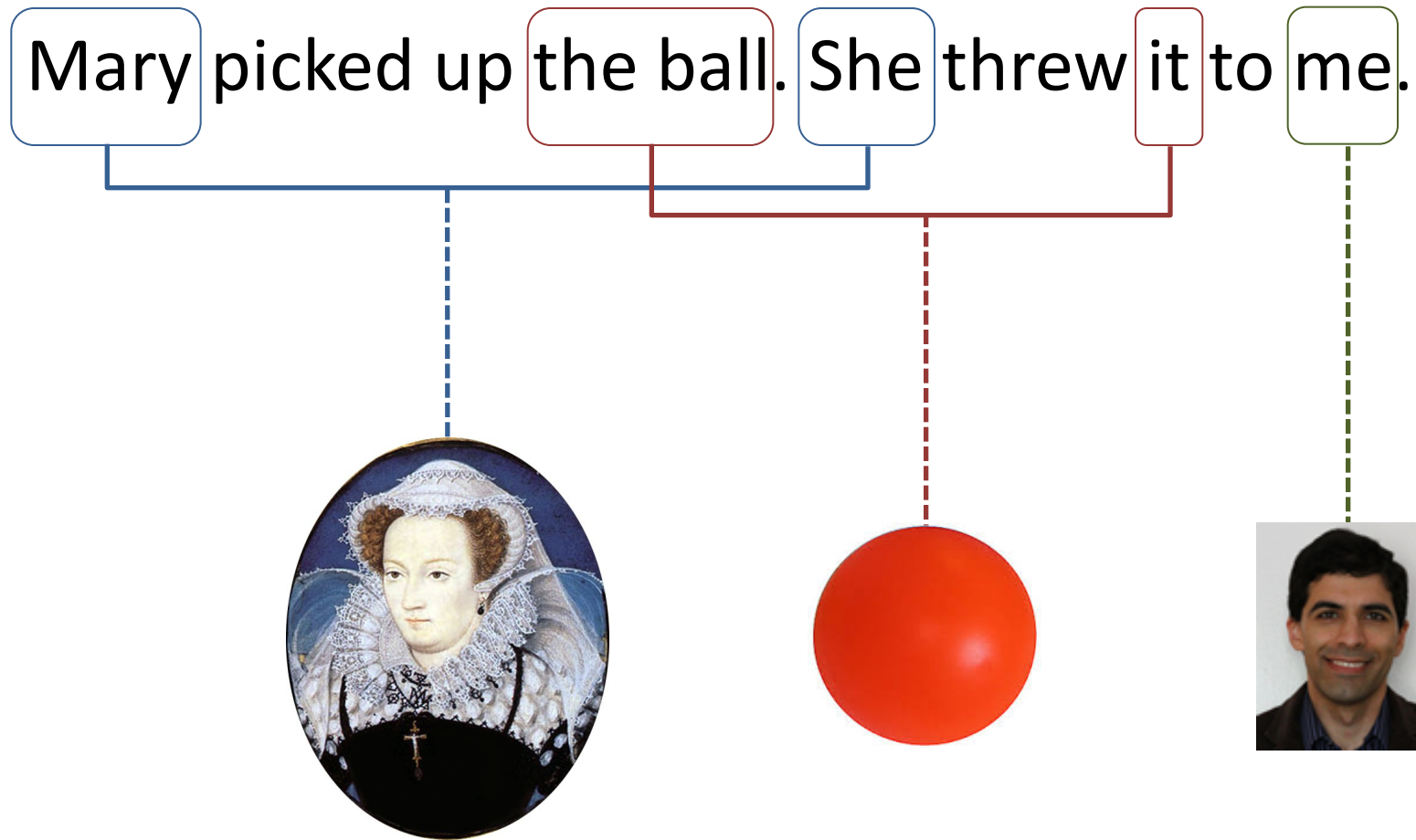
# Reference Resolution: example

- Victoria Chen, CFO of Megabucks Banking Corp since 2004, saw her pay jump 20%, to \$1.3 million, as the 37-year-old also became the Denver-based company's president. It has been ten years since she came to Megabucks from rival Lotsaloot.
- Should give 4 coreference chains:
  - {Victoria Chen, CFO of Megabucks Banking Corp since 2004, her, the 37-year-old, the Denver-based company's president, she}
  - {Megabucks Banking Corp, the Denver-based company, Megabucks}
  - {her pay}
  - {Lotsaloot}
  -

# Coreference Resolution



# Reference resolution



# (Co)Reference Resolution

- Determining the referent of a referring expression. Anaphora, antecedents corefer.
- 1961 Ford Falcon: *it, this, that, this car, the car, the Ford, the Falcon, my friend's car, ...*
- Coreference chains are part of cohesion
- Note: other kinds of referents:
  - *According to Doug, Sue just bought the Ford Falcon*
    - *But **that** turned out to be a lie*
    - *But **that** was false*
    - ***That** struck me as a funny way to describe the situation*
    - ***That** caused a financial problem for Sue*

# Types of Referring Expressions

- Indefinite NPs: *a/an, some, this*, or nothing
  - new entities; specific/non-specific ambiguity
- Definite NPs: usually *the*
  - an entity identifiable by the hearer
- Pronouns: *he, them, it*, etc. Also **cataphora**.
  - strong constraints on their use
  - can be bound: *Every student improved his grades*
- Demonstratives: *this, that*
- Names: construed to be unique, but they aren't
  - *Is that the Bob in LTI or the Bob in the Lane Center?*



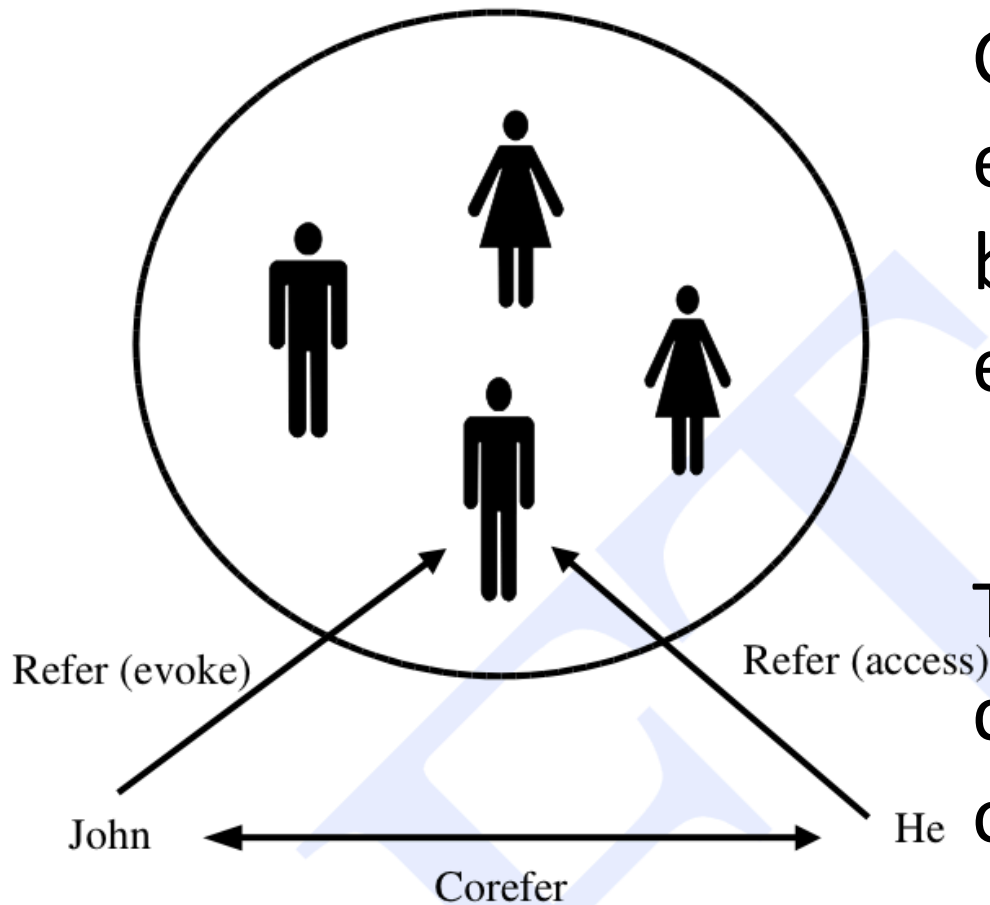
# Information structure: given/new

- *Where are my **shoes**? Your **shoes** are in the **closet***
- *What's in the **closet**?*
  - *??Your **shoes** are in the **closet**.*
  - *Your **shoes** are in the **closet**.*
- Definiteness/pronoun, length, position in S

# Complications

- Inferrables: *Some car. ... a door ... the engine ...*
- Generics: *At CMU you have to work hard.*
- Pleonastic/clefts/extraposition:
  - *It is raining. It was me who called. It was good that ...*

# Reference Resolution



Goal: determine what entities are referred to by which linguistic expressions.

The **discourse model** contains our eligible set of referents.

# Pronouns: Filters and Preferences

# Pronoun reference resolution: filters

- Agreement in number, person, gender
  - Pittsburgh dialect: *yinz=youse=y'all*
  - UK dialect: *Newcastle are a physical team.*
  - L can have >2 numbers, >3 persons, or >3 genders
- Binding theory: **reflexive** required/prohibited:
  - *John bought himself a new Ford.* [himself=John]
  - *John bought him a new Ford.* [him!=John]
  - *John said that Bill bought him a new Ford.* [him!=Bill]
  - *J said that B bought himself a new F.* [himself=Bill]
  - *He said that he bought J a new Ford.* [both he!=J]

# Pronoun reference resolution: preferences

- Recency: preference for most recent referent
- Grammatical Role: subj>obj>others
  - *Billy went to the bar with Jim. He ordered rum.*
- Repeated mention: *Billy had been drinking for days. He went to the bar again today. Jim went with him. He ordered rum.*
- Parallelism: *John went with Jim to one bar. Bill went with him to another.*
- Verb semantics: *John phoned/criticized Bill. He lost the laptop.*
- Selectional restrictions: *John parked his car in the garage after driving it around for hours.*

# Three computational approaches to pronouns

# PN ref. res. 1: Hobbs Algorithm

- Algorithm for walking through parses of current and preceding sentences
- Simple, often used as baseline
- Requires parser, morph gender and number
  - plus head rules and WordNet for NP gender
- Implements binding theory, recency, and grammatical role preferences



# PN ref. res. 2: Centering theory

- Claim: a single entity is “centered” in each S
- Backward-looking center, Forward-looking centers
- $C_b$  = most highly ranked  $C_f$  used from prev. S
- Rank: Subj>ExistPredNom>Obj>IndObj-Obl>DemAdvPP
- Defined transitions: ( $C_p$  is front of  $C_f$  list)

	$C_b(U_{n+1}) = C_b(U_n)$ 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

Rule 1: If any  $C_f$  used as  $Pro_{n+1}$ , then  $C_{b(n+1)}$  must be Pro too

Rule 2: Rank: Continue>Retain>Smooth>Rough

*U1: John saw a Ford at the dealership*

Cb: NIL

Cf: John, Ford, dealership

*U2: He showed it to Bob* [Bob!=he]

He=John, it={Ford, dealership}

Cb=John

- (it->Ford) => Cf: {John,Ford,Bob} => CONTINUE [tie-winner]
- (it->dealership) => Cf: {John,dealer,Bob} => CONTINUE

*U3: He bought it* [*dealership* is now unavailable]

He={John,Bob}, it=Ford

- (he->John) => Cb=John, Cf={John,Ford} => CONTINUE [Win]
- (he->Bob) => Cb=Bob, Cf={Bob,Ford} => SMOOTH

# Centering theory

- Same requirements as Hobbs
- Implements Grammatical Role, Recency, and Repeated Mention
- Can make mistakes:
  - *Bob opened a new dealership last week*
  - *John took a look at the Fords in his lot [Cb=Bob]*
  - *He ended up buying one*
    - He=Bob => CONTINUE, He=John => SMOOTH

# PN ref. res. 3: Log-linear model

- Supervised: hand-labelled coref corpus
- Rule-based filtering of non-referential pronouns
- Features, values for *He* in U3:

	He ( $U_2$ )	it ( $U_2$ )	Bob ( $U_2$ )	John ( $U_1$ )
<b>strict number</b>	1	1	1	1
<b>compatible number</b>	1	1	1	1
<b>strict gender</b>	1	0	1	1
<b>compatible gender</b>	1	0	1	1
<b>sentence distance</b>	1	1	1	2
<b>Hobbs distance</b>	2	1	0	3
<b>grammatical role</b>	subject	object	PP	subject
<b>linguistic form</b>	pronoun	pronoun	proper	proper

# General Reference Resolution

# General Coreference Resolution

- Victoria Chen, CFO of Megabucks Banking Corp since 2004, saw her pay jump 20%, to \$1.3 million, as the 37-year-old also became the Denver-based company's president. It has been ten years since she came to Megabucks from rival Lotsaloot.
- Should give 4 coreference chains:
  - {Victoria Chen, CFO of Megabucks Banking Corp since 2004, her, the 37-year-old, the Denver-based company's president, she}
  - {Megabucks Banking Corp, the Denver-based company, Megabucks}
  - {her pay}
  - {Lotsaloot}
  -

President Park Geun-hye of South Korea ordered the country's military on Monday to deliver a strong and immediate response to any North Korean provocation, the latest turn in a war of words that has become a test of resolve for the relatively unproven leaders in both the North and South.

“I consider the current North Korean threats very serious,” Ms. Park told the South's generals. “If the North attempts any provocation against our people and country, you must respond strongly at the first contact with them without any political consideration.

“As top commander of the military, I trust your judgment in the face of North Korea's unexpected surprise provocation,” she added.

Since Kim Jong-un took power after the death of his father, Kim Jong-il, in late 2011, the North has taken a series of provocative steps and amplified threats against Washington and Seoul to much louder and more menacing levels. The North has launched a three-stage rocket, tested a nuclear device and threatened to hit major American cities with nuclear-armed ballistic missiles. And Mr. Kim has declared that the Korean Peninsula has reverted to a “state of war.”

President Park Geun-hye of South Korea ordered the country's military on Monday to deliver a strong and immediate response to any North Korean provocation, the latest turn in a war of words that has become a test of resolve for **the relatively unproven leaders** in both the North and South.

“I consider the current North Korean threats very serious,” Ms. Park told the South's generals. “If the North attempts any provocation against our people and country, you must respond strongly at the first contact with them without any political consideration.

“As **top commander of the military**, I trust your judgment in the face of North Korea's unexpected surprise provocation,” she added.

Since Kim Jong-un took power after the death of his father, Kim Jong-il, in late 2011, the North has taken a series of provocative steps and amplified threats against Washington and Seoul to much louder and more menacing levels. The North has launched a three-stage rocket, tested a nuclear device and threatened to hit major American cities with nuclear-armed ballistic missiles. And Mr. Kim has declared that the Korean Peninsula has reverted to a “state of war.”



President Park Geun-hye of **South Korea** ordered **the country**'s military on Monday to deliver a strong and immediate response to any North Korean provocation, the latest turn in a war of words that has become a test of resolve for the relatively unproven leaders in both the North and **South**.

“I consider the current North Korean threats very serious,” Ms. Park told the **South**'s generals. “If the North attempts any provocation against **our people and country**, you must respond strongly at the first contact with them without any political consideration.

“As top commander of the military, I trust your judgment in the face of North Korea's unexpected surprise provocation,” she added.

Since Kim Jong-un took power after the death of his father, Kim Jong-il, in late 2011, the North has taken a series of provocative steps and amplified threats against Washington and **Seoul** to much louder and more menacing levels. The North has launched a three-stage rocket, tested a nuclear device and threatened to hit major American cities with nuclear-armed ballistic missiles. And Mr. Kim has declared that **the Korean Peninsula** has reverted to a “state of war.”

President Park Geun-hye of South Korea ordered the country's military on Monday to deliver a strong and immediate response to any **North Korean** provocation, the latest turn in a war of words that has become a test of resolve for the relatively unproven leaders in both **the North** and South.

“I consider the current **North Korean** threats very serious,” Ms. Park told the South's generals. “If **the North** attempts any provocation against our people and country, you must respond strongly at the first contact with them without any political consideration.

“As top commander of the military, I trust your judgment in the face of **North Korea**'s unexpected surprise provocation,” she added.

Since Kim Jong-un took power after the death of his father, Kim Jong-il, in late 2011, **the North** has taken a series of provocative steps and amplified threats against Washington and Seoul to much louder and more menacing levels. **The North** has launched a three-stage rocket, tested a nuclear device and threatened to hit major American cities with nuclear-armed ballistic missiles. And Mr. Kim has declared that **the Korean Peninsula** has reverted to a “state of war.”

President Park Geun-hye of South Korea ordered the country's military on Monday to deliver a strong and immediate response to any North Korean provocation, the latest turn in a war of words that has become a test of resolve for the relatively unproven leaders in both the North and South.

“I consider the current North Korean threats very serious,” Ms. Park told the South's generals. “If the North attempts any provocation against our people and country, you must respond strongly at the first contact with them without any political consideration.

“As top commander of the military, I trust your judgment in the face of North Korea's unexpected surprise provocation,” she added.

Since [Kim Jong-un](#) took power after the death of his father, [Kim Jong-il](#), in late 2011, the North has taken a series of provocative steps and amplified threats against Washington and Seoul to much louder and more menacing levels. The North has launched a three-stage rocket, tested a nuclear device and threatened to hit major American cities with nuclear-armed ballistic missiles. And [Mr. Kim](#) has declared that the Korean Peninsula has reverted to a “state of war.”

# High-Level Recipe for Coreference Resolution

1. Parse the text and identify NPs; then
2. For every pair of NPs, carry out binary classification: coreferential or not?
3. Collect the results into coreferential chains

What do we need?

- A choice of classifier
- Lots of labeled data
- Features

# Features?

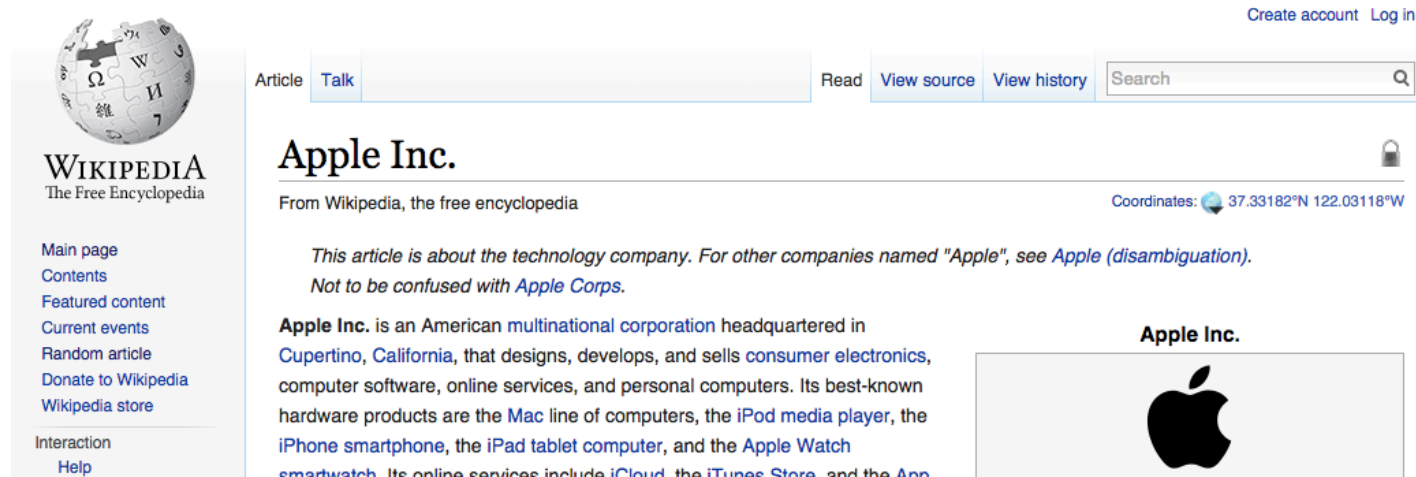
- Edit distance between the two NPs
- Are the two NPs the same NER type?
- Appositive syntax
  - “Alan Shepherd, the first American astronaut...”
- Proper/definite/indefinite/pronoun
- Gender
- Number
- Distance in sentences
- Number of NPs between
- Grammatical role
- etc.

# More Coreference Resolution

- Combine best: ENCORE (Bo Lin et al 2010)
- ML for Cross-Doc Coref (Rushin Shah et al 2011)

# Entity Linking

*Apple updated its investor relations page today to note that it will announce its earnings for the second fiscal quarter (first calendar quarter) of 2015 on Monday, April 27.*



The image shows a screenshot of the Wikipedia article for "Apple Inc.". The page layout includes a sidebar on the left with navigation links such as "Main page", "Contents", and "Featured content". The main content area features the article title "Apple Inc." with a lock icon, a search bar, and a "Coordinates" field. The introductory text states: "This article is about the technology company. For other companies named 'Apple', see [Apple \(disambiguation\)](#). Not to be confused with [Apple Corps](#)." Below this, the text begins with "Apple Inc. is an American multinational corporation headquartered in Cupertino, California, that designs, develops, and sells consumer electronics, computer software, online services, and personal computers. Its best-known hardware products are the Mac line of computers, the iPod media player, the iPhone smartphone, the iPad tablet computer, and the Apple Watch smartwatch. Its online services include iCloud, the iTunes Store, and the App Store." To the right of the text is a placeholder for an image of the Apple logo, with the text "Apple Inc." above it.

# One Approach to Entity Linking

Use supervised learning: Train on known references to each entity. Use features from context (bag of words, syntax, etc.).

## iPhone

---

From Wikipedia, the free encyclopedia

*This article is about the line of smartphones by Apple. For other uses, see [iPhone \(disambiguation\)](#).*

**iPhone** (/ˈaɪfoʊn/ *EYE-fohn*) is a line of [smartphones](#) designed and marketed by [Apple Inc.](#) It runs Apple's [iOS](#) mobile operating system.<sup>[13]</sup> The first generation iPhone was released on June 29, 2007; the most recent iPhone models are the [iPhone 6](#) and [iPhone 6 Plus](#), which were unveiled at a special event on September 9, 2014.<sup>[14]</sup>



Questions?