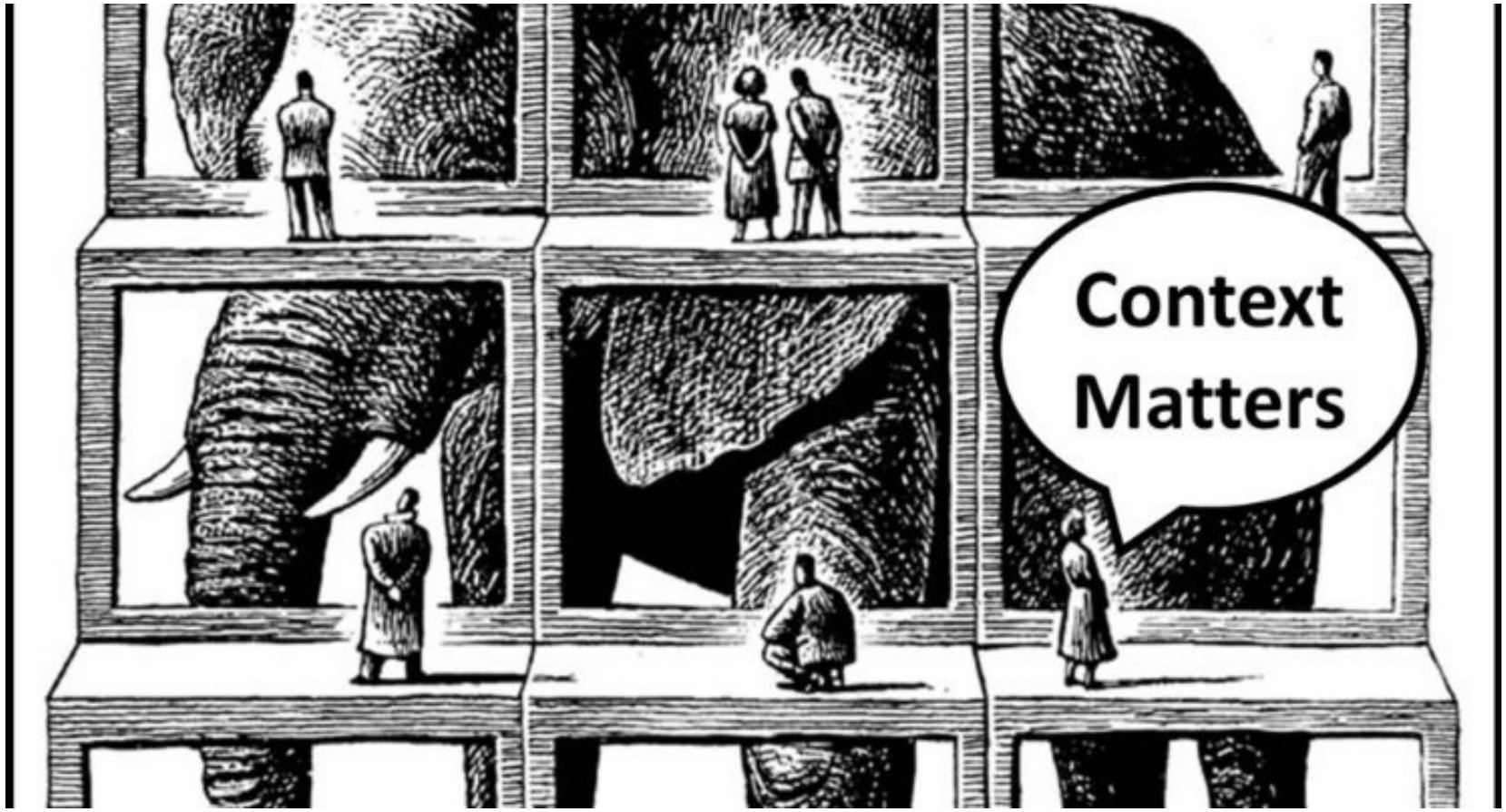
Natural Language Processing

Discourse and Pragmatics

Learning Objectives

- **Discourse**
  - Know three **coherence relations** and how to apply them
  - Know **linguistic devices** that indicate **cohesion**
  - Know how **discourse segmentation** is implemented and evaluated
  - Understand how **discourse** can be **parsed**
- **Pragmatics**
  - Know that **pragmatic meaning** is the meaning of language in context
  - Understand **speech act theory**
  - Know how to apply **Grice's cooperative principle**

With Language, as with Elephants, Context is Everything



There Are Multiple Kinds of Linguistic Context

- **Discourse context** (where an utterance sits in relation to a document, conversation, speech, etc.)
- **Physical context**
- **Social context**
  - Who is speaking
  - Who they are speaking to
  - What they are trying to acheive

Compare the Relationship between these Two Pairs of Sentences

John hid Bill’s car keys. He was drunk.

*John hid Bill’s car keys. He likes spinach.

Another Example

“Near-death experiences can help one see more clearly sometimes,” said Steve Jobs. He was speaking about struggling companies. Yet he could easily have been talking about his own life. In 1985 Mr Jobs was pushed out of Apple Computer, the firm he had helped found, only to return after a decade away. In doing so, he mounted one of capitalism’s most celebrated comebacks.

*Yet he could easily have been talking about his own life. “Near-death experiences can help one see more clearly sometimes,” said Steve Jobs. In doing so, he mounted one of capitalism’s most celebrated comebacks. In 1985 Mr Jobs was pushed out of Apple Computer, the firm he had helped found, only to return after a decade away. He was speaking about struggling companies.

http://www.economist.com/news/books-and-arts/21647593-new-book-attempts-reconstruct-one-worlds-most-celebrated-inventors-jobs-20

Definition of 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.**

DISCOURSE COHESION

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

Pyrmaid from commons.wikimedia.org

Cohesion

Relations between words in two units (sentences, paragraphs) “glue” them together.

Before winter **I** built a chimney, and **shingled** the sides of my **house**… **I** have thus a tight **shingled** and plastered **house**.

Peel, core, and slice **the** **pears** **and** **apples**. Add **the** **fruit** to the skillet.

There are Three Main Classes of Features for Discourse Cohesion

- Lexical overlap/lexical chains
- Coreference chains
- Cue words/discourse markers

Exercise: Divide this Passage into Paragraphs using Features for Cohesion

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Identify the Lexical Chains

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Exercise: Features for Cohesion

Ping worked hard to complete her last **assignment** and did her own work. But when she got **homework** scores back, she was very surprised. Because another student had *similar* answers, she was accused of *cheating*. It turns out that the other student was in the *same* program. He had *copied* down the answers during a group study **session**. The **session** ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the *cheater* instead.

Identify the Coreference Chains

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Identify the Coreference Chains

Ping worked hard to complete **her** last assignment and did **her** own work. But when **she** got homework scores back, **she** was very surprised. Because *another student* had similar answers, **she** was accused of cheating. It turns out that *the other student* was in the same program. *He* had copied down the answers during ***a group study session***. ***The session*** ran late, and **Ping** had left her computer in the room when she went to the restroom. Now **Ping** was distraught, but the professors listened to **her** story. They realized that **she** was telling the truth and punished *the cheater* instead.

Identify Discourse Markers

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Identify Discourse Markers

Ping worked hard to complete her last assignment and did her own work. **But** when she got homework scores back, she was very surprised. **Because** another student had similar answers, she was accused of cheating. **It turns out that** the other student was in the same program. He had copied down the answers during a group study session. The session ran late, **and** Ping had left her computer in the room when she went to the restroom. **Now** Ping was distraught, **but** the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Identify Discourse Markers

Ping worked hard to complete her last assignment and did her own work. But when she got homework scores back, she was very surprised. Because another student had similar answers, she was accused of cheating. ¶ It turns out that the other student was in the same program. He had copied down the answers during a group study session. The session ran late, and Ping had left her computer in the room when she went to the restroom. ¶ Now Ping was distraught, but the professors listened to her story. They realized that she was telling the truth and punished the cheater instead.

Summary of 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?

Evaluating Discourse Segmentation

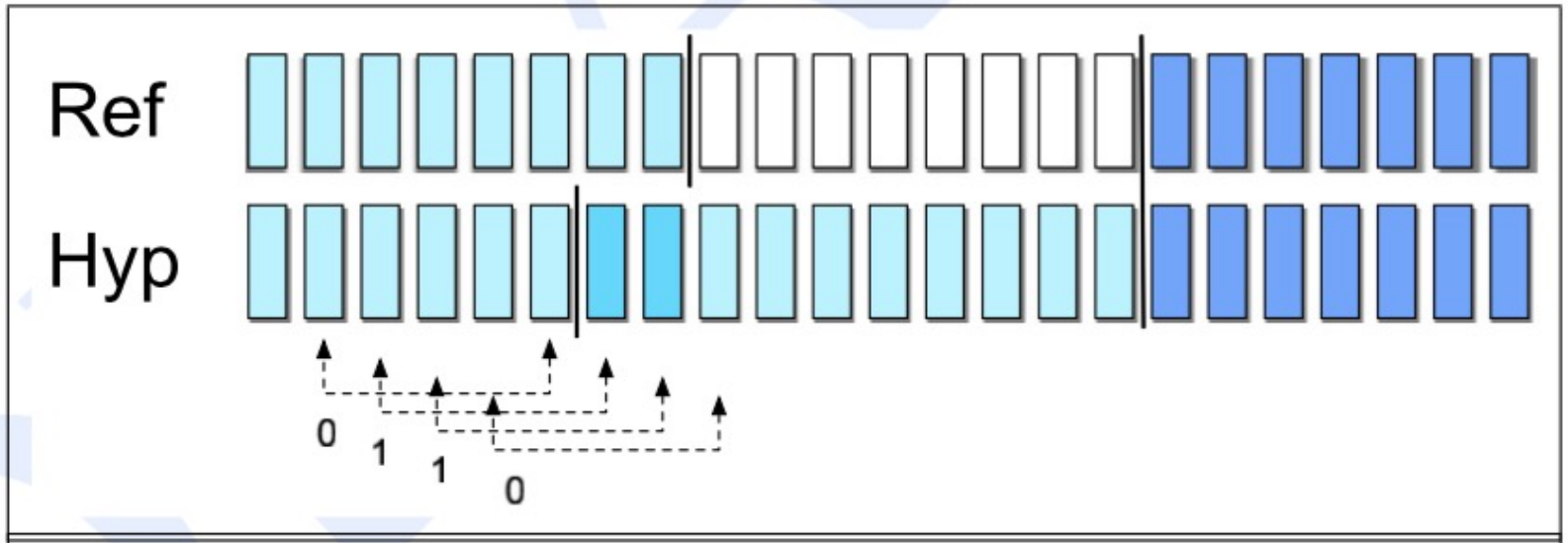


Figure 20.2 The WindowDiff algorithm, showing the moving window sliding over the hypothesis string, and the computation of $|r_i - h_i|$ at four positions. After Pevzner and Hearst (2002).

More formally, if $b(i, j)$ is the number of boundaries between positions i and j in a text, and N is the number of sentences in the text:

$$\text{WindowDiff}(ref, hyp) = \frac{1}{N-k} \sum_{i=1}^{N-k} (|b(ref_i, ref_{i+k}) - b(hyp_i, hyp_{i+k})| \neq 0)$$

DISCOURSE COHERENCE

Coherence relations are meaningful relationships between the sentences in discourse

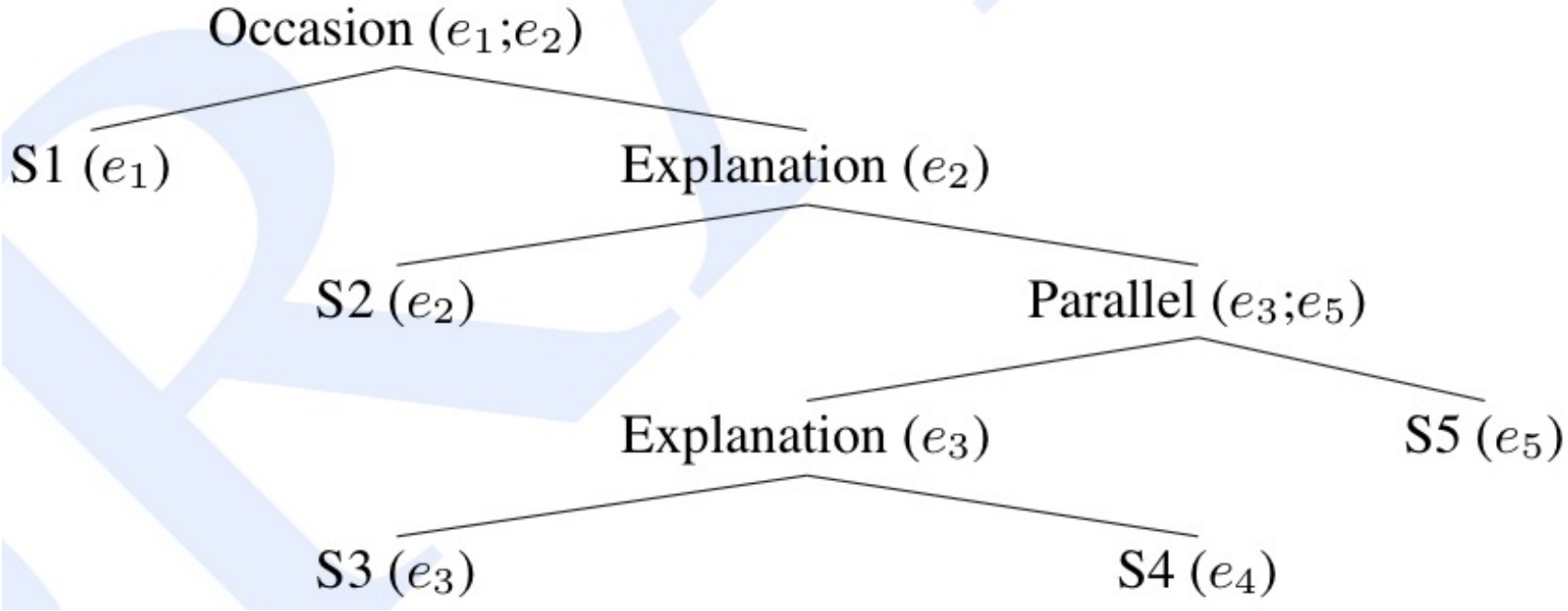
Explanation, Occasion, and Parallel are Examples of Coherence Relations

How can we label the **meaningful** 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 $p(a_1, a_2,\ldots)$ from the assertion of S_0 and $p(b_1, b_2,\ldots)$ from the assertion of S_1, where a_i and b_i are similar for all i.

Coherence Relations Reveal Discourse Structure

John went to the bank to deposit his paycheck. (S1)
He then took a train to Bill's car dealership. (S2)
He needed to buy a car. (S3)
The company he works for now isn't near any public transportation. (S4)
He also wanted to talk to Bill about their softball league. (S5)



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 is Very Difficult

One existing approach is to use cue phrases.

- John hid Bill's car keys <u>because</u> he was drunk. [explanation]
- The scarecrow came to ask for a brain. <u>Similarly</u>, the tin man wants a heart. [parallel]

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

There are now many neural approaches to discourse parsing, with models trained on PDTB or the RST Treebank, but it’s still hard

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)

Quote from http://plato.stanford.edu/entries/pragmatics/

In Context?

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

Speech Act Theory

“I’ll give the lecture today.”

“It’s cold in here.”

"This administration today, here and now, declares unconditional war on poverty in America.”

“I now pronounce you man and wife.”

Speech Act Theory in NLP: An Exercise

Let’s say that I’m building a system that will interact with people conversationally.

Is speech act theory relevant? Why?

Grice's Maxims

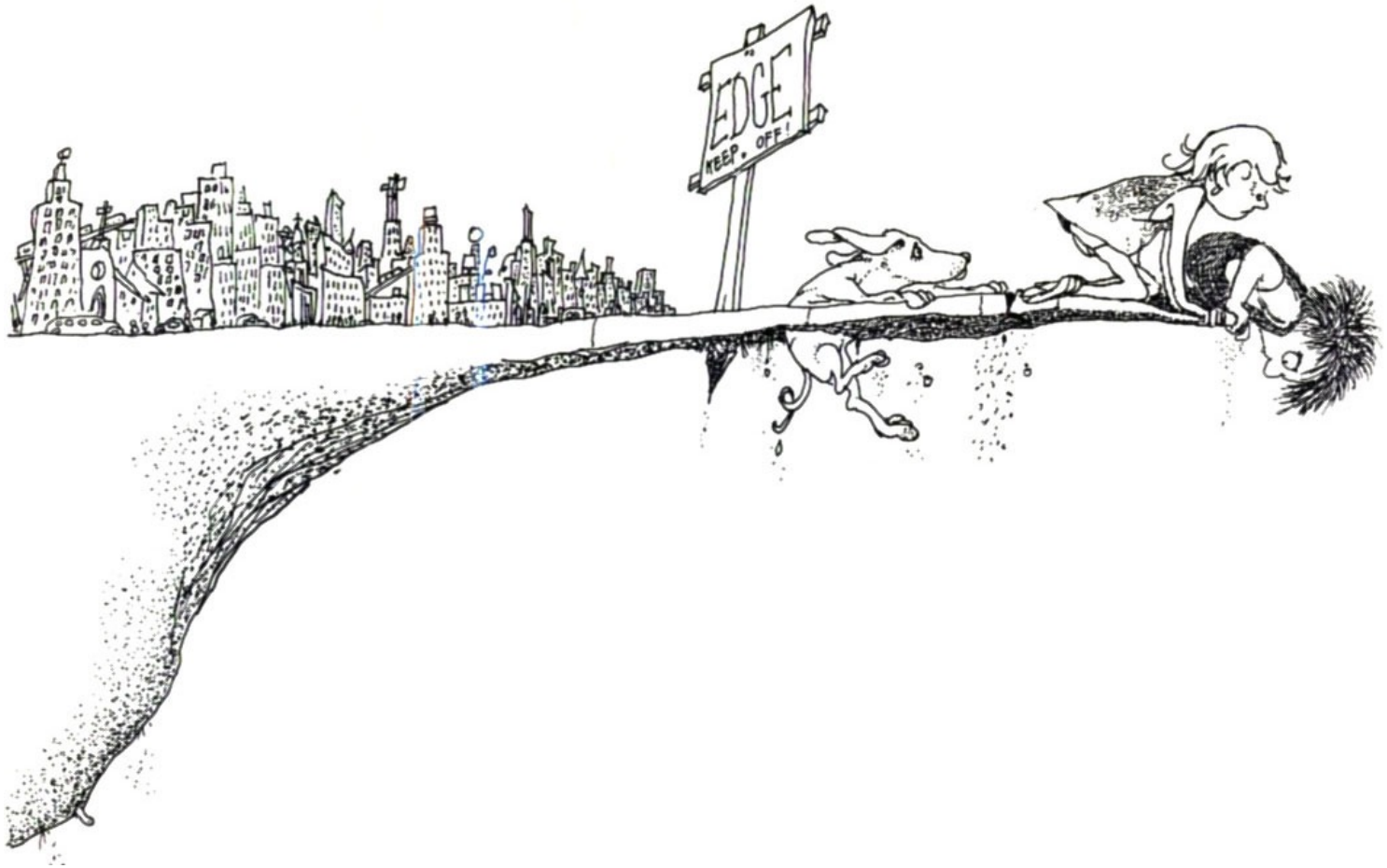
1. **Quantity**: Make your contribution as informative as required, but no more
2. **Quality**: Try to make your contribution one that is true
3. **Relation**: Be relevant
4. **Manner**:
    1. Don't be obscure
    2. Avoid ambiguity
    3. Be brief
    4. Be orderly

The Pragmatics Handbook

Grice’s Maxims in NLP: An Exercise

Let’s say that I’m building a system that will interact with people conversationally.

How are Grice’s Maxims relevant?



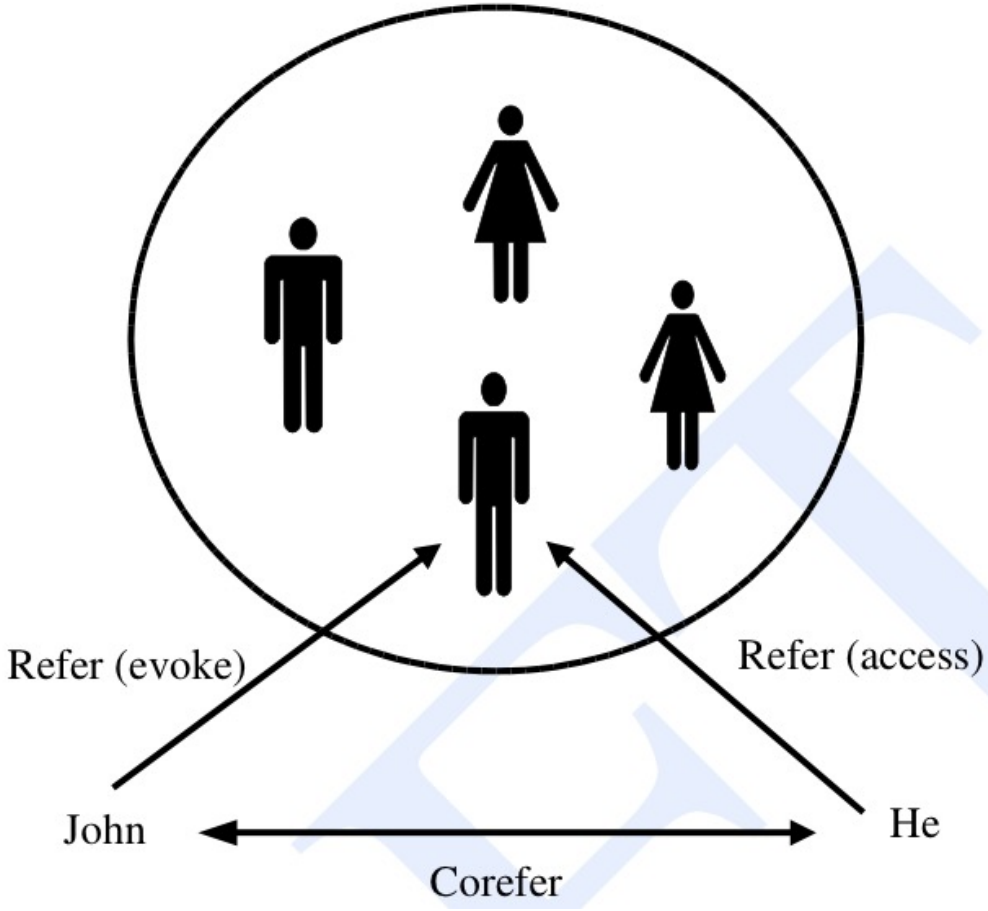
Cover of Shel Silverstein's *Where the Sidewalk Ends* (1974)

ENTITY LINKING/REFERENCE RESOLUTION

A Lead-In: Reference Resolution

John Chang, Chief Financial Officer of Megabucks Banking Corp since 2004, saw his pay jump 20%, to $1.3 million, as the 37-year-old also became the Denver-based financial-services company's president. It has been ten years since he came to Megabucks from rival Lotsabucks.

Reference Resolution



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

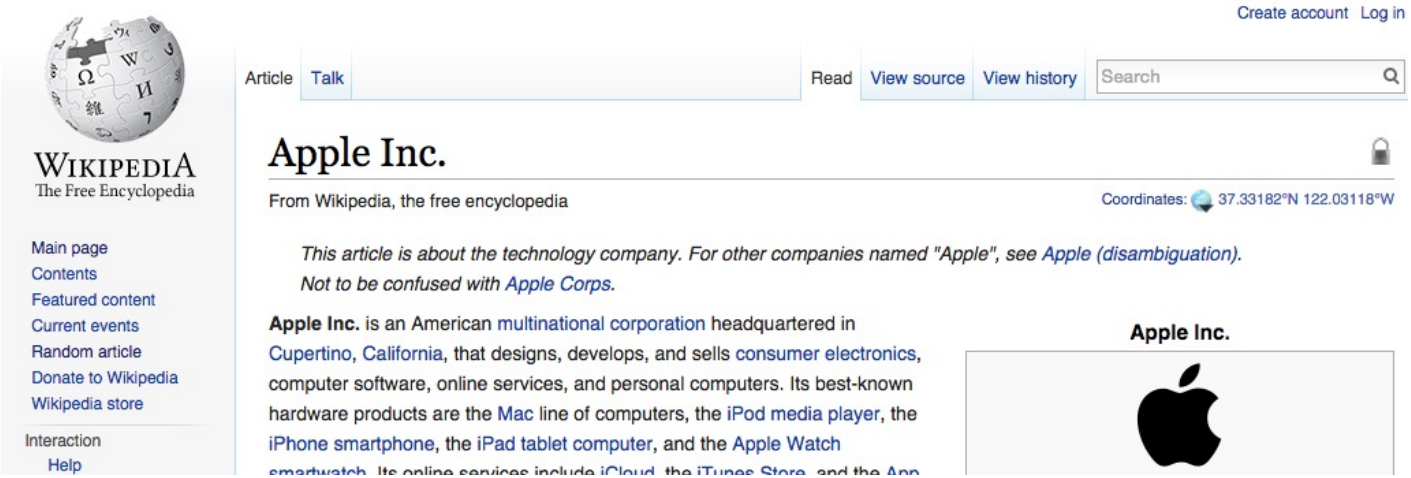
The **discourse model** contains our eligible set of referents.

Five Types of Referring Expressions

- Indefinite noun phrases
  I saw a beautiful Ford Falcon today.
- Definite noun phrases
  I read about it in the New York Times.
- Pronouns
  Emma smiled as cheerfully as she could.
- Demonstratives
  Put it back. This one is in better condition.
- Names
  Miss Woodhouse certainly had not done him justice.

Entity Linking

<u>Apple</u> 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.



News text from http://www.macrumors.com/2015/03/30/apple-to-announce-q2-2015-earnings-on-april-27/

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 iPh (disambiguation).

iPhone (/ˈaɪfoʊn/ *EYE-fohn*) is a line of smartphones designed and marketed by Apple Inc. It runs Apple's iOS mobile operating system.[13] 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.[14]

Questions?