

Topics

- Text coherence
- Discourse segmentation
- ➔ Reference resolution

The problem of reference resolution

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, my brother's wife.

George Burns and Gracie Allen in *The Salesgirl*

Reference resolution

- **Reference:** the process by which speakers use expressions like “John Simon” and “his” to denote a real-world entity
 - **Referring expressions:** NL expression used to perform reference
 - **Referent:** the entity that is referred to
 - **Shorthand form:** *his* refers to John Simon

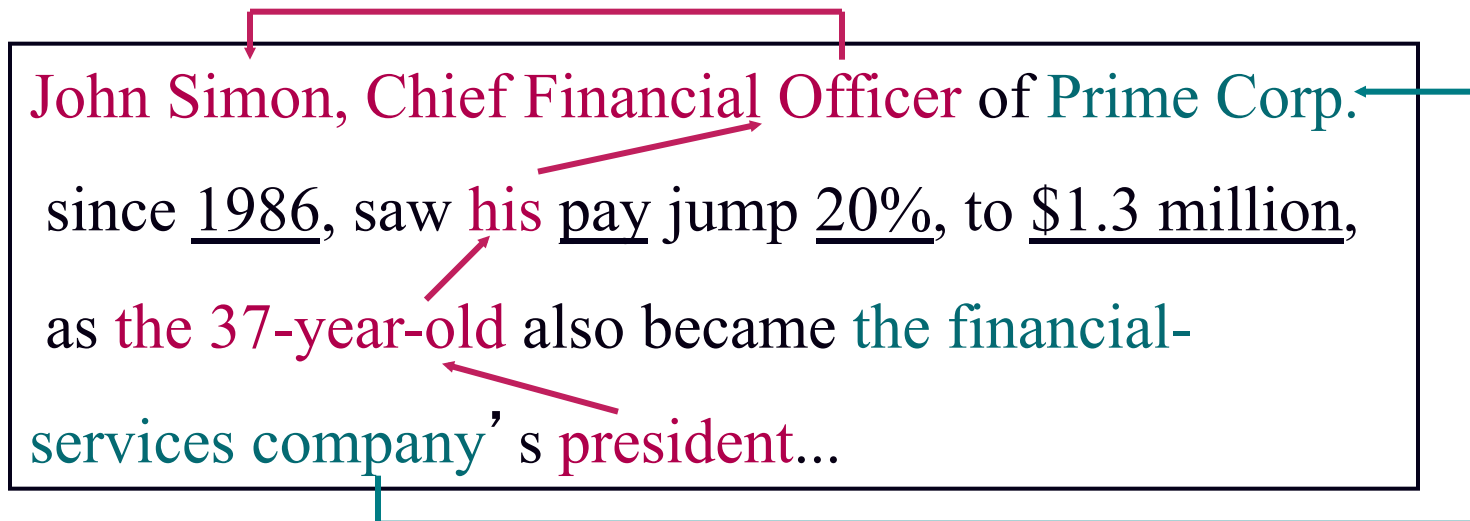


John Simon, Chief Financial Officer of Prime Corp.
since 1986, saw his pay jump 20%, to \$1.3 million,
as the 37-year-old also became the financial-
services company's president...

?

Coreference

- **Coreference:** two referring expressions that are used to refer to the same entity are said to corefer
- *John Simon* is the **antecedent** of *his*.
- Reference to an entity that has been previously introduced into the discourse is called **anaphora**; and the referring expression used is said to be **anaphoric**.



Types of referring expressions

- Indefinite noun phrases
 - Introduce entities that are new to the hearer into the discourse context
 - » I saw *a Subaru WRX* today.
 - » I saw *this awesome Subaru WRX* today.
- Definite noun phrases
 - Refer to an entity that is identifiable to the hearer
 - » It has already been mentioned in the discourse
 - » It is contained in the hearer's set of beliefs about the world
 - » The uniqueness of the object is implied by the description itself
 - ◆ I saw a Subaru WRX today. *The WRX* was blue and needed a wash.
 - ◆ *The Indy 500* is the most popular car race in the US.
 - ◆ *The fastest car in the Indy 500* was a Subaru WRX.

Types of referring expressions

■ Pronouns

- Another form of definite reference
 - » John hid Bill's car keys. *He* was drunk.
- Referent must have a high degree of activation or **salience** in the discourse model
 - » John went to Bob's party, and parked next to a beautiful Subaru WRX. He went inside and talked to Bob for more than an hour. Bob told John that he recently got engaged.
 - (a)?? He also said that he bought *it* yesterday.
 - (a') He also said that he bought *the WRX* yesterday.
- Cataphora: referring expression is mentioned before its referent
 - » Before *he* bought *it*, John checked over the WRX carefully.

Types of referring expressions

- Demonstrative pronouns

- Behave somewhat differently than simple definite pronouns

- » Can appear alone or as determiners

- » Choice of *this* or *that* depends on some notion of spatial or temporal proximity

- ◆ I bought a WRX yesterday. It's similar to the one I bought a year ago. *That one* was really nice, but I like *this one* even better.

- One-anaphora

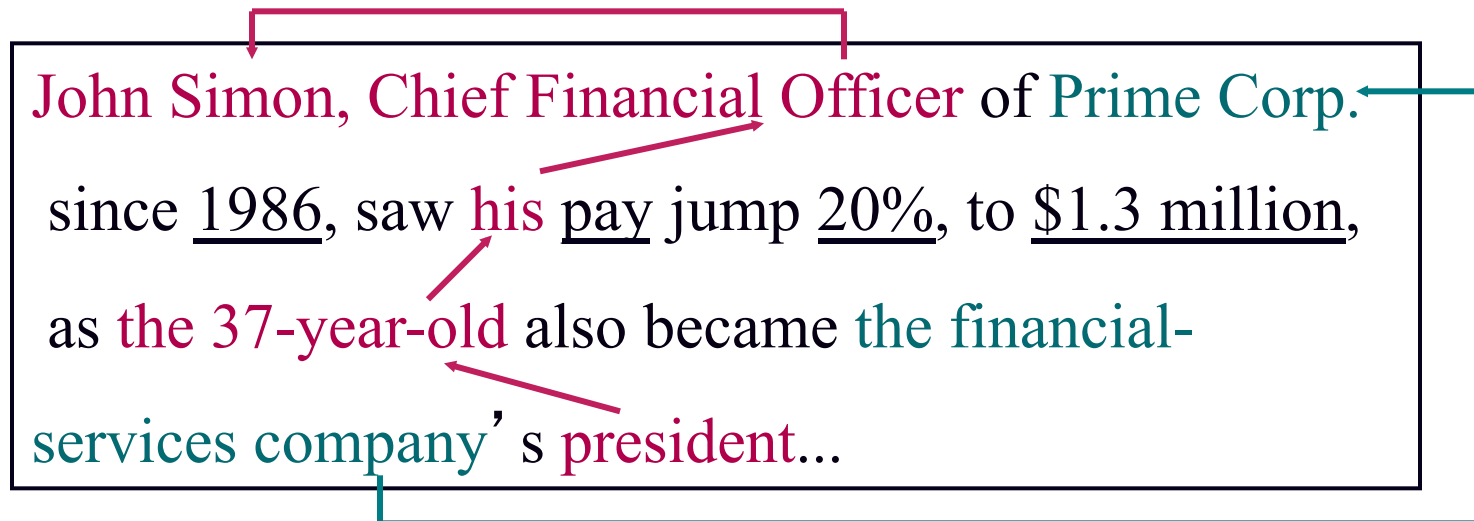
- Blends properties of definite and indefinite reference

- » I saw no fewer than 6 Subaru WRX's today. Now I want *one*.

- May introduce a new entity into the discourse, but it is dependent on an existing referent for the description of this new entity.

Noun Phrase Coreference Resolution

- Identify all phrases that refer to each real-world entity mentioned in the text



Why It's Hard

Coreference strategies differ depending on the type of referring NP

- definiteness of NPs

- » ... Then Mark saw **the man** walking down the street.

- » ... Then Mark saw **a man** walking down the street.

- pronoun resolution alone is notoriously difficult



- » resolution strategies differ for each type of pronoun

- » some pronouns refer to nothing in the text

I went outside and **it** was snowing.

Why It's Hard

Many sources of information play a role

- head noun matches
 - » IBM *executives* = the *executives*
 - » Microsoft *executives*
- syntactic constraints
 - » John helped himself to...

 - » John helped him to...

- discourse focus, recency, syntactic parallelism, semantic class, agreement, world knowledge, ...

Why It's Hard

No single source is a completely reliable indicator

- semantic preferences

- » Mr. Callahan = president =? the carrier

- number and gender

- » assassination (of Jesuit priests) = these murders

- » the woman = she = Mary =? the chairman