Queen Mary
University of London

ECS763 Natural Language Processing

Unit 10: Discourse Processing and Co-reference Resolution

Lecturer: Julian Hough
School of Electronic Engineering and Computer Science

Contents

1) Discourse
2) Coreference (anaphora) resolution
3) Real-world reference resolution
4) Ellipsis resolution
5) Discourse parsing

Contents

1) Discourse
2) Coreference (anaphora) resolution
3) Real-world reference resolution
4) Ellipsis resolution
5) Discourse parsing

Discourse (in the field of Pragmatics)

- We've looked at syntax and semantics, but we have to look at **language use** in different situations- **pragmatics**.
- **Discourse** is, approximately, sequences of sentences.
  - Monologue (one speaker)
  - Dialogue (multiple speakers)
- When we look at discourse, interesting **challenges arise**.
- One of these challenges is interpreting **pronouns**, such as **he**, **she** and **it**. Vital for many other tasks, such as automatic Question-Answering.

Understanding pronouns

- Question-Answering needs to use data like Wikipedia:

Richard Branson

From Wikipedia, the free encyclopedia

Sir Richard Charles Nicholas Branson (born 18 July 1950) is an English business magnate, investor and philanthropist.[4] He founded the Virgin Group, which controls more than 400 companies.[5]

Branson expressed his desire to become an entrepreneur at a young age. At the age of sixteen his first business venture was a magazine called *Student*.[6] In 1970, he set up a mail-order record business. In 1972, he opened a chain of record stores, Virgin Records, later known as Virgin Megastores. Branson's Virgin brand grew rapidly during the 1980s, as he set up Virgin Atlantic airline and expanded the Virgin Records music label.

- Who founded the Virgin Group?
- When did Richard Branson found Virgin Atlantic?
- What was Branson’s first business venture?

Understanding pronouns

- Question-Answering needs to use data like Wikipedia:

Richard Branson

From Wikipedia, the free encyclopedia

Sir Richard Charles Nicholas Branson (born 18 July 1950) is an English business magnate, investor and philanthropist.[4] He founded the Virgin Group, which controls more than 400 companies.[5]

Branson expressed his desire to become an entrepreneur at a young age. At the age of sixteen his first business venture was a magazine called *Student*.[6] In 1970, he set up a mail-order record business. In 1972, he opened a chain of record stores, Virgin Records, later known as Virgin Megastores. Branson's Virgin brand grew rapidly during the 1980s, as he set up Virgin Atlantic airline and expanded the Virgin Records music label.

- **Who founded the Virgin Group?**
- When did Richard Branson found Virgin Atlantic?
- What was Branson’s first business venture?

Understanding pronouns

- Question-Answering needs to use data like Wikipedia:

Richard Branson

From Wikipedia, the free encyclopedia

Sir Richard Charles Nicholas Branson (born 18 July 1950) is an English business magnate, investor and philanthropist.[4] He founded the Virgin Group, which controls more than 400 companies.[5]

Branson expressed his desire to become an entrepreneur at a young age. At the age of sixteen his first business venture was a magazine called *Student*.[6] In 1970, he set up a mail-order record business. In 1972, he opened a chain of record stores, Virgin Records, later known as Virgin Megastores. Branson's Virgin brand grew rapidly during the 1980s, as he set up Virgin Atlantic airline and expanded the Virgin Records music label.

- Who founded the Virgin Group?
- **When did Richard Branson found Virgin Atlantic?**
- What was Branson's first business venture?

Understanding pronouns

- Question-Answering needs to use data like Wikipedia:

Richard Branson

From Wikipedia, the free encyclopedia

Sir Richard Charles Nicholas Branson (born 18 July 1950) is an English business magnate, investor and philanthropist.[4] He founded the Virgin Group, which controls more than 400 companies.[5]

Branson expressed his desire to become an entrepreneur at a young age. At the age of sixteen his first business venture was a magazine called *Student*.[6] In 1970, he set up a mail-order record business. In 1972, he opened a chain of record stores, Virgin Records, later known as Virgin Megastores. Branson's Virgin brand grew rapidly during the 1980s, as he set up Virgin Atlantic airline and expanded the Virgin Records music label.

- Who founded the Virgin Group?
- When did Richard Branson found Virgin Atlantic?
- **What was Branson's first business venture?**

And what about …

- Did Branson enter the Australian aviation market? When?

The Sydney Morning Herald | Business

News & Views Markets Quotes Workplace Consumer Property Innovatio

The year was 2001, one year after the legendarily brash, anti-establishment renegade had pinpointed an opportunity to enter the Australian aviation market.Branson did so against the backdrop of a corporate landscape littered with the carcasses of other airline aspirants such as Compass and Impulse, all of whom had tried and failed to get a foothold in a duopoly market locked up the then government-owned Qantas and the Singapore Airlines-Air New Zealand-controlled Ansett.

Discourse (Pragmatics)

- Meaning beyond the sentence
- **Context-sensitivity**
  - (of meaning: not to be confused with context-sensitive grammars)
- In particular:
  - **Anaphora** (including temporal)
    - "He sold them to himself two days after that."
  - **Co-reference**
    - "The legendarily brash, anti-establishment renegade …"
  - **Ellipsis**
    - "She promised she would; but she didn't."
- Very important in semantic tasks (e.g. IE, QA)
  - Less important for others (e.g. document/sentiment classification) – **why?**
- **Implicature:**
    - Could you pass the salt? Do you have the time?
    - There's a garage round the corner
    - His handwriting is very neat.

Referring Expressions

- Five types of **referring expression (RE)**:
  - **Indefinite noun phrases** introduce *new* referents
    - “I saw a nice car today. Some other people noticed it too”
  - **Definite noun phrases** refer to *identifiable* referents
    - Sometimes identifiable from previous mention
      - “I saw a nice car today. I’d like to buy the car tomorrow.”
    - Sometimes identifiable from world knowledge or description itself:
      - “The Prime Minister is coming to tea.”
      - “The car of the year is the Ford Dustpan.”
  - **Pronouns** refer to highly *salient* referents
    - “I saw a nice car today. I’d like to buy it tomorrow.”
    - “She’s coming to tea.”
  - **Demonstratives** refer to *near* or *distant* referents (literally or metaphorically)
    - “This new car is much faster than that old one.”
  - **Names** can refer to *old or new* referents
    - “The Prime Minister will visit President Trump today, Mr Johnson’s office announced.”

Pronouns (Anaphora)

- Pronouns refer to contextual elements
- **Anaphora**: referring back to items already mentioned
  - (**antecedents**: previous referring expressions)
  - “Sue left her coat behind”
- **Cataphora**: referring forward to items to be mentioned
  - “Before she leaves, Sue always checks her coat”
- **Deixis**: referring to the current environment
  - “I want you to leave now”
- **Pleonastic** or **generic** uses:
  - “It’s raining. Always happens when you least expect it.”
- **Discourse** and **temporal** reference:
  - “We’ll be ready then. Well, that’s good.”
- **Bound** variables:
  - “Every student has their particular preference.”

Pronoun Resolution: Constraints

- For referential (non-generic) uses, we must identify (**resolve**) the antecedent
  - huge potential ambiguity: any previous RE is a candidate antecedent
- Many hard constraints on possible antecedents:
  - **Number**
    - (English: singular vs plural)
    - "I have a dog and two cats. It is/They are very fluffy."
  - **Person**
    - (English: 1^{st}, 2^{nd}, 3^{rd})
    - "I saw you with John. You were / He was happy."
  - **Gender**
    - (English: male, female, nonpersonal)
    - "Sue met John and his dog. She/he/it was happy."
  - **Binding**
    - (English: reflexives with clause subjects)
    - "John thinks Bill likes him/himself."
- (We often don't even notice these potential ambiguities, as humans!)

Pronoun Resolution: Preferences

- And many softer preferences:
  - **Recency**
    - (more recent > less recent)
    - “Sue lives in Reading. Jane lives in Havant. She has a dog.”
  - **Grammatical role**
    - (subject > object > other)
    - “Sue knows Jane. She is coming today.”
  - **Repetition**
    - (more mentions > fewer mentions)
    - “Sue is a banker. She works in the City. Jane likes her. She drives a Jaguar.”
  - **Parallelism**
    - “Sue has known Jane since 1989. Gretel has known her since 1982.”
  - **Discourse / event semantics**
    - “Sue is annoyed with Jane. She spilt her drink.”
    - “Sue loves her dog. We took her for a walk.”
- (We are more likely to notice these ambiguities)

Pronoun Resolution

- So we can use these constraints in resolution
- Rule-based:
  - e.g. using **Centering Theory** (Grosz et al, 1995)
  - Order REs and possible antecedent REs by prominence
    - e.g. subject > object > other
    - “Forward-looking centres” (FLCs) = all REs in sentence
    - “Backward-looking centres” (BLCs) = all REs mentioned in previous sentence
  - Filter possible pairings using hard constraints

- Statistical classification:
  - Supervised classification
  - Potential pronoun-antecedent pairs as instances
  - Features chosen to relate to constraints/preferences:
    - number, gender, person match
    - word/sentence/syntactic distance
    - grammatical/semantic role & parallelism

Contents

1) Discourse
2) Coreference (anaphora) resolution
3) Real-world reference resolution
4) Ellipsis resolution
5) Discourse parsing

Co-Reference Resolution

- More general version of the problem
  - Build **chains** of all co-referring expressions

Victoria Chen, Chief Financial Officer 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 financial services company's president. It has been ten years since she came to Megabucks from rival Lotsabucks.

Victoria Chen, Chief Financial Officer 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 financial services company's president. It has been ten years since she came to Megabucks from rival Lotsabucks.

Co-Reference Resolution

- More general version of the problem
  - Mentions **evoke (introduce)** or **access** referents:

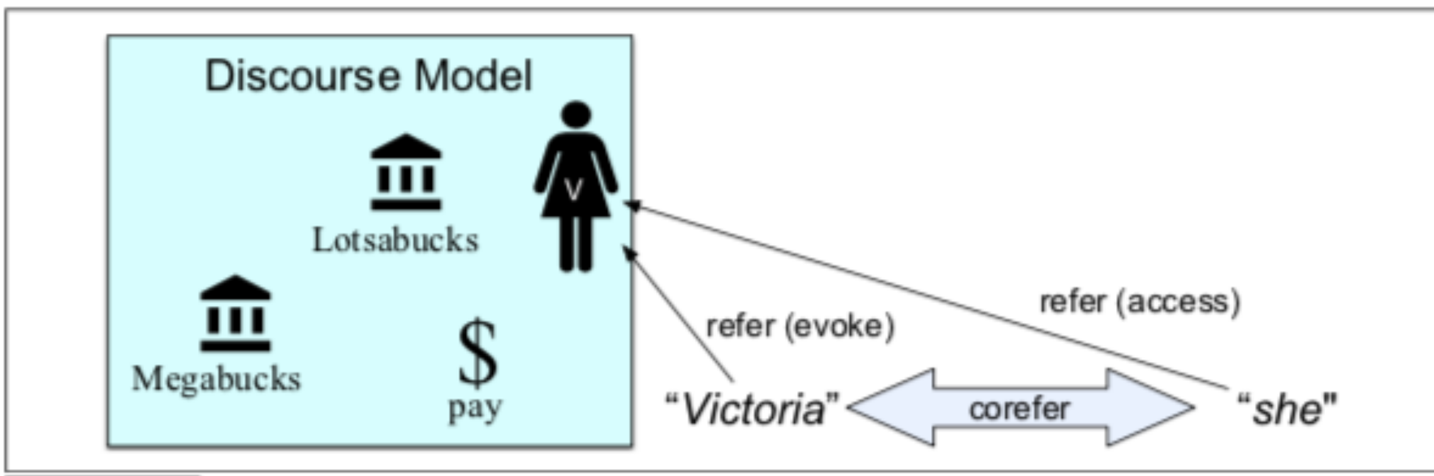


Figure 22.1 How mentions evoke and access discourse entities in a discourse model.

Co-Reference Resolution

- Candidate REs identified by POS-tagging & parsing
  - Need to identify links i.e. pair REs with antecedents
- Approach similar to pronoun resolution:
  - But we now need to add other RE types
  - Particularly definite noun phrases
- Supervised classification:
  - (same approach as for pronoun resolution)
  - Pairs of correct/incorrect examples
  - Features:
    - Features used for pronoun resolution, plus:
    - Lexical similarity (e.g. RE/antecedent edit distance)
    - Semantic co-reference (e.g. dates)
    - Syntactic relation (e.g. apposition (“X, the Y, …”)
    - POS/lexical type of head noun
  - Big search space: use e.g. graph-based methods
    - Efficient search, while finding optimal chains (not just independent pairs)

Anaphora resolution: the problem

Example

Sophia Loren says she will always be grateful to Bono. The actress revealed that the U2 singer helped her calm down when she became scared by a thunderstorm while travelling on a plane.

- *she* $\Rightarrow$ *Sophia Loren*
- *the actress* $\Rightarrow$ *Sophia Loren*
- *the U2 singer* $\Rightarrow$ *Bono*
- *her* $\Rightarrow$ *Sophia Loren*
- *she* $\Rightarrow$ *Sophia Loren*

Anaphora resolution: coreference chains

Example

Sophia Loren says she will always be grateful to Bono. The actress revealed that the U2 singer helped her calm down when she became scared by a thunderstorm while travelling on a plane.

Coreference Chains:

- {Sophia Loren, she, the actress, her, she}
- {Bono, the U2 singer }
- {a thunderstorm}
- {a plane}

Anaphora resolution as Structure Learning

- So far we have only seen examples of text analytics applications in which the task was to label a SINGLE OBJECT
- In the case of anaphora resolution/coreference, the task is to label a STRUCTURE
    - In its simplest form, the antecedent / anaphor pair (**MENTION PAIR**)
- This is an example of so-called STRUCTURED LEARNING

Factors in interpreting anaphoric expressions

- Factors:
  - Morphological features (agreement)
  - Syntactic information
  - Salience
  - Lexical and commonsense knowledge
- Distinction often made between CONSTRAINTS and PREFERENCES

Agreement

- GENDER strong CONSTRAINT for pronouns (in other languages: for other anaphors as well)
  - [Jane] blamed [Bill] because HE spilt the coffee (Ehrlich, Garnham e.a, Arnold e.a)
- NUMBER also strong constraint
  - [[Union] representatives] told [the CEO] that THEY couldn’t be reached

Lexical and commonsense knowledge

[The city council] refused [the women] a permit because <u>they</u> feared violence.
[The city council] refused [the women] a permit because <u>they</u> advocated violence.
Winograd (1974), Sidner (1979)

BRISBANE – a terrific right rip from [Hector Thompson] dropped [Ross Eadie] at Sandgate on Friday night and won <u>him</u> the Australian welterweight boxing title. (Hirst, 1981)

Problems to be resolved by an AR system: mention identification

- Effect: recall
- Typical problems:
    - Nested NPs (possessives)
        - [a city] 's [computer system] → [[a city]'s computer system]
    - Appositions:
        - [Madras], [India] → [Madras, [India]]
    - Attachments

Problems for AR: agreement extraction

- The committee are meeting / is meeting
- The Union sent a representative. They ….
- The doctor came to visit my father. SHE told him …

Problems for AR: anaphoricity determination

- Expletives:
  - <u>IT's</u> not easy to find a solution
  - Is <u>THERE</u> any reason to be optimistic at all?
- Non-anaphoric definites

Problems for AR: Complex attachments

- [The quality that's coming out of [software from [India]]
  - The quality that's coming out of software from India is now exceeding the quality of software that's coming out from the United States
- scanning through millions of lines of computer code
    - ACE/bnews/devel/ABC19981001.1830.1257

Early systems

- Hobbs 1976 Naïve Algorithm
  - Pronouns only
  - Syntax based
  - Still very competitive
- Sidner 1979
- Carter 1986

Modern Work in AR

- Availability of the first anaphorically annotated corpora circa 1993 (MUC6) made statistical methods possible.
- Most current anaphora resolution systems are based on machine learning, but there is one notable exception, the Stanford Coreference system.

MUC

- First big initiative in Information Extraction.
- Produced first sizeable annotated data for coreference.
- Developed first methods for evaluating systems.

MUC terminology:

- **MENTION:** any markable
- **COREFERENCE CHAIN:** a set of mentions referring to an entity
- **KEY:** the (annotated) solution (a partition of the mentions into coreference chains)
- **RESPONSE:** the coreference chains produced by a system

The Stanford Deterministic Coreference Resolution System

- Part of the Stanford CORE Pipeline
- The best-performing system at CONLL 2011, and used as a component by two of the top three systems at CONLL 2012
- Key to its performance are
  - A very high quality mention detection component based on the Stanford CORE pipeline
  - A PRECISION-FIRST coreference resolution component based on 10 filters called **SIEVES** that implement many of the restrictions on anaphora resolution discussed in previous slides

The Sieves

1. ***Speaker Identification*:** This sieve first identifies speakers, then matches first and second pronouns to these speakers.
2. ***Exact Match*:** This sieve links together two mentions only if they contain exactly the same text, including both determiners and modifiers.
3. ***Relaxed String Match*:** This sieve links together two mentions only if they con- tain exactly the same text after dropping the postmodifiers.
4. ***Precise Constructs*:** This sieve links together two mentions if they occur in one of a series of high precision constructs: e.g., if they are in an appositive con- struction (*[the speaker of the House], [Mr. Smith]* . . .), or if both mentions are tagged as NNP and one of them is an acronym of the other.

The Sieves

5. ***Strict Head Match*: This sieve links together a mention with a candidate antecedent entity if *all* of a number of constraints are satisfied: (a) the head of the mention matches any of the heads of the candidate antecedent; (b) all non-stop words of the mention are included in the non-stop words of the candidate antecedent; (c) all mention modifiers are included among the modifiers of the candidate antecedent; and (d) the two mentions are not in an i-within-i situation, i.e., one is not a child in the other.
6. ***Variants of Strict Head Match*:** Sieve 6 relaxes the 'compatible modifiers only' constraint in the previous sieve, whereas Sieve 7 relaxes the 'word inclusion' constraint.
7. ***Proper Head Match*:** This sieve links two proper noun mentions if their head words match and a few other constraints apply.
8. ***Relaxed Head Match*:** This sieve relaxes the requirement that the head word of the mention must match a head word of the candidate antecedent entity.
9. ***Pronoun resolution*:** Finally, pronouns are resolved, by finding candidates matching the pronoun in number, gender, person, animacy, and NER label, and at most 3 sentences distant.

Statistical approaches to AR

- **UNSUPERVISED** approaches
  - Eg Cardie & Wagstaff 1999, Ng 2008
- **SUPERVISED** approaches
  - Early (NP type specific)
  - Soon et al: general classifier + modern architecture

Soon et al 2001

- First ‘modern’ ML approach to anaphora resolution
    - Resolves ALL anaphors
    - Fully automatic mention identification
- Developed instance generation & decoding methods used in a lot of work since

Soon et al: AR as a Classification Problem

1. Classify MENTION PAIR <NP1,NP2> as coreferential or not
2. Build a complete coreferential chain

Soon et al: Mention Pairs

<ANAPHOR (j), ANTECEDENT (i)>

Soon et al: Mention Pairs

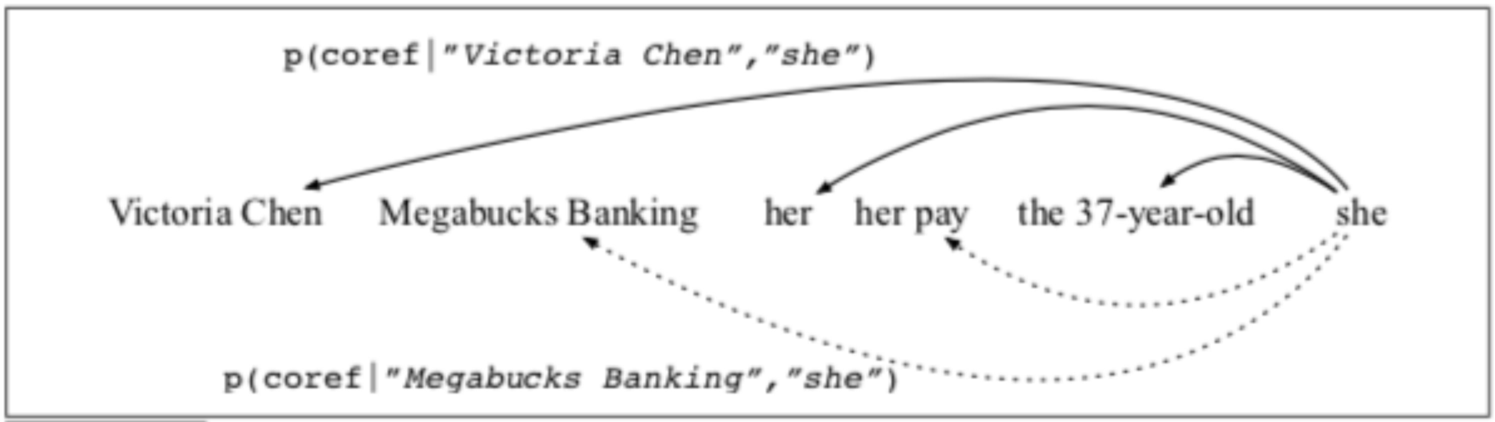


Figure 22.2 For each pair of a mention (like *she*), and a potential antecedent mention (like *Victoria Chen* or *her*), the mention-pair classifier assigns a probability of a coreference link.

Soon et al: Key Decisions

- ENCODING
  - I.e., what positive and negative instances to generate from the annotated corpus?
    - E.g. treat all elements of the coref chain as positive instances, everything else as negative.
- DECODING
  - How to use the classifier to choose an antecedent?
    - Some options: ‘sequential’ (stop at the first positive), ‘parallel’ (compare several options)

Soon et al: preprocessing

- Part-of-Speech (POS) tagger:
  - HMM-based
  - 96% accuracy
- Noun phrase identification module:
  - HMM-based
  - Can identify correctly around 85% of mentions.
- Named Entity Recognition (NER): reimplementation of Bikel Schwartz and Weischedel 1999
  - HMM based
  - 88.9% accuracy

Soon et al: Features

- NP type
- Distance
- Agreement
- Semantic class

Soon et al: NP type and distance

NP type of anaphor j (3)

`j-pronoun, def-np, dem-np (bool)`

NP type of antecedent i

`i-pronoun (bool)`

Types of both

`both-proper-name (bool)`

DIST

`0, 1, ….`

Soon et al features: string match, agreement, syntactic position

STR_MATCH
ALIAS

```
dates   (1/8 – January 8)
person  (Bent Simpson / Mr. Simpson)
organizations: acronym match
             (Hewlett Packard / HP)
```

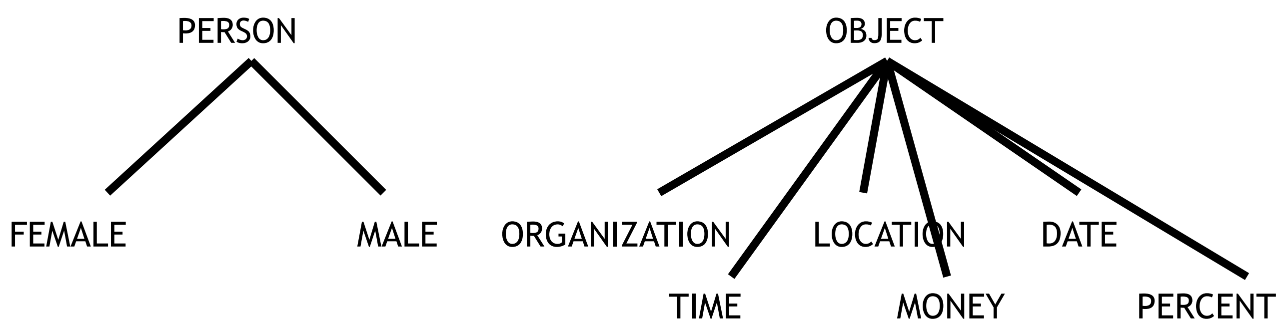
AGREEMENT FEATURES

```
number agreement
gender agreement
```

SYNTACTIC PROPERTIES OF ANAPHOR

```
occurs in appositive construction
```

Soon et al: semantic class agreement

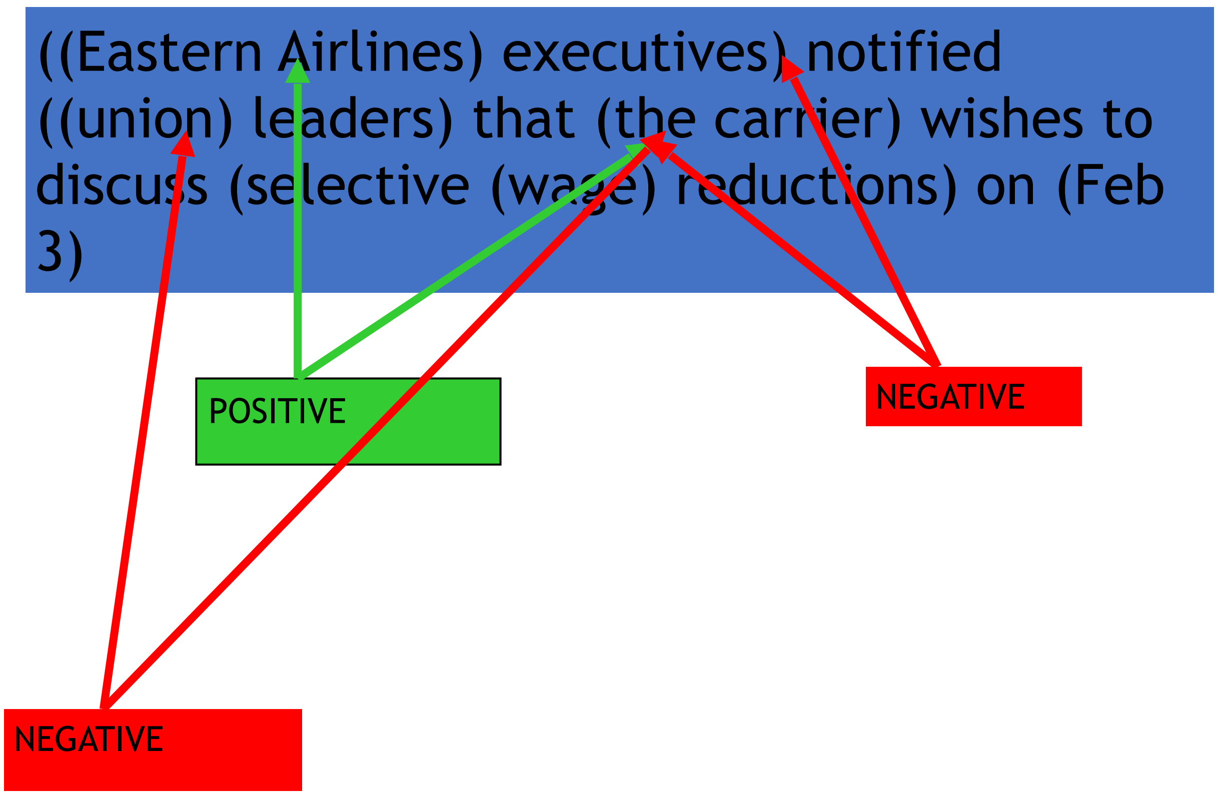


SEMCLASS = true iff semclass(i) <= semclass(j) or viceversa

Soon et al: generating training instances

- Marked antecedent used to create positive instance
- All mentions between anaphor and marked antecedent used to create negative instances

Soon et al: generating training instances



Soon et al: decoding

- Right to left, consider each antecedent until classifier returns true

Soon et al: evaluation

- MUC-6:
    - P=67.3, R=58.6, F=62.6
- MUC-7:
    - P=65.5, R=56.1, F=60.4

Soon et al: evaluation

```
STR_MATCH = +: +
STR_MATCH = -:
:...J_PRONOUN = -:
    :...APPOSITIVE = +: +
    :   APPOSITIVE = -:
    :   :...ALIAS = +: +
    :       ALIAS = -: -
    J_PRONOUN = +:
    :...GENDER = 0: -
        GENDER = 2: -
        GENDER = 1:
        :...I_PRONOUN = +: +
            I_PRONOUN = -:
            :...DIST > 0: -
                DIST <= 0:
                :...NUMBER = +: +
                    NUMBER = -: -
```

Evaluation of coreference resolution systems

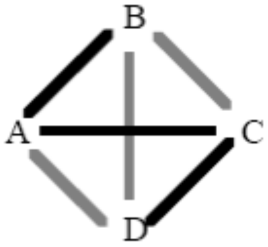
- Lots of different measures proposed
- ACCURACY:
    - Consider a mention correctly resolved if
        - Correctly classified as anaphoric or not anaphoric
        - ‘Right’ antecedent picked up
- Measures developed for the competitions:
    - Automatic way of doing the evaluation
- More realistic measures (Byron, Mitkov)
    - Accuracy on ‘hard’ cases (e.g., ambiguous pronouns)

Vilain et al 1995: Evaluation

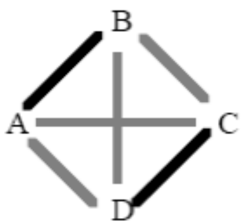
- The official MUC scorer
- Based on precision and recall of links

Vilain et al: the goal

The problem: given that A,B,C and D are part of a coreference chain in the KEY, treat as equivalent the two responses:



And as superior to:



Vilain et al: Recall

- To measure RECALL, look at how each coreference chain S_i in the KEY is partitioned in the RESPONSE, and count how many links would be required to recreate the original, then average across all coreference chains.

$$R_T = \frac{\sum\left(|S_i| - |p(S_i)|\right)}{\sum\left(|S_i| - 1\right)}$$

Vilain et al: Example recall

- In the example above, we have one coreference chain of size 4 ($|S| = 4$)
- The incorrect response partitions it in two sets ($|p(S)| = 2$)
- R = 4-2 / 4-1 = 2/3

Vilain et al: Precision

- Count links that would have to be (incorrectly) added to the key to produce the response
- I.e., ‘switch around’ key and response in the equation before

Beyond Vilain et al

- Problems:
  - Only gain points for links. No points gained for correctly recognizing that a particular mention is not anaphoric
  - All errors are equal
- Proposals:
  - Bagga & Baldwin’s B-CUBED algorithm
  - Luo recent proposal

Error analysis (Soon et al, 2001)

- Errors most affecting precision:
  - Prenominal modifiers identified as mentions and other errors in mention identification
  - String match but noun phrases refer to different entities
- Errors most affecting recall:
  - Errors in mention identification (11%)
  - Errors in SEMCLASS determination (10%)
  - Need more features (63.3%)

Soon et al examples of errors:

- Tarnoff, a former Carter administration official and president of the Council on foreign relations, is expected to be named **[undersecretary]** for political affairs … Former. Sen Tim Wirth is expected to get a newly created **[undersecretary]** post for global affairs
- [Ms Washington and Mr. Dingell] have been considered [allies] of **[the Securities exchanges],** while [banks] and **[future exchanges]** often have fought with THEM

After Soon et al 2001

- Different models of the task
- Different preprocessing techniques
- Using lexical / commonsense knowledge (particularly semantic role labelling)
- Salience
- Anaphoricity detection
- Development of AR toolkits (GATE, LingPipe, GUITAR)

Mention detection errors in GUITAR (Kabadjov, 2007)

[The bow] (see detail, below right) is decorated with a complicated arrangement of horses and lions' heads.

Above the lions' heads are four sphinxes.

Three pairs of lions clamber up the section from the point where [the sheath and bow] are joined.

More recent models

- Cardie & Wagstaff: coreference as (unsupervised) **clustering**
  - Much lower performance
- **Ranking** models:
  - Ng and Cardie 2002
  - Yang ‘twin-candidate’ model
- **Entity-mention** models
- **Joint entity detection & tracking**

Ranking models

- Idea: train a model that imposes a **ranking** on the candidate antecedents for an NP to be resolved so that it assigns the highest rank to the correct antecedent
- A ranker allows all candidate antecedents to be considered simultaneously and captures competition among them:
    - Allows us find the best candidate antecedent for an NP.
- There is a natural resolution strategy for a ranking model:
    - An NP is resolved to the highest-ranked candidate antecedent

How to train a ranking model

- Convert the problem of ranking *m* NPs into the a set of pairwise ranking problems
  - Each pairwise ranking problem involves determining which of two candidate antecedents is better for an NP to be resolved
    - Each one is essentially a classification problem

- Ranking rediscovered independently by
  - Yang et al. (2003) (twin-candidate model)
  - Iida et al. (2003) (tournament model)

- Denis & Baldridge (2007, 2008): train the ranker using maximum entropy
  - model outputs a rank value for each candidate antecedent

Entity-mention models

- Classifiers that determine whether (or how likely) an NP belongs to a preceding **coreference cluster**
- more **expressive** than the mention-pair model
  - can employ cluster-level features defined over any subset of NPs in a preceding cluster

Joint Entity Detection and Tracking

- Daume and Marcu 2005: Mention identification, classification, and linking take place at the same time
- Denis and Balridge 2007: Integer Linear Programming (ILP)

Recent state of the art in coreference: the 2012 CONLL Shared Task

- Data: OntoNotes
  - 1.6M words English, 900K words Chinese, 300K words Arabic
  - Annotated with: syntactic information, wordsenses, propositional information
- Tracks:
  - Closed
  - Open
- Metrics: MELA
  - (a combination of MUC / B3 / CEAF)

CONLL 2012 Shared Task: Results

Participant	Open			Closed			Official	Final model	
	English	Chinese	Arabic	English	Chinese	Arabic	Score	Train	Dev
fernandes				63.37	58.49	54.22	58.69	√	√
björkelund				61.24	59.97	53.55	58.25	√	√
chen		63.53		59.69	62.24	47.13	56.35	√	×
stamborg				59.36	56.85	49.43	55.21	√	√
uryupina				56.12	53.87	50.41	53.47	√	√
zhekova				48.70	44.53	40.57	44.60	√	√
li				45.85	46.27	33.53	41.88	√	√
yuan		61.02		58.68	60.69		39.79	√	√
xu				57.49	59.22		38.90	√	×
martschat				61.31	53.15		38.15	√	×
chunyang				59.24	51.83		37.02	–	–
yang				55.29			18.43	√	×
chang				60.18	45.71		35.30	√	×
xinxin				48.77	51.76		33.51	√	√
shou				58.25			19.42	√	×
xiong	59.23	44.35	44.37				0.00	√	√

Anaphora/Co-reference datasets/annotation tools:

- MUC6/MUC7 (small, old)
- ACE 2002/2005
- ONTONOTES
- Phrase Detectives (locally developed at QM and Essex) http://anawiki.essex.ac.uk/phrasedetectives
    - **Have a go at Phrase Detectives.**

Toolkits for AR

- BART (Versley et al, 2008).
- Stanford Deterministic Coreference Resolver (Lee et al 2013).
- CORT (Martschat & Strube 2015).

Contents

1) Discourse
2) Coreference (anaphora) resolution
3) Real-world reference resolution
4) Ellipsis
5) Discourse parsing

What about the physical world of referents?

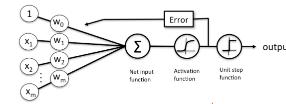
- How do process reference to real-world objects, say in a photograph?
- Recently this has become possible through improvements in **computer vision (CV).**
- One example is the words-as-classifiers model of reference res. (Kennington and Schlangen, 2015)
- Trains logistic regression classifiers for each word.
  - Input: word, CV features of object
  - Output: probability the word refers to the object
- Simple combination of classifiers gets surprisingly good results (65-70% F-score).

Words-As-Classifiers (WAC) Reference Resolution

Kennington and Schlangen, ACL 2015

Number of edges	12
Position.x	148
Position.y	198
Orientation	-9.364777
HSV.H	90.746
HSV.S	102.035
HSV.V	153.105
RGB.R	145.572
RGB.G	40.644
RGB.B	41.3999
…	…

Logistic Regression Classifier for each word



P('red'|obj_1)= 0.98

P('cross'|obj_1)= 0.90

P('blue'|obj_1)= 0.25

Words-As-Classifiers (WAC) Reference Resolution

Kennington and Schlangen, ACL 2015

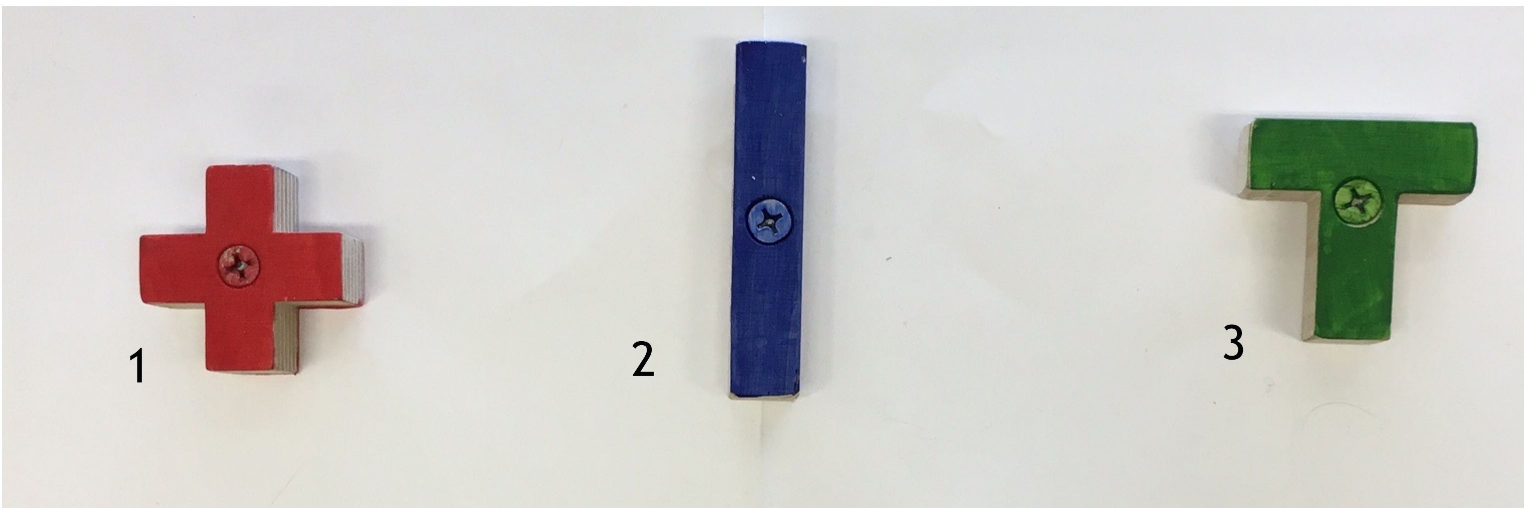
'the red cross'

P('the red cross'|obj_1) = 1/3 * [P('the'|obj_1) + P('red'|obj_1) + P('cross'|obj_1)]
= 1/3 * [0.6 + 0.9 + 0.9]
= **0.8**

Words-As-Classifiers (WAC) Reference Resolution

Kennington and Schlangen, ACL 2015

‘the red cross’



P(‘the red cross’|obj_1) = 0.8

P(‘the red cross’|obj_2) = 0.2

P(‘the red cross’,|obj_3) = 0.4

Normalize for reference resolution:

P(obj_1|‘the red cross’) = **0.57**

P(obj_2|‘the red cross’) = **0.14**

P(obj_3|‘the red cross’) = **0.28**

Contents

1) Discourse
2) Coreference (anaphora) resolution
3) Real-world reference resolution
4) Ellipsis
5) Discourse parsing

And what about …

- Did Branson enter the Australian aviation market? When?

The Sydney Morning Herald | Business

News & Views | Markets | Quotes | Workplace | Consumer | Property | Innovatio

The year was 2001, one year after the legendarily brash, anti-establishment renegade had pinpointed an opportunity to enter the Australian aviation market.Branson did so against the backdrop of a corporate landscape littered with the carcasses of other airline aspirants such as Compass and Impulse, all of whom had tried and failed to get a foothold in a duopoly market locked up the then government-owned Qantas and the Singapore Airlines-Air New Zealand-controlled Ansett.

Verb Phrase Ellipsis

- "Branson **did so**"
- Similar problem, but with some important differences
- Identifying ellipsis sites is harder:
  - "I can't today but I will try tomorrow"
  - "I can't today but I will – try tomorrow"
  - Use lexical, POS sequence, syntactic tree features
- Resolving ellipsis is more complex:
  - Not a purely lexical or syntactic operation:
    - "John <u>likes tennis</u>. So <u>do</u> you."
  - But what's the semantic operation?

$$like(john, tennis) \dashrightarrow like(you, tennis)$$

  - Can use lambda calculus to generate possible antecedents
    - abstract subject:

$$like(john, tennis) \dashrightarrow \lambda x.like(x, tennis)$$

  - Resolve VPE sites to semantic antecedent functions

$$P(you) \dashrightarrow P=\lambda x.like(x, tennis) \dashrightarrow like(you, tennis)$$

Ellipsis & Ambiguity

- Unfortunately, this brings in ambiguity again:
  - Sometimes that's appropriate:
  - "Sue thinks John likes Jane. So does Bill."

$$think(sue, like(john, jane)) \quad \text{-->} \quad \lambda x.think(x, like(john, jane))$$

$$think(sue, like(john, jane)) \quad \text{-->} \quad \lambda x.like(x, jane)$$

  - But sometimes some are spurious:
  - "John likes his teacher and his classmates. Sue does too."

$$like(john, teacher(john) \wedge mates(john)) \quad \text{-->} \quad \lambda x.like(x, teacher(x) \wedge mates(x))$$

$$like(john, teacher(john) \wedge mates(john)) \quad \text{-->} \quad \lambda x.like(x, teacher(john) \wedge mates(john))$$

$$like(john, teacher(john) \wedge mates(john)) \quad \text{-->} \quad \lambda x.like(x, teacher(x) \wedge mates(john))$$

$$like(john, teacher(john) \wedge mates(john)) \quad \text{-->} \quad \lambda x.like(x, teacher(john) \wedge mates(x))$$

- Need ambiguity resolution e.g. machine learning using suitable features:
  - Parallelism (e.g. subject-subject)
  - Discourse coherence (e.g. from Rhetorical Structure Theory)
  - Semantic plausibility & event/role restrictions

Perceived vs Potential Ambiguity

The Prime Minister announced today that the government intends to trigger Article 50 by the end of this month.

Our analyst George Snowden believes <u>she</u> may <u>do so</u> this week.

- Perceived ambiguity:
  - Is it clear who “she” is?
  - And what “do so” means?
- How many potential antecedents are there for:
  - “she”
  - “do so”?

Extreme Ellipsis: Dialogue

- *British National Corpus KSP 389-393:*

Christine	What have you been up to?
Steve	Nothing.
Michael	Eating.
Leslie	Any phone calls?
Steve	Nah.

- How could we summarise this dialogue?
  - *e.g. C asked what the others had been up to; S said he hadn't been doing anything, M said he'd been eating. L asked whether there had been any phone calls; S said there hadn't been any.*
- ("Summary" is longer than the dialogue …)

Contents

1) Discourse
2) Coreference (anaphora) resolution
3) Real-world reference resolution
4) Ellipsis resolution
5) Discourse parsing

Discourse Structure

- We can assign structure to whole texts

- Role relations between clauses/sentences
    - “Discourse parsing” (e.g. Marcu, 2012)
    - e.g. Rhetorical Structure Theory (Mann & Thompson, 1993)

- (Not core part of this course, though next week **Dialogue Acts**)

Discourse Structure

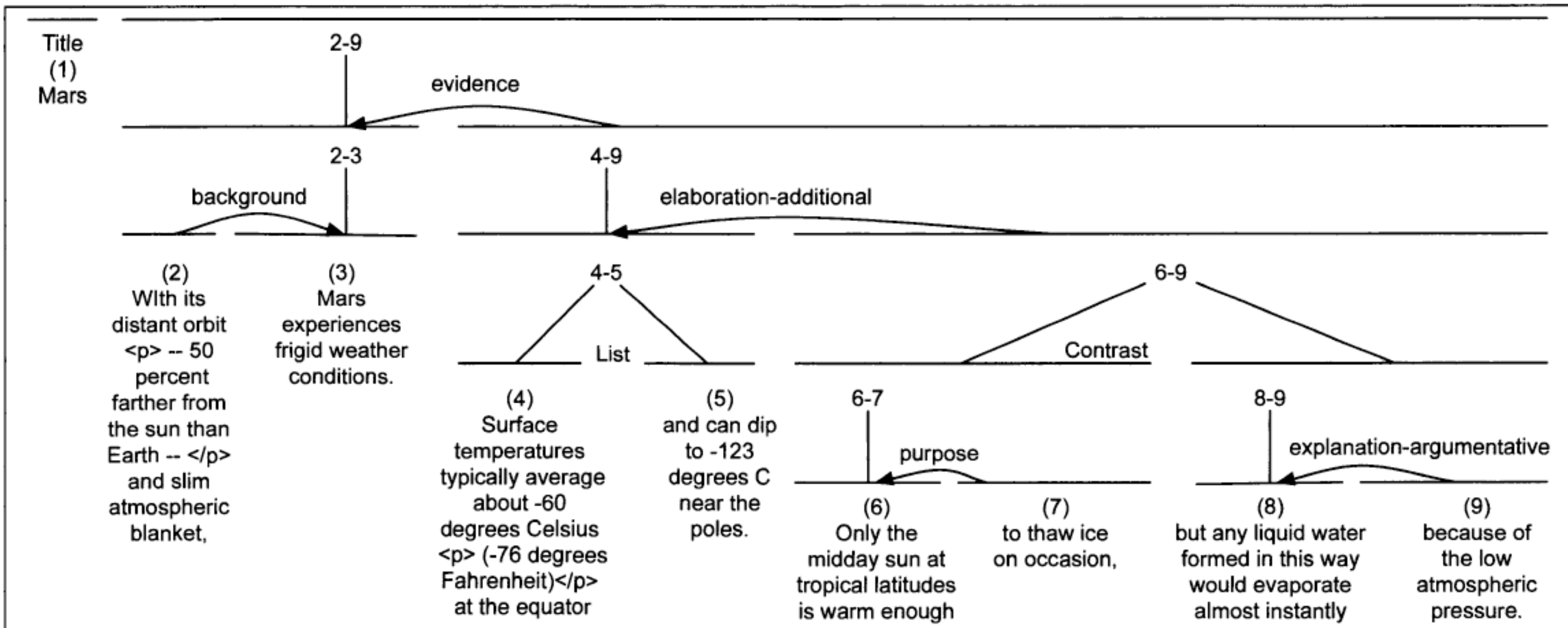


Figure 21.4 A discourse tree for the *Scientific American* text in (21.23), from Marcu (2000a). Note that asymmetric relations are represented with a curved arrow from the satellite to the nucleus.

Reading

- Jurafsky and Martin (3rd Ed.) Chapters 21 and 22
- Soon, W. M., Ng, H. T., and Lim, D. C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. Computational Linguistics, 27(4), 521–544.