Coreference Resolution

Asif Ekbal
IIT Patna
Email: asif@iitp.ac.in

COREFERENCE (OR ANAPHORA) RESOLUTION: THE TASK

IDENTIFYING WHICH **MENTIONS** REFER TO THE SAME (DISCOURSE) **ENTITY**

Contents

- **Definitions,** Issues and Corpora
- Machine learning models for coreference resolution: the standard (= mention pair) model

Introduction

What is

Anaphora

Antecedent

Anaphora Resolution

Sabeer Bhatia arrived at Los Angeles International Airport at 6 p.m. on September 23, 1998. **His** flight from Bangalore had taken 22hrs and **he** was starving.

[RD, NOV 2000]

Etymology of Anaphora

ANA-Back, Upstream, Back upstream

Phora- Act of Carrying

Anaphora-Act of Carrying Back

Chains of mentions in text (COREFERENCE CHAIN)

Toni Johnson pulls a tape measure across the front of what was once a stately Victorian home.
A deep trench now runs along its north wall, exposed when the house lurched two feet off its foundation during last week's earthquake.
Once inside, she spends nearly four hours measuring and diagramming each room in the 80-year-old house, gathering enough information to estimate what it would cost to rebuild it.
While she works inside, a tenant returns with several friends to collect furniture and clothing.
One of the friends sweeps broken dishes and shattered glass from a countertop and starts to pack what can be salvaged from the kitchen.

(WSJ section of Penn Treebank corpus)

WHY

- Original motivation (MUC, ACE): Information extraction
  - **Ang-2** blocks the ability of **Ang-1** to activate **Tie2** in ECs, but it activates **Tie2** expressed in hemangioblast
- Newer applications
  - Question answering (Morton, 2000)
  - Summarization (Steinberger et al, 2007)
- Other
  - Spoken dialogue systems
  - NLG
  - Machine Translation

Cataphora

- When “anphora” precedes the antecedent

  Because **she** was going to the departmental store, **Mary** was asked to pick up the vegetables.

Types of Anaphors

Pronominal anaphora

Vajpayee hits back forcefully when ***he*** told the opposition today "sometimes we fall prey to the media and sometimes you do.

[Indian Express 2001]

Possessive

MI owner Ambani wished all the best to Hardik for his stint with Gujrat Titans

[TOI, Feb 14 2022]

Types of Anaphors

Reflexive Pronoun

Finally, *Messi* decided himself to quit *Barcelona*

Demonstrative Pronoun

John had lots of ***packing*** to do before he shifted his house. ***This*** was something he never liked….

Relative Pronoun

Stumper ***Sameer Dige***, ***who*** made his test debut, failed to show fast reflexives when it mattered.

Types of Anaphors

- VP Anaphor

John screamed, as ***did*** Mary .

Non Anaphoric Usage of Pronouns

Pleonastic It (***refers to an occurrence of 'it' that does not refer to any entity and usually occurs in a sentence about state***)

Cognative

a. It is believed that…..

b. It appears that…..

Modal Adjectives

c. It is dangerous……

d. It is important…..

Temporal

e. It is five o'clock

f. It is winter

Weather verbs

g. It is raining

f. It is snowing

Distance

h. How far it is to Kolkata?

Non-anaphoric uses of pronouns

He that plants thorns must never expect to gather roses.

He who dares wins

Deictic: determination of referents depends on who says the sentence

I want him to come here now

I, here, him, **and** ***now*** **are deictic because the determination of their referents depends on who says that sentence, and where, when, and of whom it is said**

Noun Phrase Anaphora

Definite descriptions and Proper names

Roy Kaene has warned Manchester United ***he*** may snub their pay deal. ***United's skipper*** is even hinting that unless the future Old Trafford Package meets his demands, he could quit the club in June 2000. ***Irishman Keane***, 27, still has 17 months to run on his current 23,000 pound a week contract and wants to commit ***himself*** to United for life. ***Alex Ferguson's No 1 player*** confirmed: If it's not the contract ***I*** want, ***I*** won't sign".

Anaphora ≠ Coreference

- COREFERENT, not ANAPHORIC
  - two mentions of same object in different documents
- ANAPHORIC, not COREFERENT
  - identity of sense: John bought a shirt, and Bill got ONE, too
  - Dependence on non-referring expressions: EVERY CAR had been stripped of ITS paint

Chains of object mentions in text

Toni Johnson pulls a tape measure across the front of what was once a stately Victorian home. A deep trench now runs along its north wall, exposed when the house lurched two feet off its foundation during last week's earthquake.

Once inside, she spends nearly four hours measuring and diagramming each room in the 80-year-old house, gathering enough information to estimate what it would cost to rebuild it.

While she works inside, a tenant returns with several friends to collect furniture and clothing

One of the friends sweeps broken dishes and shattered glass from a countertop and starts to pack what can be salvaged from the kitchen

(WSJ section of Penn Treebank corpus)

RESEARCH ON ANAPHORA RESOLUTION: A QUICK SUMMARY

1970-1995

- Primarily theoretical
- Emphasis: commonsense knowledge, salience
- Exception: Hobbs 1977, Shalom Lappin

1995-2005

- First annotated corpora used to develop, evaluate and compare systems (MUC, Geand Charniak, ACE)
- First robust systems
- Heuristic-based: Mitkov
- ML: Vieira & Poesio 1998, 2000; Soon et al 2001, Ng and Cardie2002
- Emphasis: surface features
- Exceptions: Poesio & Vieira, Harabagiu, Markert

2005-present

- More sophisticated ML techniques (global models, kernels)
- Richer features –especially semantic information

Some well-known Tools

- GATE
- Java-RAP (pronouns)
- GUITAR (Poesio & Kabadjov, 2004; Kabadjov, 2007)
- BART (Versleyet al, 2008)
- CoreNLP suite
- LingPipe

Application of Anaphora Resolution

- Segmentation tasks: require determining the coherence of (segments of) text
- Summarization-
  - Post-hoc coherence check (Steinberger et al, 2007)
  - Sentence selection (Steinberger et al 2005, 2007)
- Tasks that require identifying the most important information in a text

- Information extraction: recognize which expressions refer to objects in the domain
- Relation extraction from biomedical text (Sanchez-Grailletand Poesio, 2006, 2007)
- Multimodal interfaces: recognize which objects in the visual scene are being referred to
- Machine translation

Different Approaches in Anaphora Resolution

- Rule based
- Statistical based
- Machine Learning based

Hard Constraints on Coreference

- Number agreement
- Person and case
- Gender Agreement
- Syntactic Agreement
- Selectional Restrictions

Number Agreement

Singular	Plural	Unspecified
She,her, he, him, his, it	We, us, they, them	you

John and Mary loaned Sue a cup of coffee. Little did they know the magnitude of her addiction.

Person and Case Agreement

	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 a coffee machine. She loves it.

Constraints (for English)

- **Syntactic Constraints**:
  - Syntactic relationships between a referring expression and a possible antecedent noun phrase
    - John bought himself a new car. [himself=John]
    - John bought him a new car. [him ≠ John]

- **Selection Restrictions**:
  - A verb places restrictions on its arguments
    - John parked his Tyoto in the garage. He had driven it around for hours. [it=Tyoto, it ≠ garage];
    - I picked up the book and sat in a chair. It broke. [it=Chair, it ≠ book]

Also: Preferences

- Recency
- Grammatical Role
- Repeated Mention
- Parallelism
- Verb Semantics

Based on Salience

Preferences in Pronoun Interpretation

- **Recency**:
  - Entities introduced recently are more salient than those introduced before.
    - Ram has a Maruti. Govind has a Hyundai. Anjali likes to drive it.

- **Grammatical Role**:
  - Entities mentioned in subject position are more salient than those in object position.
    - Ram went to the Maruti dealership with John. He bought an Hyundai. [he=Ram]

Preferences in Pronoun Interpretation

- **Repeated Mention**:
  - Entities that have been focused on in the prior discourse are more salient

Ram needed a car to get to his new job.

He decided that he wanted something sporty.

Govind went to the Maruti dealership with him.

He bought an Hyundai. [he=Ram]

Preferences in Pronoun Interpretation

- **Parallelism (more generally – discourse structure)**:
  - There are also strong preferences that appear to be induced by parallelism effects
    - Govind went with Anjali to the cinema. Reshma went with her to the mall. [her = Anjali]
    - Govind surprised Dhruv and then Surabhi shocked him. (*him* = Dhruv)

Preferences in Pronoun Interpretation

- **Verb Semantics**:
  - Certain verbs appear to place a semantically-oriented emphasis on one of their argument positions.
    - **Deepak** telephoned Shad. **He** had lost the book in the mall. [He = Deepak]
    - Dhruv criticized **Nalin**. **He** had lost the book in the mall. [He = Nalin]
    - Rahul praised **Sachin** because **he** … [he = Sachin]
    - **Dhruv** apologized to Nalini because **he**… [he = Dhruv]

Preferences in Pronoun Interpretation

- **World knowledge in general**:

  E.g. The city council denied the demonstrators a permit because they {feared|advocated} violence.

  - **The city council** denied the demonstrators a permit because **they** {**feared**|advocated} violence.

  - The city council denied **the demonstrators** a permit because **they** {feared|**advocated**} violence.

Chains of mentions in text (COREFERENCE CHAIN)

Toni Johnson pulls a tape measure across the front of what was once a stately Victorian home.
A deep trench now runs along its north wall, exposed when the house lurched two feet off its foundation during last week's earthquake.
Once inside, she spends nearly four hours measuring and diagramming each room in the 80-year-old house, gathering enough information to estimate what it would cost to rebuild it.
While she works inside, a tenant returns with several friends to collect furniture and clothing.
One of the friends sweeps broken dishes and shattered glass from a countertop and starts to pack what can be salvaged from the kitchen.

(WSJ section of Penn Treebank corpus)

Resources for coreference resolution

- Existing resources
  - MUC/ACE
  - MATE/GNOME/ARRAU

  - Ontonotes (Release 5.0 available)
  - AnaWiki (Anaphorically Annotated Resources Through Web Cooperation)

- Others
  - Penn Dependency TreeBank, VENEX, TueBa-D/Z
- Indian Languages: ICON-2011 Shared Task Datasets

Resources for coreference resolution

- **MATE** ‘meta-scheme’ for anaphora annotation (Poesio et al., 1999)
  - Developed as part of the MATE project (McKelvie et al., 2001), whose goal was to develop annotation tools suitable for different types of dialogue annotation
- **GNOME**
  - focused on identifying a subset of bridging relations that could be reliably annotated
- **ARRAU**
  - marking of ambiguous anaphoric expressions, and expressions which refer to abstract entities such as events, actions and plans

Resources for coreference resolution

- Different schemes for annotating coreference!
  - WHICH TYPES OF MENTIONS GET ANNOTATED
    - MUC, ACE: a subset of NPs
    - GNOME, ARRAU, OntoNotes: all NPS
  - WHICH TYPES OF RELATIONS
    - Mostly focusing on Identity relations
    - GNOME, ARRAU, also associative relations
    - OntoNotes: references to events
- Issues
  - Reliability
  - Ambiguity

Outline

- **Definitions, Issues and Corpora**
- **Machine learning models for coreference, 1**
  - **The standard (= mention pair) model**

Interpreting anaphoric expressions

- Interpreting (‘resolving’) an anaphoric expressions involves at least three aspects:
  1. Deciding whether the expression is in fact anaphoric
  2. Identifying its antecedent (possibly not introduced by a nominal)
  3. Determining its meaning (identity of sense vs. identity of reference)

(not necessarily taken in this order!)

A nominal is a word which differs grammatically from a noun but functions as one

In *the poor are many*, the word *poor* is a nominal. It functions as a noun; however, it does not pluralize.

Factors that affect the interpretation of anaphoric expressions

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

Binding

- Binding is the distribution of anaphoric elements
- Goal of binding theory is to identify the syntactic relationships that can or must hold between a given pronoun or noun and its antecedent

e.g.

1. <u>John</u> said <u>he</u> would help (Possible)

2. <u>He</u> said <u>John</u> would help (Not possible!)

Binding Theory (Chomsky, 1981)

- Principle A: Reflexives must have local antecedents

  John washed **himself**

  John asked Mary to wash **himself** (NOT)

- Principle B: Personal pronouns must not have local antecedents

  John asked Mary to wash **him**

  John washed **him** **(NOT)**

- Principle C: A referring expression can not have an antecedent that c-commands it

  He asked Mary to wash **John** **(NOT)**

  The car had a **trailer**, behind **it**

A c-commands B if and only if neither A dominates B nor B dominates A; and every branching node that dominates A also dominates B

Hand-coded algorithms for Coreference / anaphora

- Using lexical and commonsense knowledge
  - Charniak 72, Hobbs et al 1992
- All NPs
  - Webber 1978, Sidner 1979
- Pronouns
  - Syntax: Hobbs 1974
  - Salience: Sidner, Brennan et al, Strube, Tetreault
  - Lappin & Leass 1994
  - Heuristic: Mitkov 1998, Baldwin
- Definite descriptions
  - Vieira and Poesio 2000

Machine learning for coreference resolution

- Started in the mid-90s
  - Connolly et al. (1994)
  - Aone & Bennett (1995)
  - McCarthy & Lehnert (1995)
- Made possible and boosted by the availability of labeled data (MUC 6/7)
- Most of the work on **supervised learning**

Anaphora Resolution: A Classification Problem

- Classify NP1 and NP2 as coreferential or not
- Build a complete coreferential chain

Supervised learning for coreference resolution

- Learn a model of coreference from training labeled data
- need to specify
  - learning algorithm
  - feature set
  - clustering algorithm

Some 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)

Example

ACE-02, NWIRE APW19980213.1305:

Israel will ask the United States to delay a military strike against Iraq until **the Jewish state** is fully prepared for a possible Iraqi attack with non-conventional weapons, the defense minister said in remarks published Friday. [. . .] **Israel** is equipping **its residents** with gas masks and preparing kits with antidotes. [. . .] **Israel**'s armed forces chief, Lt. Gen. Amnon Lipkin-Shahak, appeared on national television Friday night in attempt to reassure **a jittery public**. “When events are concrete and actually unfolding, we will include **the entire population** in the measures we decide are appropriate to take,” Shahak said in an interview with Channel Two Television.

Supervised learning for coreference resolution

- Classification
  - Train a classifier to determine whether two mentions are *coreferent* or *not coreferent*

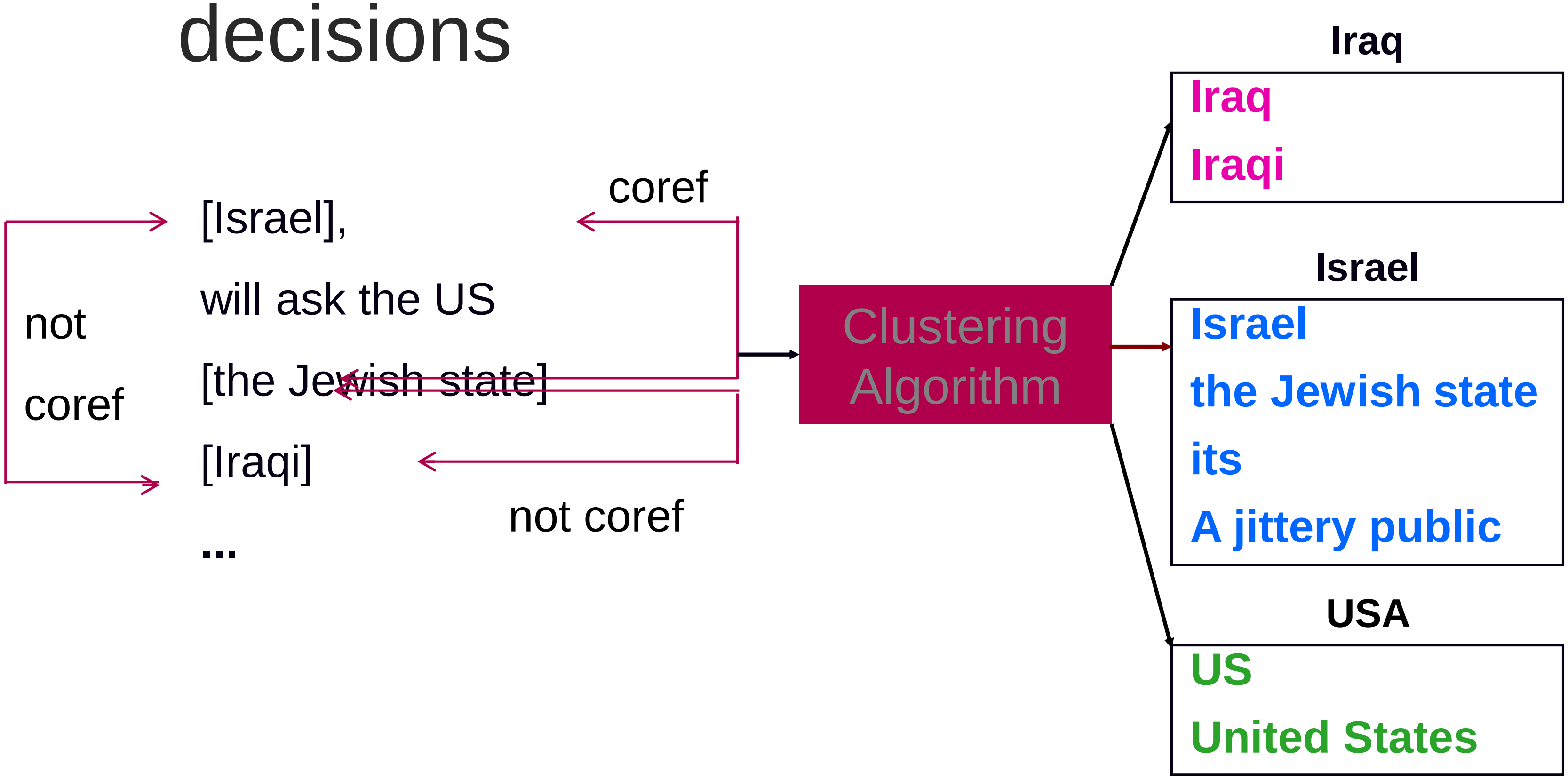
coref ? *coref ?*

[Israel] will ask the US to … [the Jewish state] … [Iraqi], ...

not coref ?

Supervised learning for coreference resolution

- Clustering pairwise coreference decisions

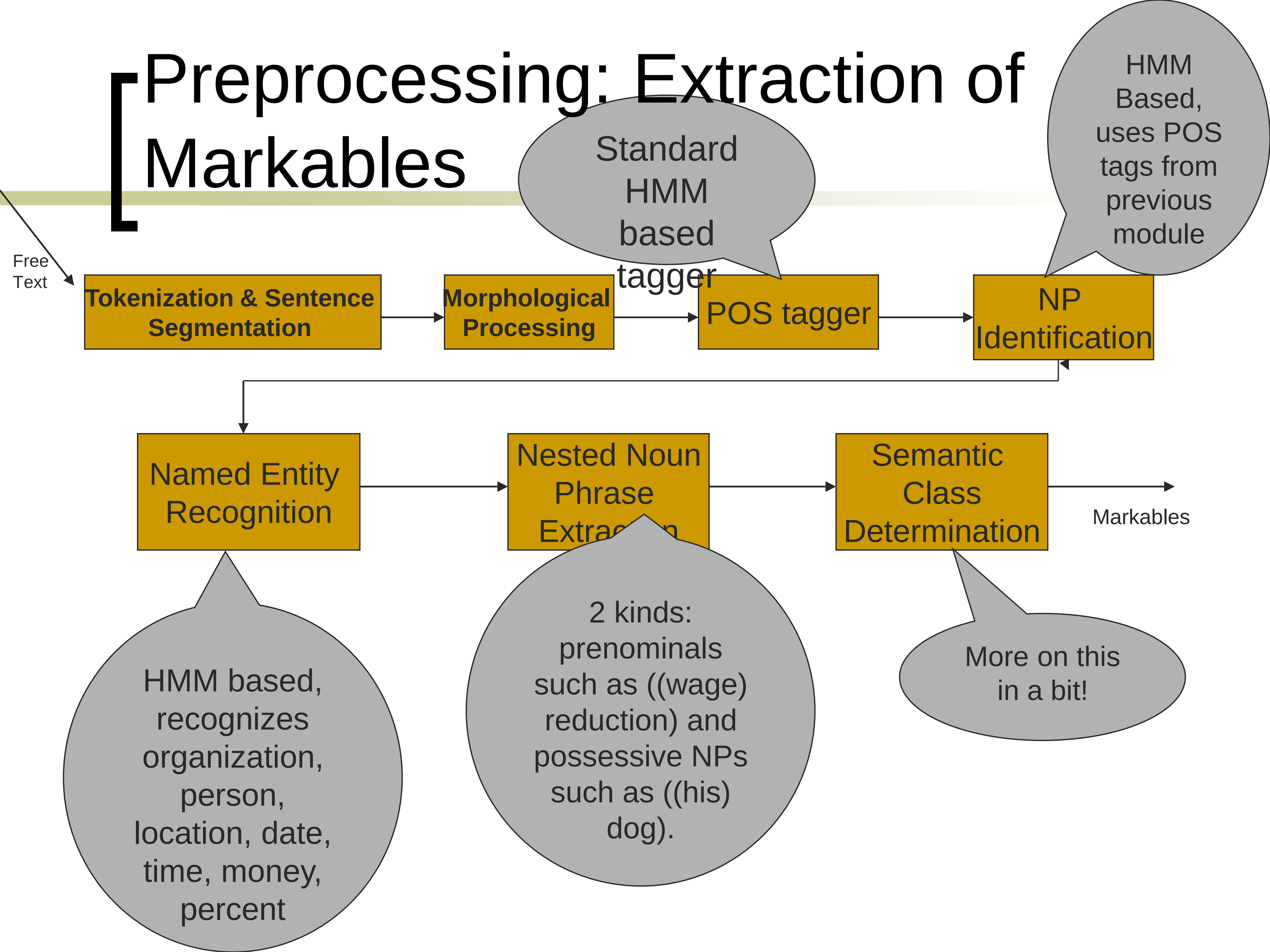


Soon et al. (2001)

Wee Meng Soon, Hwee Tou Ng, Daniel Chung Yong Lim, *A Machine Learning Approach to Coreference Resolution of Noun Phrases*, Computational Linguistics 27(4):521–544

- A reference methodology for supervised learning of coreference chains in text
- Used by many as baseline system

Preprocessing: Extraction of Markables



Soon et al. (2001): preprocessing

- POS tagger: HMM-based
  - 96% accuracy
- Noun phrase identification module
  - HMM-based
  - Can retrieve correctly around 85% of mentions
- NER: reimplementation of Bikel Schwartz and Weischedel (1999)
  - HMM based
  - 88.9% accuracy

Soon et al. (2001): preprocessing

- Nested noun phrase extraction
  - Nested noun phrases from possessive noun phrases
    - [[its] residents]
    - [[Eastern]'s parent]
  - Nested noun phrases that are modifier nouns (or prenominals)
    - [[Iraqi] attack]

Soon et al. (2001): preprocessing

Israel will ask the United States to delay a military strike against Iraq until **the Jewish state** is fully prepared for a possible Iraqi attack with non-conventional weapons, the defense minister said in remarks published Friday. [. . .] **Israel** is equipping **its residents** with gas masks and preparing kits with antidotes. [. . .] **Israel**'s armed forces chief, Lt. Gen. Amnon Lipkin-Shahak, appeared on national television Friday night in attempt to reassure **a jittery public**. "When events are concrete and actually unfolding, we will include **the entire population** in the measures we decide are appropriate to take," Shahak said in an interview with Channel Two Television. (NWIRE/APW19980213.1305)

Soon et al. (2001): preprocessing

Israel will ask the United States to delay a military strike against Iraq until **the Jewish state** is fully prepared for a possible Iraqi attack with non-conventional weapons, the defense minister said in remarks published Friday. [. . .] **Israel** is equipping **its residents** with gas masks and preparing kits with antidotes. [. . .] **Israel**'s armed forces chief, Lt. Gen. Amnon Lipkin-Shahak, appeared on national television Friday night in attempt to reassure **a jittery public**. "When events are concrete and actually unfolding, we will include **the entire population** in the measures we decide are appropriate to take," Shahak said in an interview with Channel Two Television. (NWIRE/APW19980213.1305)

Soon et al: training instances

<ANAPHOR (j), ANTECEDENT (i)>

Soon et al. (2001): 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. (2001): generating training instances

function GET-TRAINING-INSTANCES(T) **returns** *instances*

1: $instances \leftarrow \{\emptyset\}$
2: **for all** training document $D_i \in T = \{D_1, D_2, \ldots D_n\}$ **do**
3:     **for all** coref set S_j in document D_i, $S_j = \{S_1, S_2, \ldots S_m\}$ **do**
4:         $instances \leftarrow instances \cup$ GET-POSITIVE-INSTANCES(S_j)
5:         $instances \leftarrow instances \cup$ GET-NEGATIVE-INSTANCES(S_j, D_i)
6: **return** *instances*

function GET-POSITIVE-INSTANCES(S) **returns** *positive instances*

1: $positive\ instances \leftarrow \{\emptyset\}$
2: **for all** $RE_{S,i} \in S = \{RE_{S,1}, RE_{S,2}, \ldots RE_{S,n}\}$, $i < n$ **do**
3:     $positive\ instances \leftarrow positive\ instances \cup \langle RE_{S,i}, RE_{S,i+1}\rangle$
4: **return** *positive instances*

Soon et al. (2001): generating training instances

function GET-NEGATIVE-INSTANCES(S, D) **returns** *negative instances*

$negative\ instances \leftarrow \{\emptyset\}$

for all $RE_{S,i} \in S = \{RE_{S,1}, RE_{S,2}, \ldots RE_{S,n}\}, i < n$ **do**

  $inter\ REs \leftarrow$ GET-INTERVENING-REs$(\langle RE_{S,i}, RE_{S,i+1}\rangle, S, D)$

  for all $RE_j \in inter\ REs$ **do**

    $negative\ instances \leftarrow negative\ instances \cup \langle RE_j, RE_{S,i+1}\rangle$

return *negative instances*

function GET-INTERVENING-REs$(\langle RE_{S,1}, RE_{S,2}\rangle, S, D)$ **returns** *inter REs*

$inter\ REs \leftarrow \{\emptyset\}$

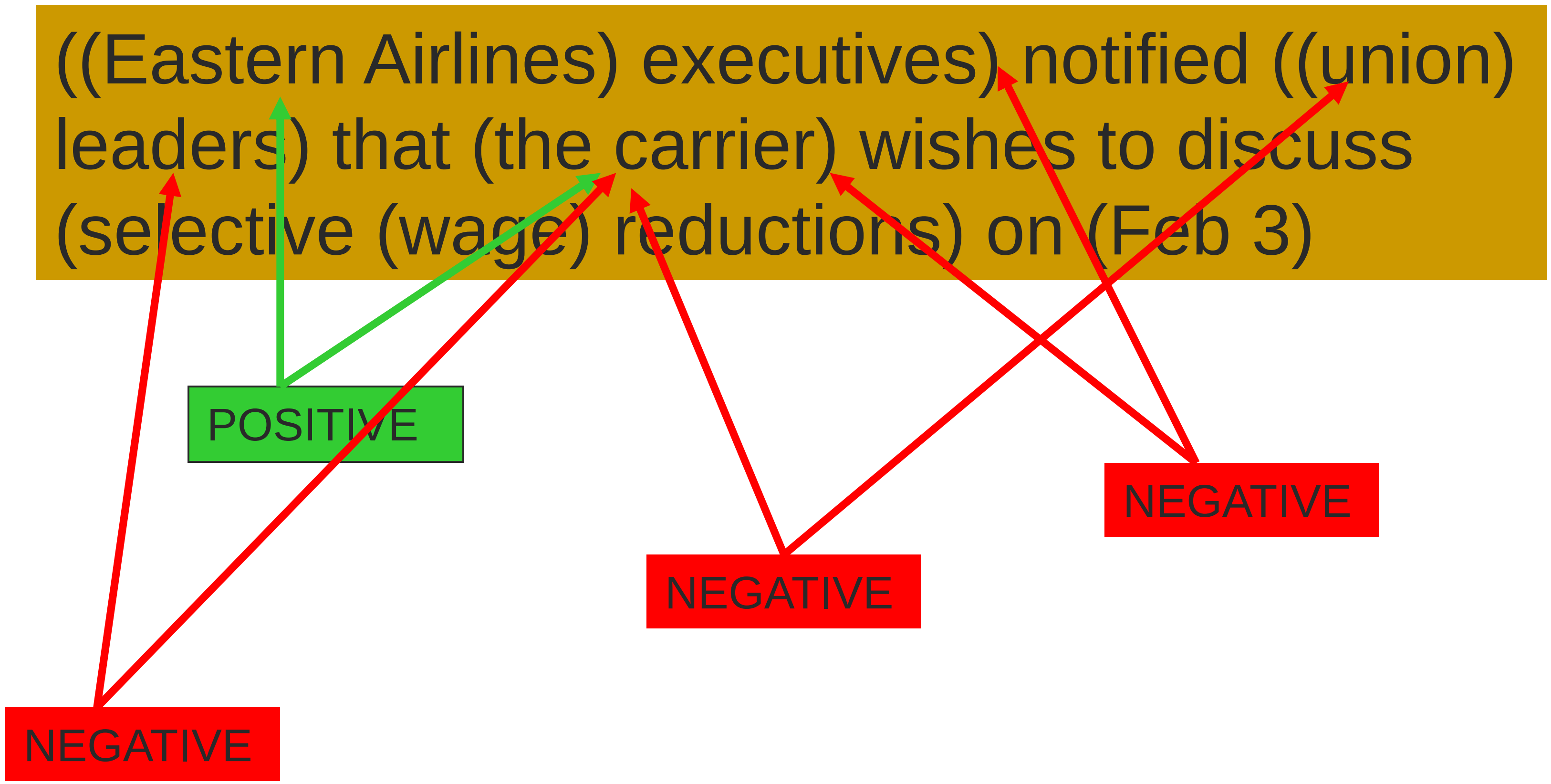
for all RE_i in document D, $RE_i \notin S$ **do**

  if $\mathrm{onset}_{RE_i} > \mathrm{offset}_{RE_1} \wedge \mathrm{offset}_{RE_i} < \mathrm{onset}_{RE_2}$ **then**

    $inter\ REs \leftarrow inter\ REs \cup RE_i$

return *inter REs*

Generating training instances



Soon et al. (2001): preprocessing

Israel will ask the United States to delay a military strike against Iraq until **the Jewish state** is fully prepared for a possible Iraqi attack with non-conventional weapons, the defense minister said in remarks published Friday. [. . .] **Israel** is equipping **its residents** with gas masks and preparing kits with antidotes. [. . .] **Israel**'s armed forces chief, Lt. Gen. Amnon Lipkin-Shahak, appeared on national television Friday night in attempt to reassure **a jittery public**. "When events are concrete and actually unfolding, we will include **the entire population** in the measures we decide are appropriate to take," Shahak said in an interview with Channel Two Television. (NWIRE/APW19980213.1305)

Generating training instances

antecedent	anaphor	label
Israel	**the Jewish state**	$\oplus$
United States	**the Jewish state**	$\ominus$
a military strike	**the Jewish state**	$\ominus$
Iraq	**the Jewish state**	$\ominus$
Jewish	**the Jewish state**	$\ominus$
the Jewish state	**Israel**	$\oplus$
a possible Iraqi attack	**Israel**	$\ominus$
Iraqi	**Israel**	$\ominus$
non-conventional weapons	**Israel**	$\ominus$
the defense minister	**Israel**	$\ominus$
defense	**Israel**	$\ominus$
remarks	**Israel**	$\ominus$
Friday	**Israel**	$\ominus$

Soon et al. (2001): Features

- NP type
- Distance
- Agreement
- Semantic class

Soon et al. (2001): 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)
```

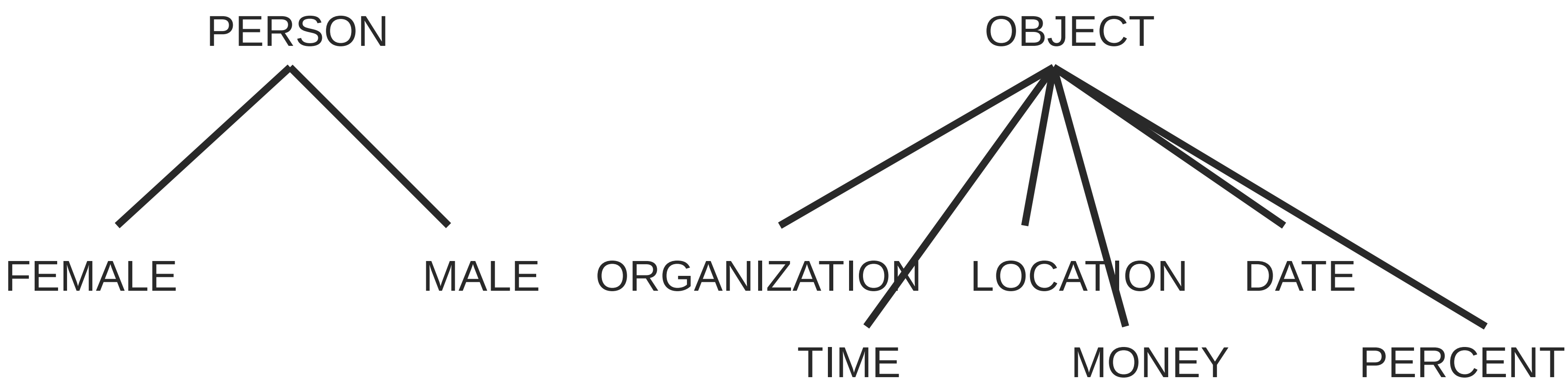
AGREEMENT FEATURES

```
number agreement
gender agreement
```

SYNTACTIC PROPERTIES OF ANAPHOR

```
occurs in appositive contruction
```

Soon et al. (2001): semantic class agreement



SEMCLASS = true iff semclass(i) <= semclass(j) or viceversa (parent of the other or same)

Chairman with semantic class Person and Mr. Lim with semantic class "Male"→ True

IBM with semantic class Organization and Mr. Lim with semantic class Male->False

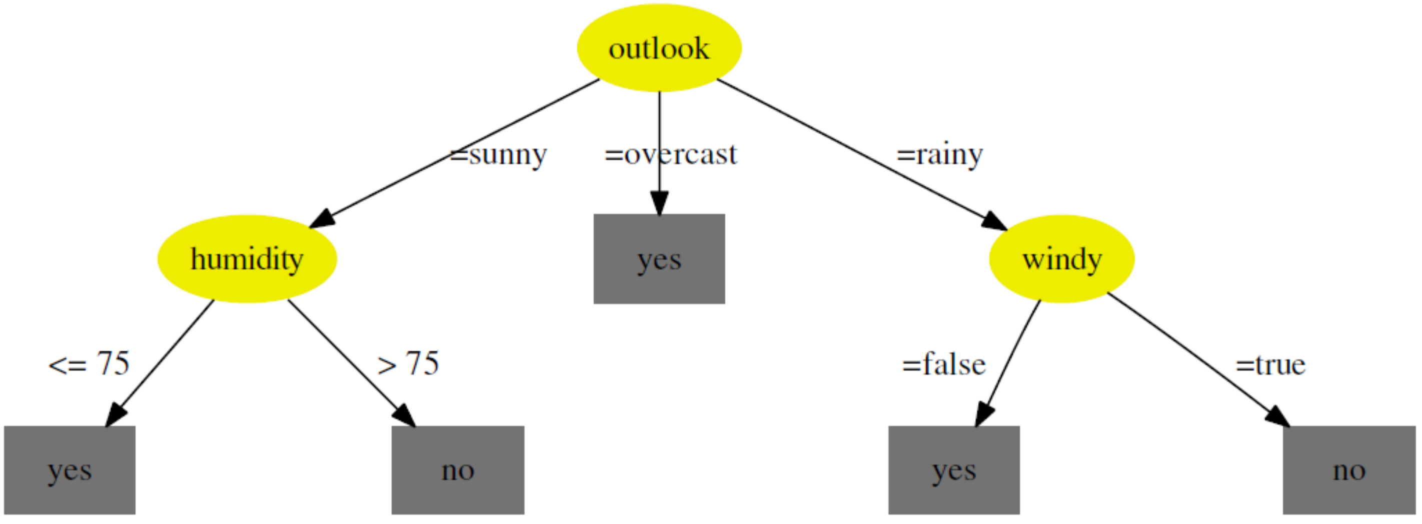
Soon et al. (2001): Features

- Example <Israel, the Jewish state>

Feature	Value	Comment
STRING_MATCH	FALSE	REs do not match lexicographically.
ALIAS	FALSE	RE_j is not an alias of RE_i since RE_i is not a named entity of type PER or ORG.
I_PRONOUN	FALSE	RE_i is not a pronoun.
J_PRONOUN	FALSE	RE_j is not a pronoun.
J_DEF	TRUE	RE_j is a definite NP.
J_DEM	FALSE	RE_j is not a demonstrative NP.
NUMBER	TRUE	RE_i and RE_j are both singular.
GENDER	TRUE	RE_i and RE_j are both neutral.
PROPER_NAME	FALSE	RE_i is a proper name, RE_j is not.
APPOSITIVE	FALSE	RE_i is not in apposition to RE_j.
WN_CLASS	TRUE	RE_i and RE_j are both OBJECT(s).
DISTANCE	0	RE_i and RE_j are in the same sentence.

Decision Trees

- Tree-based classifiers for instances represented as feature-vectors
  - Nodes test features, there is one branch for each value of the feature
  - Leaves specify the category
- ‘Play Tennis’ example



Decision Trees

- Can represent arbitrary conjunction and disjunction
- Can represent any classification function over discrete feature vectors
- Can be rewritten as a set of rules, i.e. disjunctive normal form (DNF)

$$
\begin{aligned}
P_1(x) &:= outlook(x) = overcast \\
P_2(x) &:= outlook(x) = sunny \wedge humidity(x) \leq 75 \\
P_3(x) &:= outlook(x) = rainy \wedge \neg windy(x)
\end{aligned}
$$

Decision tree learning

function LEARN-TREE($examples, attributes, class$) **returns** decision tree

if $examples = \emptyset$ **then**
    return $class$
else
    $majority \leftarrow$ MAJORITY-CLASS($examples$)
    $majorityExamples \leftarrow \{e | e \in examples \wedge \text{CLASS-OF}(e) = m\}$
    if $|examples| = |majorityExamples| \vee attributes = \emptyset$ **then**
        return $majority$
    else
        $bestAttribute \leftarrow$ CHOOSE-ATTRIBUTE($examples, attributes$)
        $otherAttributes \leftarrow attributes - bestAttribute$
        $tree \leftarrow$ tree with root node $bestAttribute$
        for all value v of $bestAttribute$ **do**
            $examples_v \leftarrow \{e | e \in examples \wedge e_{bestAttribute} = v\}$
            $subtree \leftarrow$ LEARN-TREE($examples_v, otherAttributes, majority$)
            add an edge to $tree$ from $bestAttribute$ to $subtree$ with label v
        return $tree$

Decision Tree (MUC-6)

```
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 = -: -
```

Soon et al. (2001): decoding

- Right to left, consider each antecedent **until classifier returns true**
  - → First-positive resolution strategy

Soon et al. (2001): decoding

function GET-COREFERENCE-SETS(D) **returns** *disjoint set*

1: *disjoint set* := a disjoint-set data structure, maintains a collection $\mathcal{S} = \{S_1, S_2 \ldots S_k\}$ of disjoint dynamic sets

2: **for all** $RE_i \in \{RE_1, RE_2, \ldots RE_n\}$ **do**

3:     *disjoint set*.MAKE-SET(RE_i)

4: *coreferent instances* $\leftarrow$ GET-COREFERENCE-PAIRS(D)

5: **for all** $\langle RE_i, RE_j \rangle \in$ *coreferent instances* **do**

6:     *disjoint set*.UNION(FIND(RE_i), FIND(RE_j))

7: **return** *disjoint set*

function GET-COREFERENCE-PAIRS(D) **returns** *coreferent instances*

1: *coreferent instances* $\leftarrow \{\emptyset\}$

2: **for all** $RE_j \in \{RE_1, RE_2, \ldots RE_n\}$, $j > 1$ **do**

3:     **for all** $RE_i \in \{RE_{j-1}, RE_{j-2}, \ldots RE_1\}$ **do**

4:         *test instance* $\leftarrow \langle RE_i, RE_j \rangle$

5:         **if** $p(\oplus|\mathit{test\ instance}) > p(\ominus|\mathit{test\ instance})$ **then**

6:             *coreferent instances* $\leftarrow$ *coreferent instances* $\cup$ *test instance*, **goto** (2)

7: **return** *coreferent instances*

Soon et al. (2001): Decoding

Israel will ask the United States to delay a military strike against Iraq until **the Jewish state** is fully prepared for a possible Iraqi attack with non-conventional weapons, the defense minister said in remarks published Friday. [. . .] **Israel** is equipping **its residents** with gas masks and preparing kits with antidotes. [. . .] **Israel**'s armed forces chief, Lt. Gen. Amnon Lipkin-Shahak, appeared on national television Friday night in attempt to reassure **a jittery public**. "When events are concrete and actually unfolding, we will include **the entire population** in the measures we decide are appropriate to take," Shahak said in an interview with Channel Two Television. (NWIRE/APW19980213.1305)

Decoding example

antecedent	anaphor	corefer?
Jewish	the Jewish state	no
Iraq	the Jewish state	no
a military strike	the Jewish state	no
United States	the Jewish state	no
Israel	**the Jewish state**	yes
Friday	Israel	no
remarks	Israel	no
the defense minister	Israel	no
defense	Israel	no
non-conventional weapons	Israel	no
a possible Iraqi attack	Israel	no
Iraqi	Israel	no
the Jewish state	**Israel**	yes

Soon et al. (2001): evaluation

- MUC-6:
  - P=67.3, R=58.6, F=62.6
- MUC-7:
  - P=65.5, R=56.1, F=60.4
- Results about 3rd or 4th amongst the best MUC-6 and MUC-7 systems

More analysis: contribution of the features (MUC-6)

Baseline systems using just one feature				
DIST	0.0	0.0	0.0	Only "distance" feature is used
SEMCLASS	0.0	0.0	0.0	Only "semantic class agreement"
NUMBER	0.0	0.0	0.0	Only "number agreement"
GENDER	0.0	0.0	0.0	Only "gender agreement"
PROPER_NAME	0.0	0.0	0.0	Only "both proper names"
ALIAS	24.5	88.7	38.4	Only "alias"
J_PRONOUN	0.0	0.0	0.0	Only "j-pronoun"
DEF_NP	0.0	0.0	0.0	Only "definite noun phrase"
DEM_NP	0.0	0.0	0.0	Only "demonstrative noun phrase"
STR_MATCH	45.7	65.6	53.9	Only "string match"
APPOSITIVE	3.9	57.7	7.3	Only "appositive"
I_PRONOUN	0.0	0.0	0.0	Only "i-pronoun"
Other baseline systems				
ALIAS_STR	51.5	66.4	58.0	Only the "alias" and "string match" features are used
ALIAS_STR_APPOS	55.2	66.4	60.3	Only the "alias," "string match," and "appositive" features are used
ONE_CHAIN	89.9	31.8	47.0	All markables form one chain
ONE_WRD	55.4	36.6	44.1	Markables corefer if there is at least one common word
HD_WRD	56.4	50.4	53.2	Markables corefer if their head words are the same

More analysis: contribution of the features (MUC-7)

Baseline systems using just one feature				
DIST	0.0	0.0	0.0	Only "distance" feature is used
SEMCLASS	0.0	0.0	0.0	Only "semantic class agreement"
NUMBER	0.0	0.0	0.0	Only "number agreement"
GENDER	0.0	0.0	0.0	Only "gender agreement"
PROPER_NAME	0.0	0.0	0.0	Only "both proper names"
ALIAS	25.6	81.1	38.9	Only "alias"
J_PRONOUN	0.0	0.0	0.0	Only "j-pronoun"
DEF_NP	0.0	0.0	0.0	Only "definite noun phrase"
DEM_NP	0.0	0.0	0.0	Only "demonstrative noun phrase"
STR_MATCH	43.8	71.4	54.3	Only "string match"
APPOSITIVE	2.4	60.0	4.6	Only "appositive"
I_PRONOUN	0.0	0.0	0.0	Only "i-pronoun"
Other baseline systems				
ALIAS_STR	49.4	70.4	58.1	Only the "alias" and "string match" features are used
ALIAS_STR_APPOS	51.6	69.9	59.4	Only the "alias," "string match," and "appositive" features are used
ONE_CHAIN	87.5	30.5	45.2	All markables form one chain
ONE_WRD	55.9	38.7	45.7	Markables corefer if there is at least one common word
HD_WRD	55.2	55.6	55.4	Markables corefer if their head words are the same

Contribution of the features

- The decision tree shows that only 8 features are being used
- When used with 3 features (**alias, apposition, string match**) the scores (F-measure) were only 1-2.3% worse than when used with all of them
  - → <u>only 3 features really contribute</u>

Ng & Cardie

- 2002:
  - Changes to the model:
    - Positive: first NON PRONOMINAL
    - Decoding: choose MOST HIGH PROBABILITY
  - Many more features:
    - Many more string features
    - Linguistic features (binding, etc)
- Subsequently:
  - Discourse new detection (see below)

Better positive examples generation

- Soon et al.
  - Use *most recent* antecedent
- Ng & Cardie
  - Use the **most confident antecedent**
  - if the anaphor is non-pronominal, use the most recent non-pronominal antecedent

Ng & Cardie (2002): decoding

- Right to left, consider each antecedent and select the **highest likely antecedent**
  - → Best-first resolution strategy

More features #1: Improved string match

- Soon et al.
  - Same string match feature for all expressions
- Ng & Cardie
  - 3 different features for
    - Pronominals
    - Common nouns
    - Proper names

Ng & Cardie: learning framework improvements

	MUC-6	MUC-7
Soon et al.	62.6	60.4
Better preprocessing	66.3	61.2
Best-first clustering	66.3	62.3
Most confident antecedent	65.8	61.1
String features	66.7	62.0
Combined	67.5	63.0

More features #2

Added 41 more features:

- lexical
- grammatical
- semantic

Ng & Cardie (2002): additional features

Lexical features

Lexical		PRO_STR*	C if both NPs are pronominal and are the same string; else I.
		PN_STR*	C if both NPs are proper names and are the same string; else I.
		WORDS_STR	C if both NPs are non-pronominal and are the same string; else I.
		SOON_STR_NONPRO*	C if both NPs are non-pronominal and the string of NP_i matches that of NP_j; else I.
		WORD_OVERLAP	C if the intersection between the content words in NP_i and NP_j is not empty; else I.
		MODIFIER	C if the prenominal modifiers of one NP are a subset of the prenominal modifiers of the other; else I.
		PN_SUBSTR	C if both NPs are proper names and one NP is a proper substring (w.r.t. content words only) of the other; else I.
		WORDS_SUBSTR	C if both NPs are non-pronominal and one NP is a proper substring (w.r.t. content words only) of the other; else I.

Ng & Cardie (2002): additional features

Grammatical features I

Grammatical	NP type	BOTH_DEFINITES	C if both NPs start with "the;" I if neither start with "the;" else NA.
		BOTH_EMBEDDED	C if both NPs are prenominal modifiers ; I if neither are prenominal modifiers; else NA.
		BOTH_IN_QUOTES	C if both NPs are part of a quoted string; I if neither are part of a quoted string; else NA.
		BOTH_PRONOUNS*	C if both NPs are pronouns; I if neither are pronouns, else NA.
	role	BOTH_SUBJECTS	C if both NPs are grammatical subjects; I if neither are subjects; else NA.
		SUBJECT_1*	Y if NP_i is a subject; else N.
		SUBJECT_2	Y if NP_j is a subject; else N.
	lin-gui-stic con-stra-ints	AGREEMENT*	C if the NPs agree in both gender and number; I if they disagree in both gender and number; else NA.
		ANIMACY*	C if the NPs match in animacy; else I.
		MAXIMALNP*	I if both NPs have the same maximal NP projection; else C.
		PREDNOM*	C if the NPs form a predicate nominal construction; else I.
		SPAN*	I if one NP spans the other; else C.
		BINDING*	I if the NPs violate conditions B or C of the Binding Theory; else C.
		CONTRAINDICES*	I if the NPs cannot be co-indexed based on simple heuristics; else C. For instance, two non-pronominal NPs separated by a preposition cannot be co-indexed.
		SYNTAX*	I if the NPs have incompatible values for the BINDING, CONTRAINDICES, SPAN or MAXIMALNP constraints; else C.

Ng & Cardie (2002): additional features

Grammatical features II

ling. prefs	INDEFINITE*	I if NP_j is an indefinite and not appositive; else C.
	PRONOUN	I if NP_i is a pronoun and NP_j is not; else C.
heur-istics	CONSTRAINTS*	C if the NPs agree in GENDER and NUMBER and do not have incompatible values for CONTRAINDICES, SPAN, ANIMACY, PRONOUN, and CONTAINS_PN; I if the NPs have incompatible values for any of the above features; else NA.
	CONTAINS_PN	I if both NPs are not proper names but contain proper names that mismatch on every word; else C.
	DEFINITE_1	Y if NP_i starts with "the;" else N.
	EMBEDDED_1*	Y if NP_i is an embedded noun; else N.
	EMBEDDED_2	Y if NP_j is an embedded noun; else N.
	IN_QUOTE_1	Y if NP_i is part of a quoted string; else N.
	IN_QUOTE_2	Y if NP_j is part of a quoted string; else N.
	PROPER_NOUN	I if both NPs are proper names, but mismatch on every word; else C.
	TITLE*	I if one or both of the NPs is a title; else C.

Ng & Cardie (2002): additional features

Other features

Semantic		CLOSEST_COMP	C if NP_i is the closest NP preceding NP_j that has the same semantic class as NP_j and the two NPs do not violate any of the linguistic constraints; else I.
		SUBCLASS	C if the NPs have different head nouns but have an ancestor-descendent relationship in WordNet; else I.
		WNDIST	Distance between NP_i and NP_j in WordNet (using the first sense only) when they have an ancestor-descendent relationship but have different heads; else infinity.
		WNSENSE	Sense number in WordNet for which there exists an ancestor-descendent relationship between the two NPs when they have different heads; else infinity.
Pos		PARANUM	Distance between the NPs in terms of the number of paragraphs.
Other		PRO_RESOLVE*	C if NP_j is a pronoun and NP_i is its antecedent according to a naive pronoun resolution algorithm; else I.
		RULE_RESOLVE	C if the NPs are coreferent according to a rule-based coreference resolution algorithm; else I.

Selecting features

- Full feature set gives very low precision on nominal anaphors
  - Too many features for too little data
  - Manually eliminate those features which gives low precision (on training data)

Ng & Cardie: feature set expansion improvements

	MUC-6	MUC-7
Intermediate system	67.5	63.0
All features	63.8	61.6
Hand-selected features	69.1	63.4

Based on:

- Ponzetto and Poesio (2009). State-of-the-art NLP Approaches to Coreference Resolution: Theory and Practical Recipes-Tutorial
- Few slides from
  - Sobha Lalitha Devi (Summer School, IIIT Hyderabad)