EMPIRICAL INVESTIGATIONS OF ANAPHORA AND SALIENCE

Massimo Poesio
Università di Trento and
University of Essex

Vilem Mathesius Lectures
Praha, 2007

CONTEXT DEPENDENCE

1.1 M: all right system
1.2 : we've got a more complicated problem
1.4 : first thing _I'd_ like you to do
1.5 : is send engine E2 off with a boxcar to Corning to pick up oranges
1.6 : uh as soon as possible
2.1 S: okay
3.1 M: and while it's there it should pick up the tanker
4.1 S: okay
4.2 : and that can get
4.3 : we can get that done by three
5.1 M: good
5.3 : can we please send engine E1 over to Dansville to pick up a boxcar
5.4 : and then send it right back to Avon
6.1 S: okay
6.2 : it'll get back to Avon at 6

CONTEXT DEPENDENCE

- The interpretation of most expressions depends on the context in which they are used
  - Studying the semantics & pragmatics of context dependence a crucial aspect of linguistics
- Developing methods for interpreting context dependent expressions useful in many applications
  - Information extraction: recognize which expressions are mentions of the same object
  - Multimodal interfaces: recognize which objects in the visual scene are being referred to
- We focus here on dependence of nominal expressions on context introduced LINGUISTICALLY, for which I'll use the term ANAPHORA

Plan of these lectures

- Today: Annotating context dependence, and particularly anaphora
- Tomorrow: Using anaphorically annotated corpora to investigate local & global salience ('topic tracking')
- Friday: Using anaphorically annotated corpora to investigate anaphora resolution

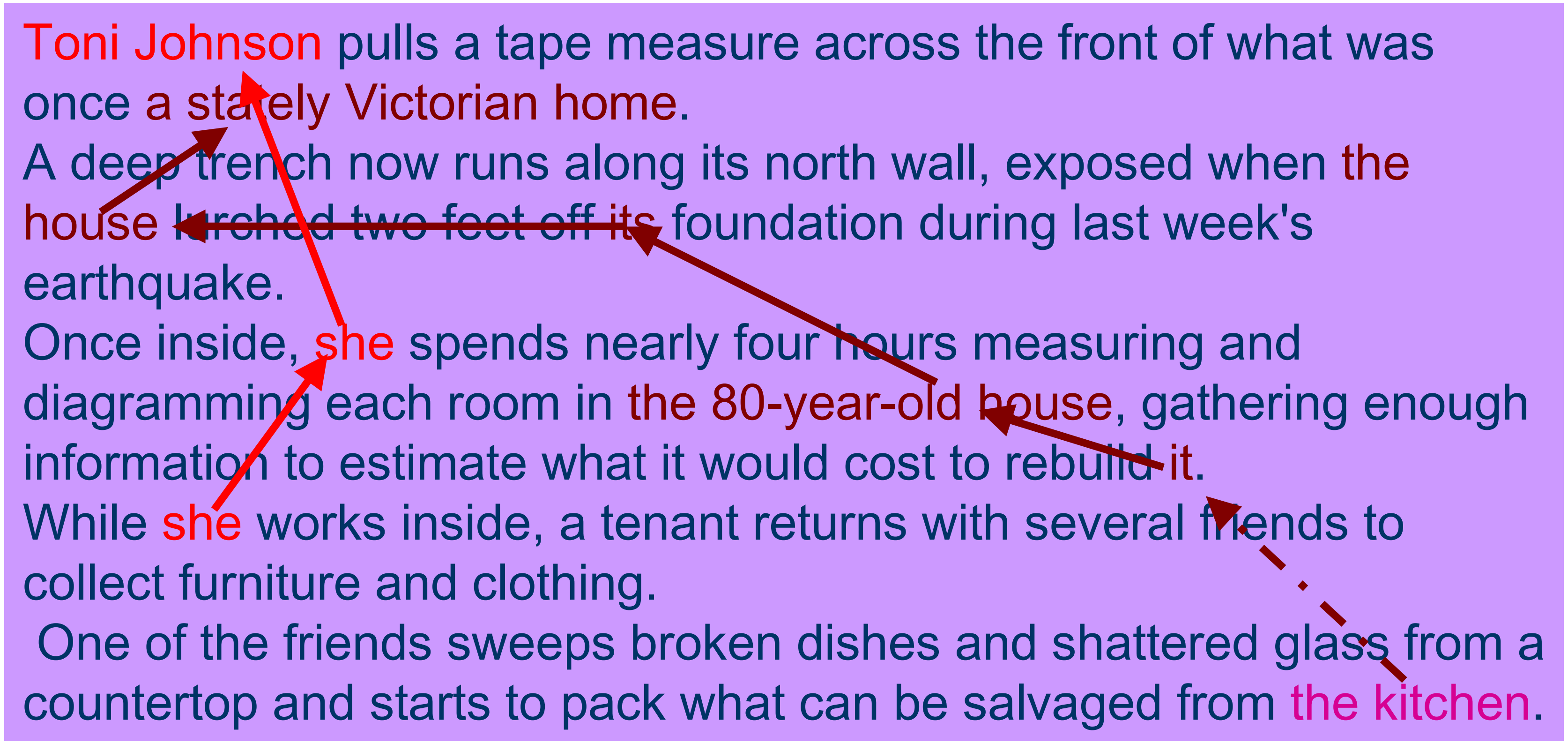
Objectives of today’s lecture

- Methods we and others have developed to annotate various types of linguistic context dependence for a variety of purposes
- Some lessons we learned

MOTIVATIONS FOR ANNOTATING ANAPHORIC INFORMATION

- Linguistic research
  - E.g., work on information structure in Prague (Haijcova, Sgall, Kruijff-Korbayova) and elsewhere (Prince, Gundel et al, Fraurud)
  - Also in Computational Linguistics (e.g., work by Passonneau, Walker)
  - Example: tomorrow, our work on salience
- System building
  - E.g., development of anaphora resolution / NLG systems
  - Example: Friday, our work on bridging and anaphora resolution
- Applications
  - Information extraction (MUC, ACE, GENIA)
  - Other applications: segmentation, summarization

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.

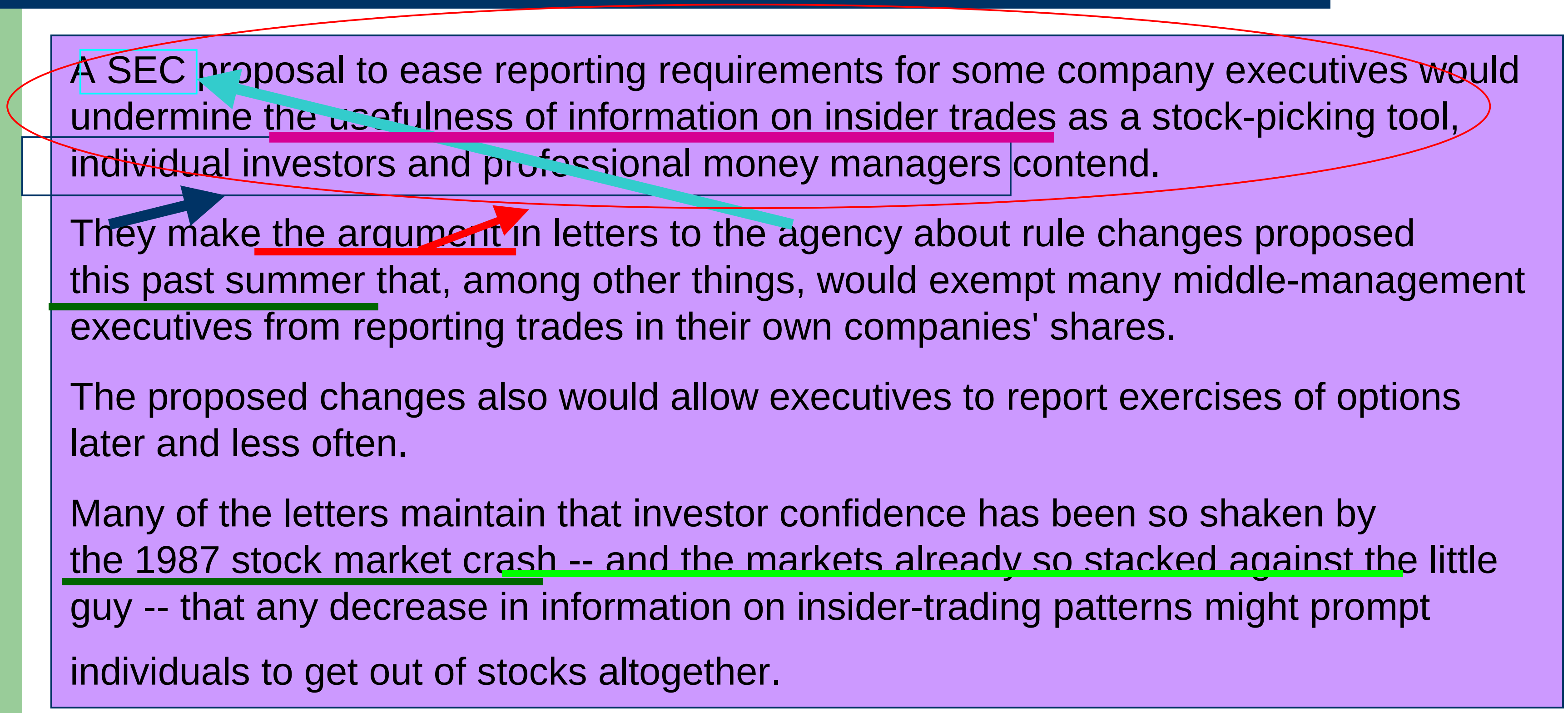
One of the friends sweeps broken dishes and shattered glass from a countertop and starts to pack what can be salvaged from the kitchen.

(WSJ section of Penn Treebank corpus)

The Big Issue

- More than with shallower annotations (POS tags, constituency / dependency) purpose of annotation may affect decisions as to what annotate and how
  - MUC vs. MapTask
  - Coref vs anaphora

More difficult choices



A SEC proposal to ease reporting requirements for some company executives would undermine the usefulness of information on insider trades as a stock-picking tool, individual investors and professional money managers contend.

They make the argument in letters to the agency about rule changes proposed this past summer that, among other things, would exempt many middle-management executives from reporting trades in their own companies' shares.

The proposed changes also would allow executives to report exercises of options later and less often.

Many of the letters maintain that investor confidence has been so shaken by the 1987 stock market crash -- and the markets already so stacked against the little guy -- that any decrease in information on insider-trading patterns might prompt individuals to get out of stocks altogether.

WSJ section of Penn Treebank corpus

Today’s lecture

- Linguistic background on anaphora
- A survey of some of the best-known schemes for annotating linguistic context-dependence
  - Mostly focusing on identity relations
  - GNOME: annotating bridging relations
- Reliability
- Ambiguity
- (If time allows) Annotating discourse deixis

Nominal anaphoric expressions

- REFLEXIVE PRONOUNS:
  - John bought himself an hamburger
- PRONOUNS:
  - Definite pronouns: Ross bought {a radiometer | three kilograms of after-dinner mints} and gave {it | them} to Nadia for her birthday. (Hirst, 1981)
  - Indefinite pronouns: Sally admired Sue's jacket, so she got one for Christmas. (Garnham, 2001)
- DEFINITE DESCRIPTIONS:
  - A man and a woman came into the room. The man sat down.
  - Epiteths: A man ran into my car. The idiot wasn't looking where he was going.
- DEMONSTRATIVES:
  - Tom has been caught shoplifting. That boy will turn out badly.

Interpretive differences between nominal expressions

Put the apple on the napkin and then move it to the side.

Put the apple on the napkin and then move that to the side. (Gundel)

John thought about {becoming a bum}.

It would hurt his mother and it would make his father furious.

It would hurt his mother and that would make his father furious. (Schuster, 1988)

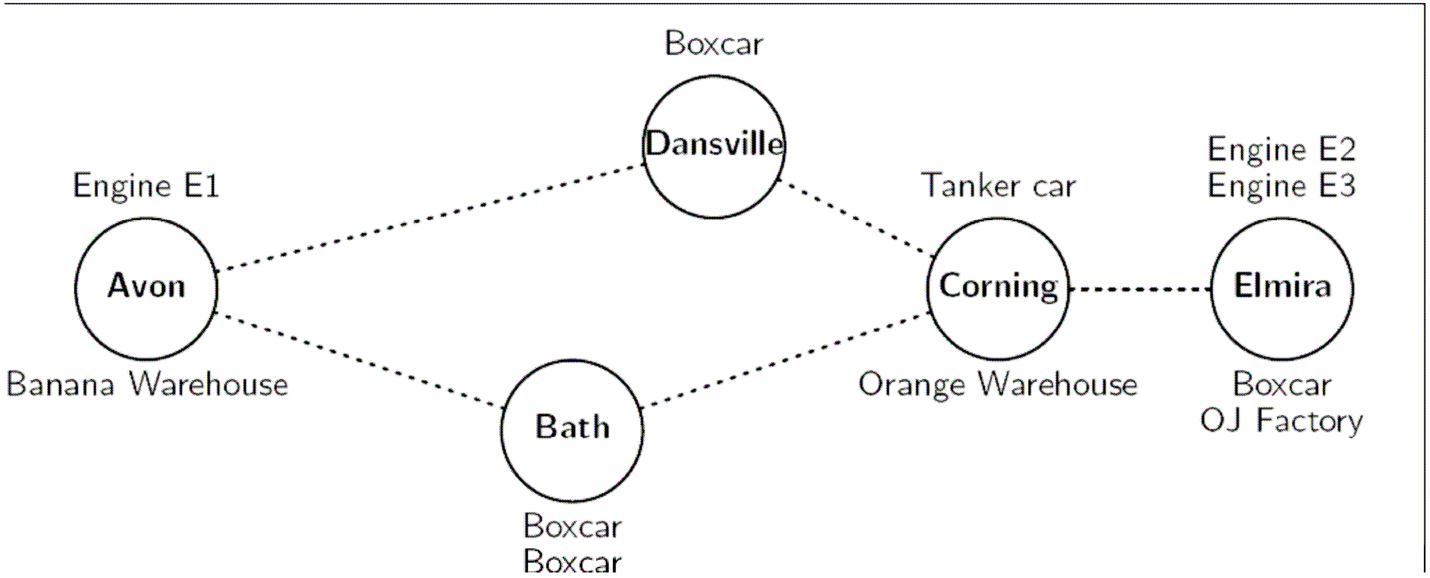
Non-nominal anaphoric expressions

- PRO-VERBS:
  - Daryel thinks like I do.
- GAPPING:
  - Nadia brought the food for the picnic, and Daryel _ the wine.
- TEMPORAL REFERENCES:
  - In the mid-Sixties, free love was rampant across campus. It was then that Sue turned to Scientology. (Hirst, 1981)
- LOCATIVE REFERENCES:
  - The Church of Scientology met in a secret room behind the local Colonel Sanders’ chicken stand. Sue had her first dianetic experience there. (Hirst, 1981)

Not all ‘anaphoric’ expressions always anaphoric

- Expletives
  - It is half past two.
- References to visual situation (‘exophora’)
  - pick that up and put it over there.
- Discourse deixis
- First mention definites

REFERENCES TO VISUAL SITUATION (`EXOPHORA’) IN TRAINS



References to visual situation (‘exophora’ / deixis)

S hello can I help you

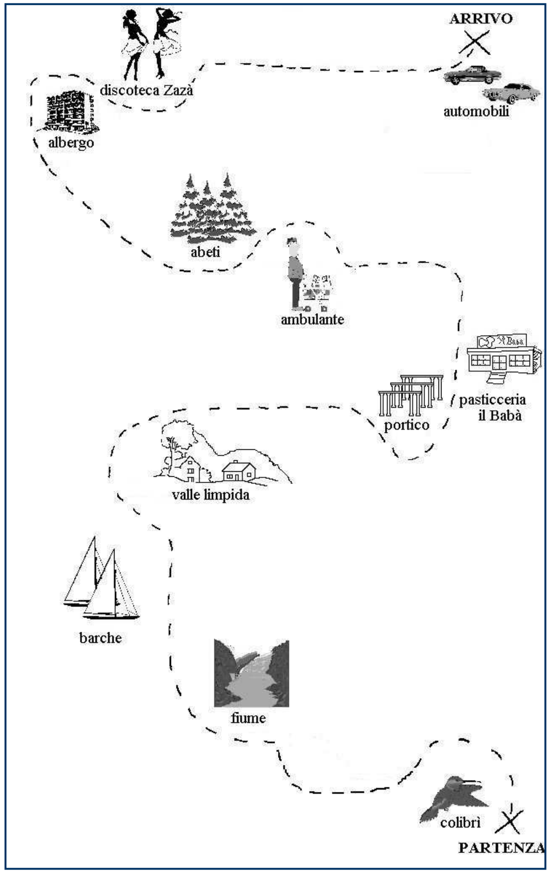
U yeah I want t- I want to determine the maximum number of boxcars of oranges that I can get to Bath by 7 a.m. tomorrow morning so hm so I guess all the boxcars will have to go through oran- through Corning because that's where the orange juice factory is

TRAINS corpus 1993 (Heeman & Allen)
(example reported by J. Gundel)

(Speaker sees addressee looking at a picture) She looks just like her mother, doesn’t she?

(Gundel 1980)

EXOPHORA IN THE MAPTASK



Discourse deixis

"We believe her, the court does not, and **that** resolves the matter,"

[NY Times, 5/24/ 00] (from Gundel)

(Dentist to patient) Did **that** hurt?

(Jackendoff 2002)

First-mention definites

S hello can I help you

U yeah I want t- I want to determine **the maximum number of boxcars of oranges that I can get to Bath by 7 a.m. tomorrow morning** so hm so I guess all the boxcars will have to go through oran- through Corning because that's where **the orange juice factory** is

1993 TRAINS corpus, Heeman & Allen
(example reported by J. Gundel)

Not all ‘anaphoric’ expressions always anaphoric

- Expletives
- References to visual situation (‘exophora’)
- Discourse deixis
- First mention definites
  - Fraurud 1990, Poesio & Vieira 1998: first mention definites more than 50% of all definites (more in newspaper style)

Types of anaphoric relations

- Identity of REFERENCE
  - Ross bought {a radiometer | three kilograms of after-dinner mints} and gave {it | them} to Nadia for her birthday.
- Identity of SENSE
  - Sally admired Sue's jacket, so she got one for Christmas. (Garnham, 2001)
  - (PAYCHECK PRONOUNS): The man who gave his paycheck to his wife is wiser than the man who gave it to his mistress. (Karttunen, 1976?)
- BOUND anaphora
  - No Italian believes that World Cup referees treated his team fairly
- ASSOCIATIVE / indirect anaphoric relations ('bridging')
  - The house …. the kitchen

Associative anaphora

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.

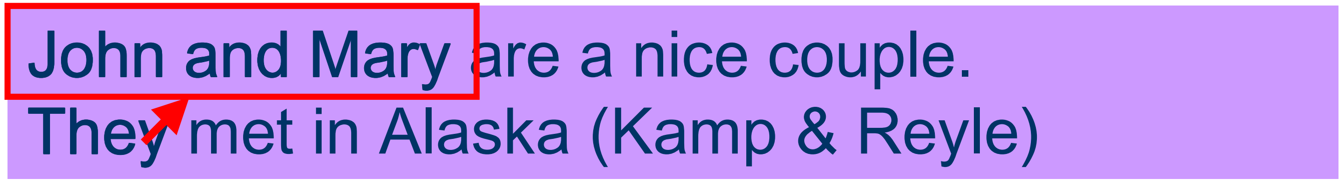
Once inside, she spends nearly four hours measuring and diagramming each room in the 80-year-old house, gathering enough information to estimate what it would cost to rebuild it.

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

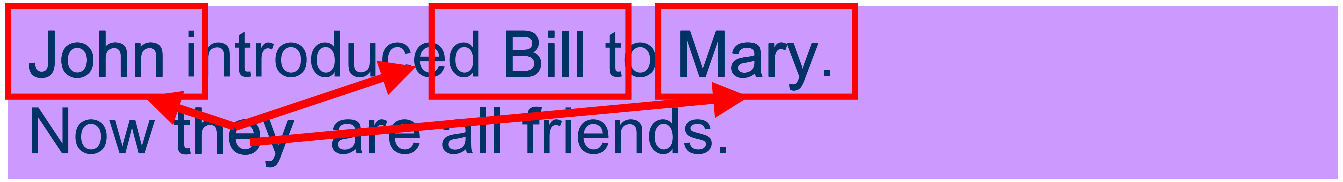
One of the friends sweeps broken dishes and shattered glass from a countertop and starts to pack what can be salvaged from the kitchen.

(WSJ section of Penn Treebank corpus)

Explicit and implicit antecedents



John and Mary are a nice couple.
They met in Alaska (Kamp & Reyle)



John introduced Bill to Mary.
Now they are all friends.

Explicit and implicit antecedents

We believe her, the court does not, and **that** resolves the matter,” [NY Times, 5/24/ 00]

Anyway , going back from the kitchen then is a little hallway leading to a window, and across from the kitchen is **a big walk-through closet**. On the other side of **that** is another little hallway leading to a window…[personal letter, from Gundel et al 1993]

Theoretical foundations

- Although one of the goals of corpus annotation is to uncover linguistic evidence, it cannot be done in the complete absence of any theoretical framework
- Problem with annotating context dependence: even less theoretical agreement than with parsing
- Our own work on context dependence based on ideas developed in ‘dynamic’ theories of the ‘discourse model’ as developed by Heim, Kamp and Reyle, Webber, et al

ANAPHORIC RELATIONS IN A DISCOURSE MODEL

We're gonna take engine E3

and shove IT to Corning

DE1

DE1=E3

take(we,DE1)

ANAPHORIC RELATIONS IN A DISCOURSE MODEL

We're gonna take engine E3

and shove IT to Corning

DE1 DE2 DE3 ….

DE1=E3
take(we,DE1)

DE2=DE1
DE3=Corning
shove(we,DE2,DE3)

IMPLICIT OBJECTS IN A DISCOURSE MODEL: PLURALS

John introduced Bill to Mary.
Now they are all friends.

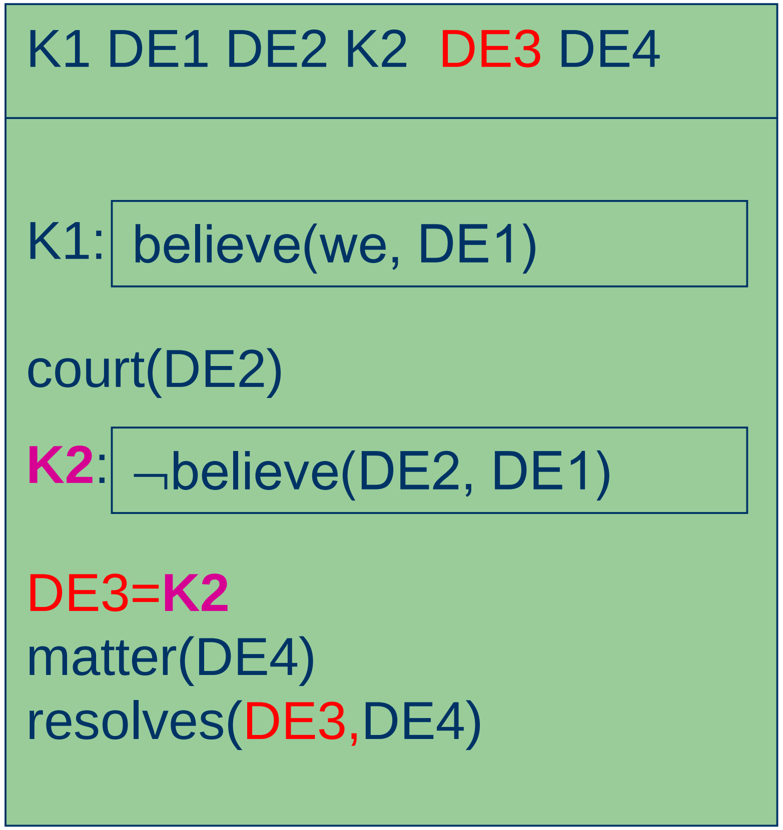
DE1 DE2 DE3 **DE4** DE5

DE1 = John
DE2 = Bill
DE3 = Mary
introduce (DE1, DE2, DE3)

DE4 = DE1+DE2+DE3
DE5=DE4
friends(DE5)

IMPLICIT OBJECTS IN A DISCOURSE MODEL: DISCOURSE DEIXIS

We believe her,
the court does not,
and that resolves the matter



K1 DE1 DE2 K2 DE3 DE4

K1: believe(we, DE1)

court(DE2)

K2: ¬believe(DE2, DE1)

DE3=**K2**
matter(DE4)
resolves(DE3,DE4)

EXOPHORA / DEIXIS

We're gonna take engine E3

and shove IT to Corning

DE2 DE7 DE8 DE9 ….

DE7=E3
take(we,DE7)

DE8=DE7
DE2=Corning
shove(we,DE8,DE2)

EXOPHORA / DEIXIS?

- E.g., MapTask

Some terminology

- CONTEXT-DEPENDENCE: meaning of expression depends on context
  - More specifically: depends on DISCOURSE ENTITY introduced in context
- COREFERENCE: two expressions denote the same object
- ANAPHORA:
  - `textual' definition: a 'linguistic' relation between surface expressions / syntactic expressions (asymmetric)
    - Problem: can't always mark the closest antecedent
  - Discourse-model based definition: the DISCOURSE ENTITIES realized by the expressions are linked by a NON-EXPLICIT relation

Problems with taking `linguistic’ view of ‘anaphora’ as basis for annotation

- Can’t always choose closest antecedent

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

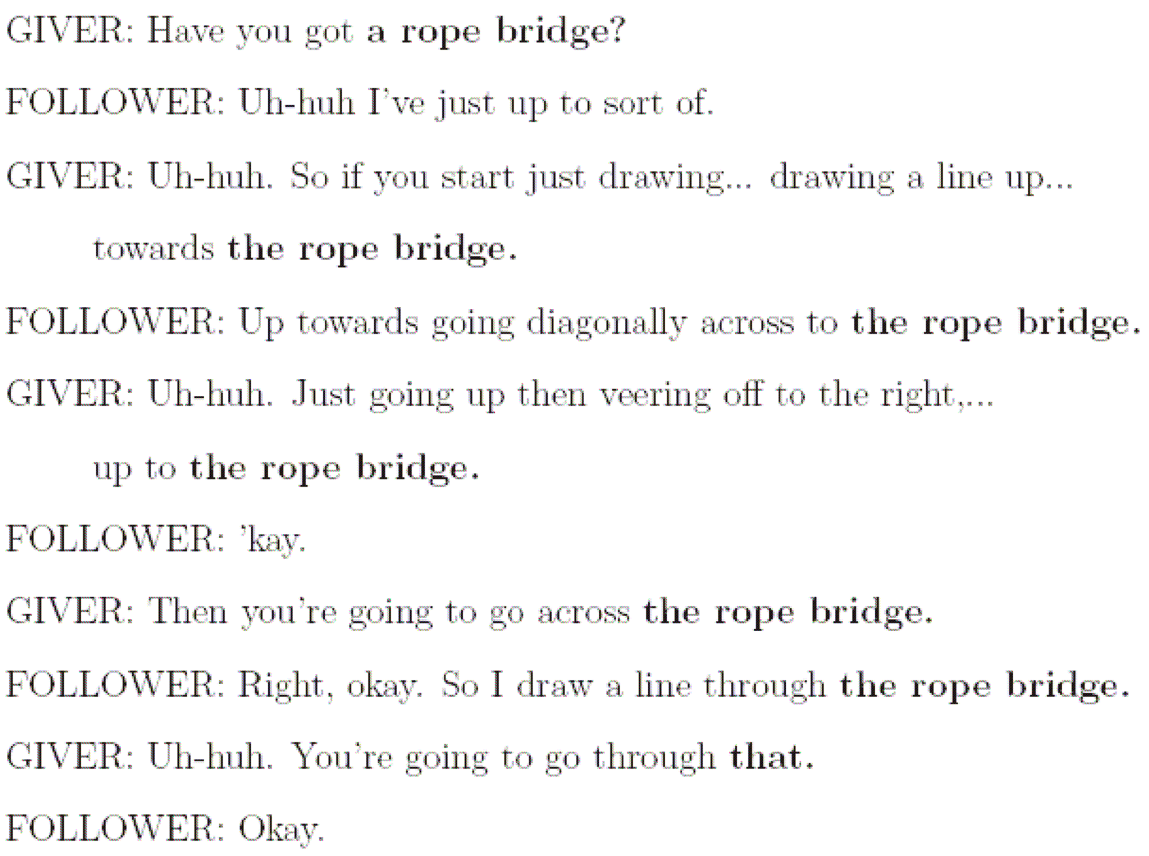
Coding schemes for context-dependence

- MapTask (non linguistic)
- MUC (coreference)
- MATE
- GNOME
- (Some schemes for marking familiarity)
- Prague Dependency Treebank
- ONTONOTES

Differences between coding schemes

- Type of anaphoric expressions and context dependence relations that were annotated
  - Most proposals concentrate on nominal anaphoric expressions (but see work by Hardt)
  - Most proposals avoid bridging relations (but: DRAMA, MATE, GNOME, MULI)
- Coding instructions and their level of formalization
  - E.g., which markables (full nominal expression including postmodifiers / only up to head)
  - Whether markables identified by hand or automatically
- Markup scheme
  - Since MapTask & MUC, most SGML / XML
  - But: some schemes use attributes, other elements

MapTask Reference Coding (Aylett, 2000)



GIVER: Have you got **a rope bridge**?

FOLLOWER: Uh-huh I've just up to sort of.

GIVER: Uh-huh. So if you start just drawing... drawing a line up...
towards **the rope bridge**.

FOLLOWER: Up towards going diagonally across to **the rope bridge**.

GIVER: Uh-huh. Just going up then veering off to the right,...
up to **the rope bridge**.

FOLLOWER: 'kay.

GIVER: Then you're going to go across **the rope bridge**.

FOLLOWER: Right, okay. So I draw a line through **the rope bridge**.

GIVER: Uh-huh. You're going to go through **that**.

FOLLOWER: Okay.

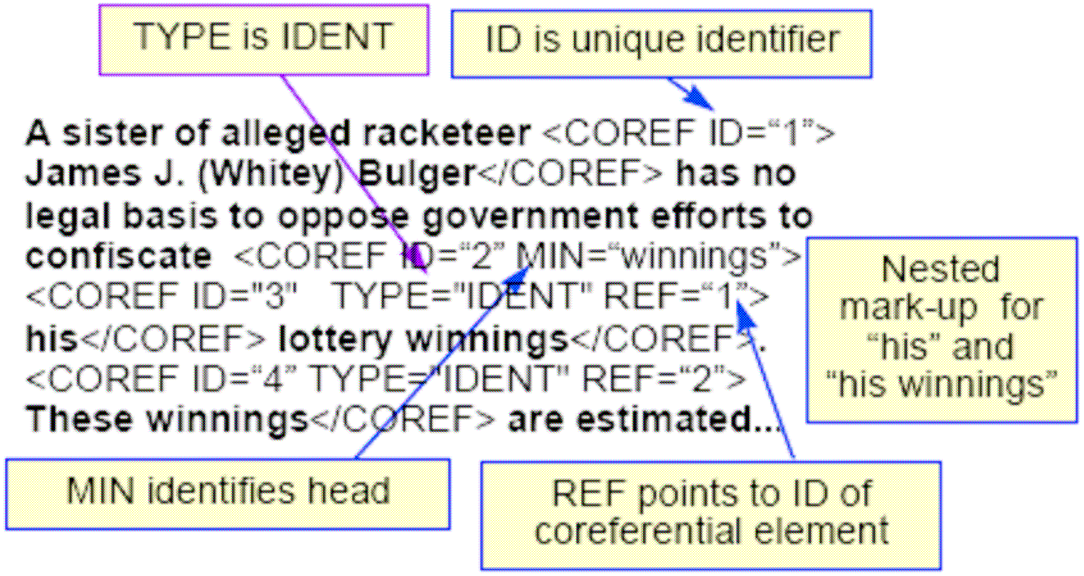
MapTask Reference Coding (Aylett, 2000)

- Type of context dependence annotated: reference to landmarks
  - an example of exophora / deixis
  - Not unlike ‘TIMEX’ markup
- Markup scheme:
  - XML
  - Using attribute to specify landmark
- Coding manual: unknown

MUC coreference scheme (Hirschman & Sundheim, 1997)

- The most popular scheme for linguistic context-dependence in text (used in MUC-6, MUC-7, and ACE)
- Two key design decisions:
  - Goal of the annotation: evaluating subtask of information extraction → attempt to maximise links (also mark predications)
  - Practical focus → concentrate on what can be annotated quickly and reliably → ignore bridging relations
- A very detailed coding scheme
- Markup scheme: SGML, using attributes to indicate coref links

The coding scheme



TYPE is IDENT

ID is unique identifier

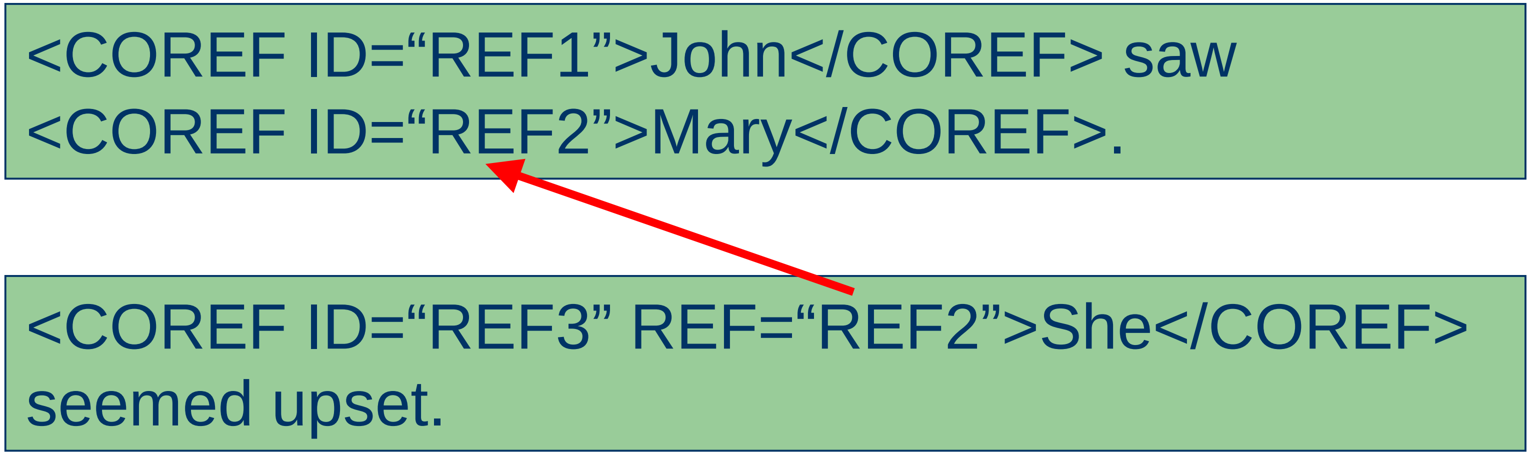
A sister of alleged racketeer <COREF ID="1">
James J. (Whitey) Bulger</COREF> has no
legal basis to oppose government efforts to
confiscate <COREF ID="2" MIN="winnings">
<COREF ID="3" TYPE="IDENT" REF="1">
his</COREF> lottery winnings</COREF>.
<COREF ID="4" TYPE="IDENT" REF="2">
These winnings</COREF> are estimated...

Nested mark-up for "his" and "his winnings"

MIN identifies head

REF points to ID of coreferential element

Coreference in XML: MUC (Hirschman, 1997)



<COREF ID=“REF1”>John</COREF> saw <COREF ID=“REF2”>Mary</COREF>.

<COREF ID=“REF3” REF=“REF2”>She</COREF> seemed upset.

Problems with the MUC scheme

- Linguistic limitation: Notion of ‘coreference’ not well defined (van Deemter and Kibble, 2001)
- Limitations of the markup scheme:
  - Only one type of anaphoric relation
  - No way of marking ambiguous cases

‘Extended coreference’ in MUC

the IRS's position was that
<COREF ID=“REF1”> the stock's value </COREF>
was
<COREF REF=“REF1”> $144.5 million </COREF>
on the alternative valuation date

Problems with ‘extended coreference’

News that the Italian government is going to sell its remaining 45% participation in Alitalia have caused increased trading. The stock's value, yesterday €2 a share, went up to €3 a share.

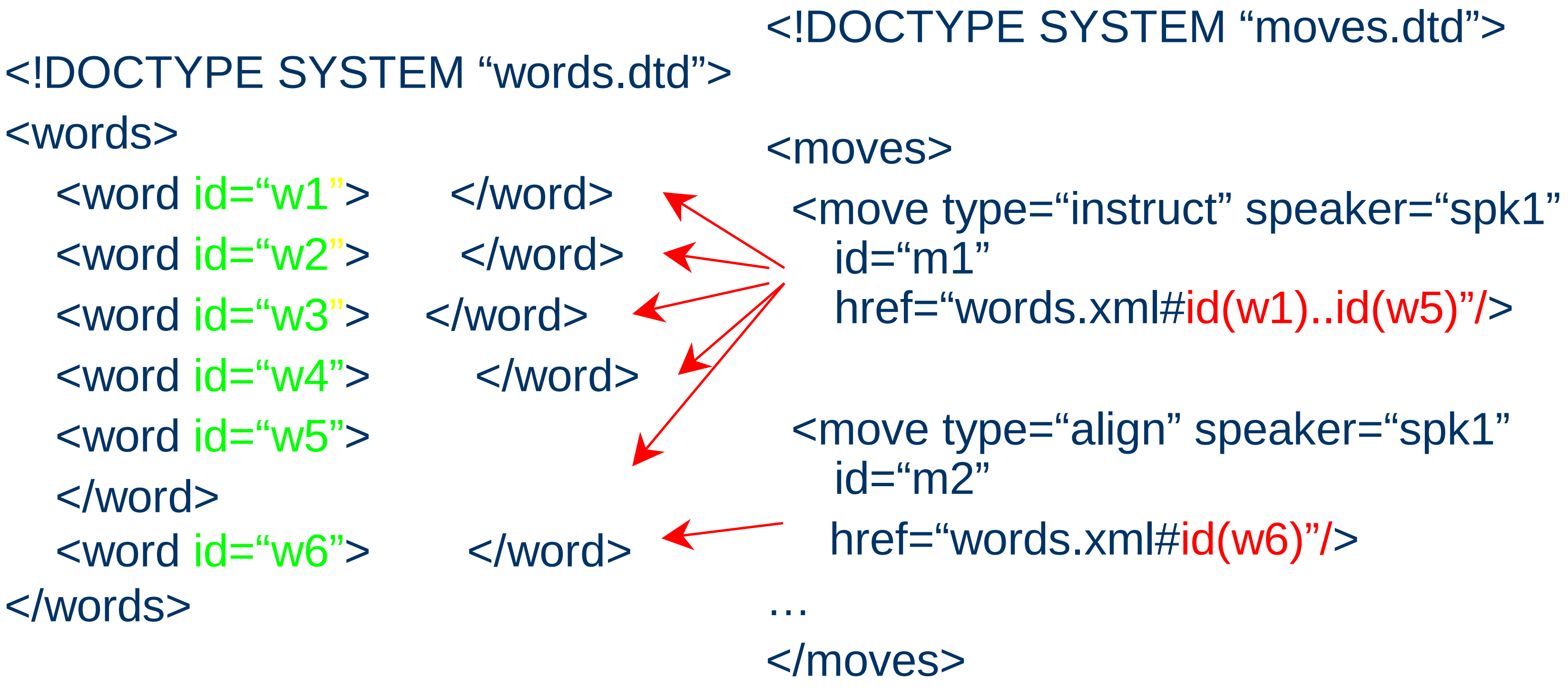
Problems

The company had already entered into negotiations to sell the company and had ample reason to believe that
the stock's value was much closer to
$2 a share than it was to
10 cents a share.

THE MATE PROJECT

- Goal: develop general tools for dialogue annotation (parsing, dialogue acts, coreference)
  - AND ‘codes of good practice’
- Markup:
  - XML
  - Standoff
- The workbench: McKelvie et al, 2001
- URL: mate.nis.sdu.dk
- Continuation: NITE (and NXT)

EXAMPLE OF STANDOFF



```
<!DOCTYPE SYSTEM “words.dtd”>
<words>
    <word id=“w1”>      </word>
    <word id=“w2”>       </word>
    <word id=“w3”>    </word>
    <word id=“w4”>        </word>
    <word id=“w5”>
    </word>
    <word id=“w6”>        </word>
</words>
```

```
<!DOCTYPE SYSTEM “moves.dtd”>

<moves>
  <move type=“instruct” speaker=“spk1”
     id=“m1”
     href=“words.xml#id(w1)..id(w5)”/>

  <move type=“align” speaker=“spk1”
     id=“m2”
     href=“words.xml#id(w6)”/>
…
</moves>
```

COREFERENCE IN MATE

- The problem with coreference (and any higher-level annotation): different tasks require different annotation
  - E.g., MUC-style annotation INSTRUCTIONS appropriate for IE but problematic from a semantic point of view
- Conclusions:
  - Unlikely that single annotation instructions useful for all types of ‘coreference annotation’
  - But it should be possible to develop a universal MARKUP SCHEME (supported by a general-purpose tool)
- Proposal:
  - markup scheme
  - suggestions for using markup tools for different types of annotation: MUC-style, DRAMA-style, MapTask-style

MATE coreference markup

- Key ideas of the markup scheme:
  - separate coreference LINKS from coreference MARKABLES
  - Use standoff
  - Specify different types of relations
- Motivation: Multiple relations
- From TEI (via Bruneseaux / Romary)

Links in the Text Encoding Initiative

```
<seg lang=FRA id=FR001>Jean aime Marie</seg>
<seg lang=ENG id=EN001>John loves Mary</seg>
<link type=translation targets="EN001 FR001">
```

ANAPHORIC RELATIONS IN A DISCOURSE MODEL

We're gonna take engine E3

and shove IT to Corning

DE1 DE2 DE3 ….

DE1=E3
take(we,DE1)

DE2=DE1
DE3=Corning
shove(we,DE2,DE3)

INDEPENDENT LINKS IN MATE

coref.xml:

```
…
<de ID="de00">we</de>'re gonna take
     <de ID="de01"> the engine E3 </de>
and shove <de ID="de02"> it </de> over
     to <de ID="de03">Corning</de>,
hook <de ID="de04"> it </de> up to
     <de ID="de05">the tanker car</de>...

<link href="coref.xml#id(de02)"  type="ident">
      <anchor href="coref.xml#id(de01)"/>
</link>
```

IDENTITY AND PREDICATION

```
<de ID="de01">Henry Higgins</de>,
    who was formerly
        <de ID="de02"> sales director of
            Sudsy soap </de>,
     became
        <de ID="de03"> president of Dreamy
            Detergents </de>

<link href="coref.xml#id(de02)"  type="REL">
      <anchor href="coref.xml#id(de01)"/>
</link>
```

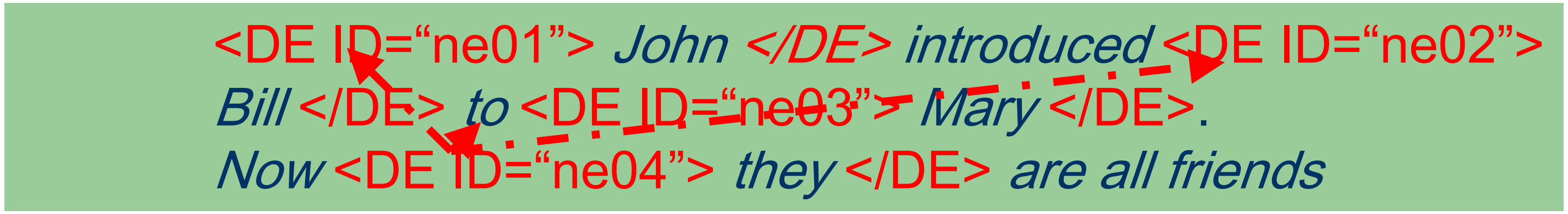
MUC:
IDENT

PROP

INDEPENDENT LINKS AND BRIDGING

- Independent links make it possible to have
  - Both identity link and bridging link
  - Multiple bridging links

Marking multiple semantic relations



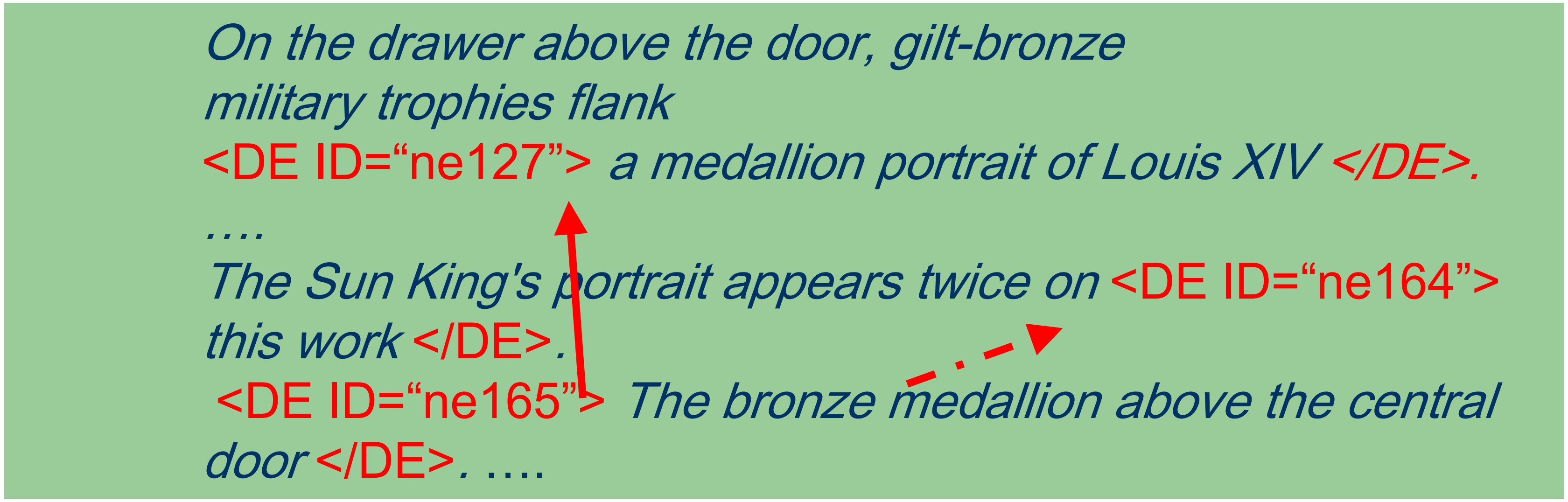
<DE ID="ne01"> *John </DE>* *introduced* <DE ID="ne02"> *Bill* </DE> *to* <DE ID="ne03"> *Mary* </DE>.
Now <DE ID="ne04"> *they* </DE> *are all friends*

```
<LINK  HREF="ne04" REL="has-element">
          <ANCHOR ANTECEDENT="ne01" /> </LINK>
```

```
<LINK  HREF="ne04" REL="has-element">
          <ANCHOR ANTECEDENT="ne02" /> </LINK>
```

```
<LINK  HREF="ne04" REL="has-element">
          <ANCHOR ANTECEDENT="ne03" /> </LINK>
```

Marking multiple semantic relations

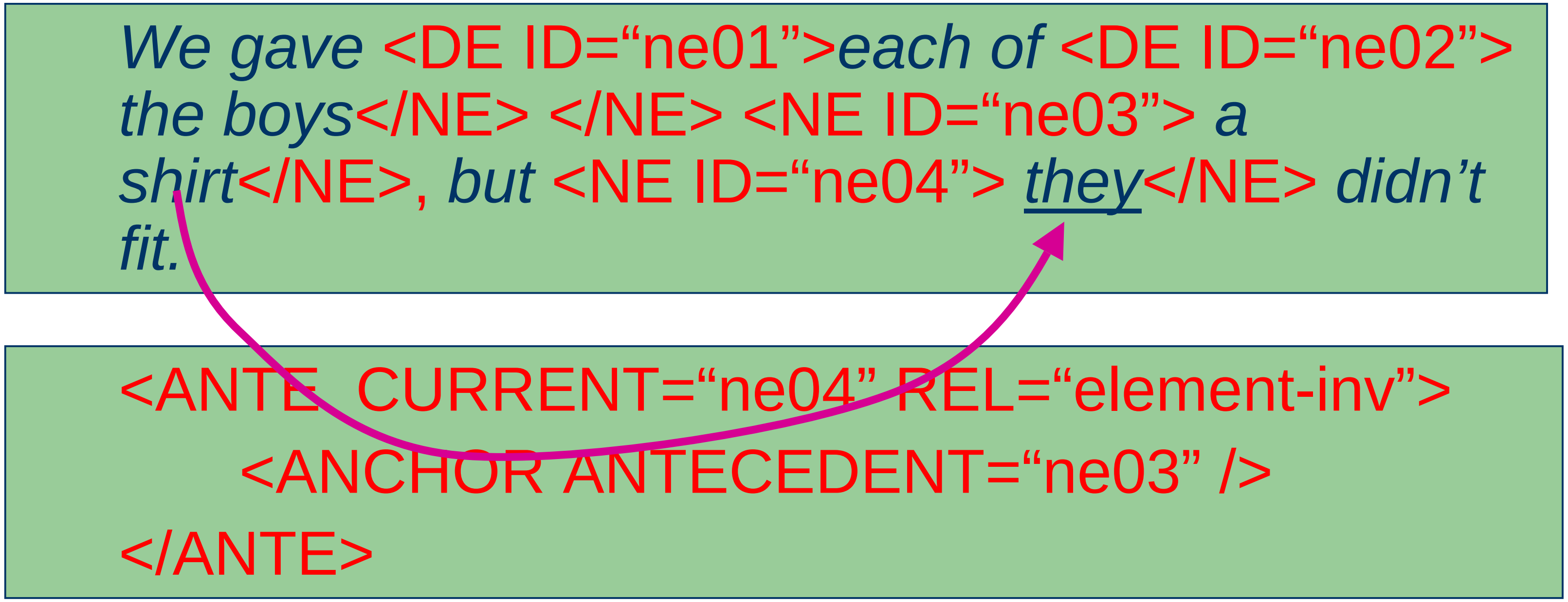


On the drawer above the door, gilt-bronze military trophies flank
<DE ID=“ne127”> *a medallion portrait of Louis XIV* </DE>.
….
The Sun King's portrait appears twice on <DE ID=“ne164”>
this work </DE>.
<DE ID=“ne165”> *The bronze medallion above the central door* </DE>. ….

<LINK HREF=“ne165” REL=“ident”>
<ANCHOR ANTECEDENT=“ne127” /> </LINK>

<LINK HREF=“ne165” REL=“part”>
<ANCHOR ANTECEDENT=“ne164” /> </LINK>

Marking bridging relations



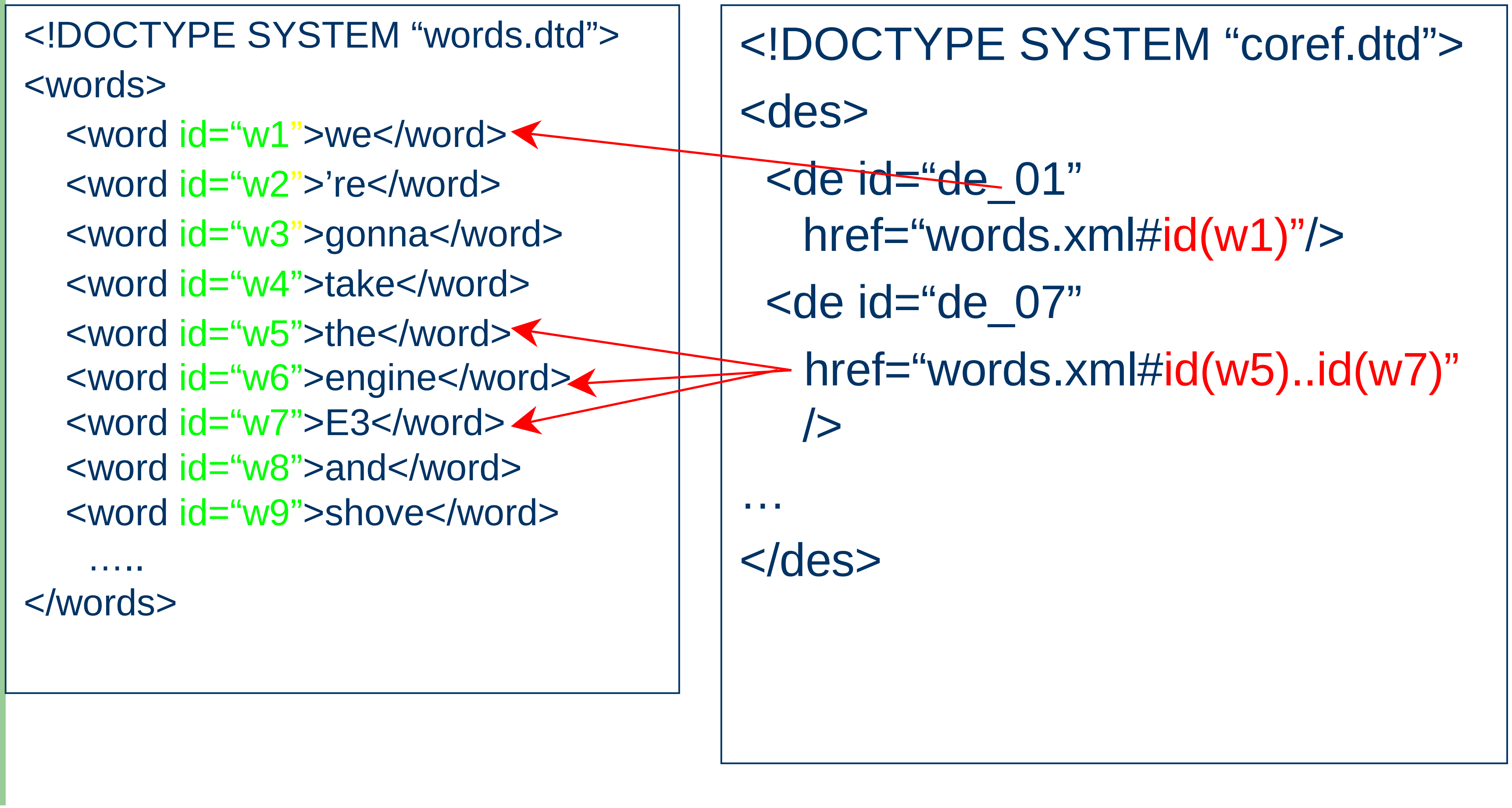
We gave <DE ID=“ne01”>*each of* <DE ID=“ne02”> *the boys*</NE> </NE> <NE ID=“ne03”> *a shirt*</NE>, *but* <NE ID=“ne04”> *they*</NE> *didn’t fit.*

<ANTE CURRENT=“ne04” REL=“element-inv”>
    <ANCHOR ANTECEDENT=“ne03” />
</ANTE>

TYPES OF BRIDGING RELATIONS

- Perhaps later when talking about GNOME

COREFERENCE STANDOFF



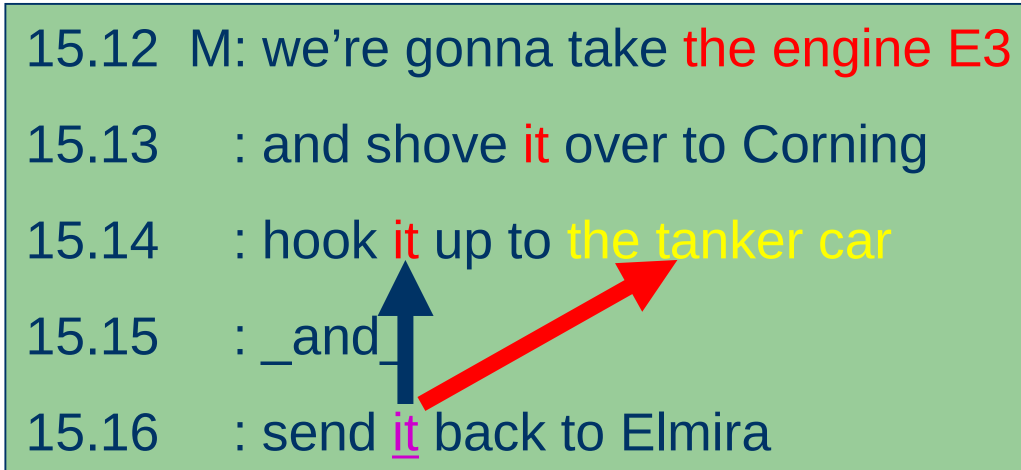
```
<!DOCTYPE SYSTEM “words.dtd”>
<words>
   <word id=“w1”>we</word>
   <word id=“w2”>’re</word>
   <word id=“w3”>gonna</word>
   <word id=“w4”>take</word>
   <word id=“w5”>the</word>
   <word id=“w6”>engine</word>
   <word id=“w7”>E3</word>
   <word id=“w8”>and</word>
   <word id=“w9”>shove</word>
     …..
</words>
```

```
<!DOCTYPE SYSTEM “coref.dtd”>
<des>
  <de id=“de_01”
     href=“words.xml#id(w1)”/>
  <de id=“de_07”
     href=“words.xml#id(w5)..id(w7)”
     />
…
</des>
```

AMBIGUITY VS. MULTIPLE RELATIONS

- The MATE markup scheme included methods for distinguishing between MULTIPLE RELATIONS and AMBIGUITY
  - (More on ambiguity below)

AMBIGUOUS ANAPHORIC EXPRESSIONS



15.12 M: we’re gonna take the engine E3

15.13 : and shove it over to Corning

15.14 : hook it up to the tanker car

15.15 : _and_

15.16 : send it back to Elmira

(from the TRAINS-91 dialogues collected at the University of Rochester)

Ambiguous anaphoric expressions in the MATE/GNOME scheme

3.3: <NE ID="ne01">*engine E2*</NE> to
<NE ID="ne02">*the boxcar at … Elmira*</NE>

5.1: and send <NE ID="ne03">*it*</NE> to
<NE ID="ne04">*Corning*</NE>

```
<ANTE  CURRENT="ne03" REL="ident">
      <ANCHOR ANTECEDENT="ne01" />
      <ANCHOR ANTECEDENT="ne02" />
</ANTE>
```

Other markup ideas in MATE

- Exophora:
  - <UNIVERSE> elements
- Discourse deixis:
  - <SEG> elements
- Multiple languages
  - Some suggestions about how to deal with zero anaphora in Italian etc

THE GNOME ANNOTATION

- Goal: study factors that affect sentence planning, particularly the form of referring expressions
- The corpus used to study:
  - Salience (Poesio et al 2000, 2004; Poesio and Nissim 2001; Poesio and Modjeska 2002, 2006)
  - Statistical generation (Poesio et al, 1999; Poesio, 2000; Cheng, Poesio and Henschel, 2001; Karamanis et al, 2004a, 2004b)
  - Bridging references (Poesio et al, 2002; Poesio, 2003; Poesio et al, 2004)
  - Anaphora resolution (Poesio and Alexandrov-Kabadjov, 2004; Poesio et al, 2005)

FROM MATE TO GNOME

- Annotation manual
  - Detailed instructions for several types of annotation, including anaphora
  - Agreement studies, particularly for bridging relations
- Markup scheme:
  - based on MATE, but no standoff (no tools!)
  - added UNIT (and other tags – e.g., MOD)
    - Mostly to compare several definitions of UTTERANCE Requires second type of MARKABLE

The GNOME markup scheme for anaphoric information

```
<NE ID=“ne07”>Scottish-born, Canadian based jeweller, Alison Bailey-Smith</NE>
<NE ID=“ne08”> <NE ID=“ne09”>Her</NE> materials</NE>
```

```
<ANTE  CURRENT=“ne09” REL=“ident”>
     <ANCHOR ANTECEDENT=“ne07” />
</ANTE>
```

GUIDELINES

- A crucial part of the task of defining an annotation is the development of guidelines
  - What counts as markable
  - Resolving ambiguities
- Two main objectives:
  - Ensure reliability
  - Limit amount of work

MUC guidelines

- From Hirschman & Sundheim
- E.g., markable guidelines

The GNOME annotation manual

- ONLY ANAPHORIC RELATIONS IN WHICH BOTH ANAPHORA AND ANTECEDENT REALIZED USING NPs
  - No ellipsis
  - No discourse deixis
- DETAILED INSTRUCTIONS FOR MARKABLES
  - ALL NPs are treated as markables, INCLUDING PREDICATIVE NPS AND EXPLETIVES (use attributes to identify non-referring expressions)
  - Markables identified by hand!!
- Online version:
  - http://www.hcrc.ed.ac.uk/~poesio/GNOME/anno_manual_4.html

Limiting the amount of work

- Restrict the extent of the annotation:
  - ALWAYS MARK AT LEAST ONE ANTECEDENT FOR EACH EXPRESSION THAT IS ANAPHORIC IN SOME SENSE, BUT NO MORE THAN ONE IDENT AND ONE BRIDGE;
  - ALWAYS MARK THE RELATION WITH THE CLOSEST PREVIOUS ANTECEDENT OF EACH TYPE;
  - ALWAYS MARK AN IDENTITY RELATION IF THERE IS ONE; BUT MARK AT MOST ONE BRIDGING RELATION

RELIABILITY OF COREF

Agreement on annotation

- Crucial requirement for the corpus to be of any use, is to make sure that annotation is RELIABLE (I.e., two different annotators are likely to mark in the same way)
- E.g., make sure they can agree on part-of-speech tag
  - … we walk in SNAKING lines (JJ? VBG?)
- Or on attachment
- Agreement more difficult the more complex the judgments asked of the annotators
  - E.g., on givenness status
- The development of the annotation likely to follow a develop / test / redesign test
  - Task may have to be simplified

A measure of agreement: the K statistic

- Carletta, 1996: in order for the statistics extracted from an annotation to be reproducible, it is crucial to ensure that the coding distinctions are understandable to someone other than the person who developed the scheme
- Simply measuring the percentage of agreement does not take chance agreement into account
- The K statistic (Siegel and Castellan, 1988):
  - K=0: no agreement
  - .6 <= K < .8: tentative agreement
  - .8 <= K <= 1: OK agreement

Agreement on familiarity (Poesio and Vieira, 1998)

Annotators asked to classify about 1,000 definite descriptions from the ACL/DCI corpus (Wall Street Journal texts) into three classes:

- DIRECT ANAPHORA: *a house … the house*
- DISCOURSE-NEW: *the belief that ginseng tastes like spinach is more widespread than one would expect*
- BRIDGING DESCRIPTIONS: *the flat … the living room; the car … the vehicle*

A `knowledge-based' classification of bridging descriptions (Vieira, 1998)

- Based on LEXICAL RELATIONS such as synonymy, hyponymy, and meronimy, available from a lexical resource such as WordNet
  the flat … the living room

- The antecedent is introduced by a PROPER NAME
  Bach … the composer

- The anchor is a NOMINAL MODIFIER introduced as part of the description of a discourse entity:
  selling discount packages … the discounts

… continued

- The anchor is introduced by a VP:
  Kadane oil is currently drilling two oil wells. The *activity…*

- The anchor is not explicitly mentioned in the text, but is a `discourse topic’
  the industry (in a text about oil companies)

- The resolution depends on more general commonsense knowledge
  last week’s earthquake … the suffering people

Results

- Agreement over three classes: K=.68
  - K=.63 if make further distinction between LARGER SITUATION and UNFAMILIAR
  - K = .73 for first mention / subsequent mention
- Subjects didn’t always agree on the classification of an antecedent
- Bridging descriptions:
  - Disagreement = 70%
  - K (bridging / non bridging) = .24

Achieving agreement (but not completeness) in GNOME

- RESTRICTING THE NUMBER OF RELATIONS
  - IDENT (*John … he, the car … the vehicle*)
  - ELEMENT (*Three boys … one (of them)*)
  - SUBSET (*The vases … two (of them) …*)
  - Generalized POSSession (*the car … the engine*)
  - OTHER (when no other connection with previous unit)

GNOME: Agreement results on bridging references

- RESULTS (2 annotators, anaphoric relations for 200 NPs)
  - Only 4.8% disagreements ON ANCHORS
  - But 73.17% of relations marked by only one annotator

Problem: K for antecedents

- Problem: the most obvious ‘labels’ for measuring agreement over antecedents are the anaphoric chains
- But the longer the chain, the less likely that all coders will include all mentions in it
  - Stats: how many cases of perfect agreement in our study?
- Need a coefficient of agreement that takes into account partial agreement

The GNOME corpus

- Initiated at the University of Edinburgh, HCRC / continued at the University of Essex
- 3 Genres
    - Descriptions of museum pages (including the ILEX/SOLE corpus)
    - ICONOCLAST corpus (500 pharmaceutical leaflets)
    - Tutorial dialogues from the SHERLOCK corpus
- Small size
    - 3000 NPs in each genre, 10000 NPs total
    - Around 1500 sentences

An example museum text

Cabinet on Stand

The decoration on this monumental cabinet refers to the French king Louis XIV's military victories. A panel of marquetry showing the cockerel of France standing triumphant over both the eagle of the Holy Roman Empire and the lion of Spain and the Spanish Netherlands decorates the central door. On the drawer above the door, gilt-bronze military trophies flank a medallion portrait of Louis XIV. In the Dutch Wars of 1672 - 1678, France fought simultaneously against the Dutch, Spanish, and Imperial armies, defeating them all. This cabinet celebrates the Treaty of Nijmegen, which concluded the war. Two large figures from Greek mythology, Hercules and Hippolyta, Queen of the Amazons, representatives of strength and bravery in war, appear to support the cabinet.

The fleurs-de-lis on the top two drawers indicate that the cabinet was made for Louis XIV. As it does not appear in inventories of his possessions, it may have served as a royal gift. The Sun King's portrait appears twice on this work. The bronze medallion above the central door was cast from a medal struck in 1661 which shows the king at the age of twenty-one. Another medallion inside shows him a few years later.

Other information marked up in the GNOME corpus

- Syntactic features: grammatical function, agreement
- Semantic features:
  - Logical form type (term / quantifier / predicate)
  - `Structure’: Mass / count, Atom / Set
  - Ontological status: abstract / concrete, animate
  - Genericity
  - ‘Semantic’ uniqueness (Loebner, 1985)
- Discourse features:
  - Deixis
  - Familiarity (discourse new / inferrable / discourse old) (using anaphoric annotation)
- A number of additional features automatically computed (e.g., is an entity the current CB, if any)

The GNOME annotation of NEs

```
<ne id="ne109"
cat="this-np" per="per3" num="sing" gen="neut“
gf="np-mod"
lftype="term" onto="concrete“ ani="inanimate"
structure="atom" count="count-yes"
generic="generic-no“deix="deix-yes"
reference="direct" loeb="disc-function" >  this
monumental cabinet </ne>
```

Coding for familiarity

- Poesio / Vieira: tried to classify all types of familiarity, including hearer old (‘larger situation’)
  - Serious problems
- GNOME: only discourse old
- The problem remain of how to mark the rest RELIABLY
- More recent efforts:
  - MULI project (Baumann et al 2004)
  - Nissim et al 2004

Follow-up: VENEX, ARRAU

- Looking at DIALOGUE
  - Marking EXOPHORA
- Semi-automatic identification of markables
- Using more modern tools (MMAX)

VENEX (Poesio, Bristot, Delmonte, Tonelli 2004)

- A corpus of anaphoric information in Italian
- Both written (WSJ-style) and spoken (MapTask-style) text
- Both corpora automatically parsed using the GETARUN parser (Delmonte and Pianta)
- Annotated using MMAX
- Issues of interest:
  - Clitics in Italian
  - Misunderstandings

DEVELOPMENTS FOR THE VENEX ANNOTATION

- Annotation of deictic references to landmarks in MapTask-style dialogues
  - Developing techniques for marking both anaphoric and deictic differences in interpretation
- Annotation of empty anaphors
- Additional distinction in bridging references between PART-OF (the wheel) and ATTRIBUTES (the width)

MMAX (Mueller and Strube, 2002, 2003)

- A tool for annotation especially of anaphoric information
- Based on XML technology and (a simplified form of) standoff markup
- Implemented in Java
- Available from the European Media Lab, Heidelberg

Standoff in MMAX: Words

```
<?xml version='1.0' encoding='ISO-8859-1'?>
<!DOCTYPE words SYSTEM "words.dtd">
<words>   <word id="word_1">Leben</word>
          <word id="word_2">und</word>
          <word id="word_3">Wirken</word>
          <word id="word_4">von</word>
          <word id="word_5">Georg</word>
          <word id="word_6">Philipp</word>
          <word id="word_7">Schmitt</word>
          <word id="word_8">.</word>
          <word id="word_9">Am</word>
          <word id="word_10">28.</word>
          <word id="word_11">Oktober</word>
          <word id="word_12">1808</word>
          <word id="word_13">wurde</word>
          <word id="word_14">Georg</word>
          <word id="word_15">Philipp</word>
          <word id="word_16">Schmitt</word>
```

Standoff in MMAX: Markables

```
<?xml version="1.0"?>
<markables>
……
<markable id="markable_36"
span="word_5,word_6,word_7“
np_form="NE" agreement="3M" grammatical_role="other">
</markable>
….
<markable id="markable_37"
 span="word_14,word_15,word_16"
np_form="NE" agreement="3M" grammatical_role="other">
</markable>

</markables>
```

Standoff in MMAX: Anaphoric information

```
<?xml version="1.0"?>
<markables>
……
<markable id="markable_36"
span="word_5,word_6,word_7“
np_form="NE" agreement="3M" grammatical_role="other"
member="set_22" > </markable>
….
<markable id="markable_37"
 span="word_14,word_15,word_16"
np_form="NE" agreement="3M" grammatical_role="other"
member="set_22" ></markable>
…….
</markables>
```

Standoff in MMAX: Markables

```
<?xml version='1.0' encoding='ISO-8859-1'?>
<markables>
<markable id="markable_1" form="NP" span="word_0"></markable>
<markable id="markable_2" form="NP“ span="word_4..word_8">
</markable>
<markable id="markable_3" form="NP" span="word_10"></markable>
<markable id="markable_4" form="NP" span="word_18..word_21">
</markable>
<markable id="markable_5" form="NP" span="word_16..word_21">
</markable>
<markable id="markable_6" form="NP" span="word_23..word_24">
</markable>
<markable id="markable_7" form="NP" span="word_13..word_24">
</markable>
```

Other annotation efforts

- Large-scale annotation of identity relations:
  - Prague Dependency Treebank
  - The Tuebingen Treebank (Kuebler, Versley, Hinrichs)
  - Ontonotes
- Associative relations:
  - Gardent (French)
  - Caselli (Italian)

PRAGUE DEPENDENCY TREEBANK

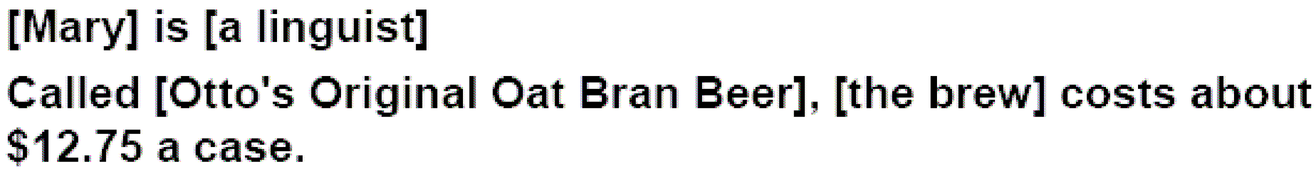
- Using DEEP SYNTACTIC STRUCTURE to define markables
  - Cleanest solution for zero anaphora
- Full MATE scheme:
  - Exophora
  - Discourse deixis (SEG)

ONTONOTES

- Large effort to create corpus semantically annotated at different levels:
  - Wordsense (using Omega Ontology)
  - Propbank
  - Coreference
- Started November 2005

Ontonotes coreference (Ramshaw & Weischedel)

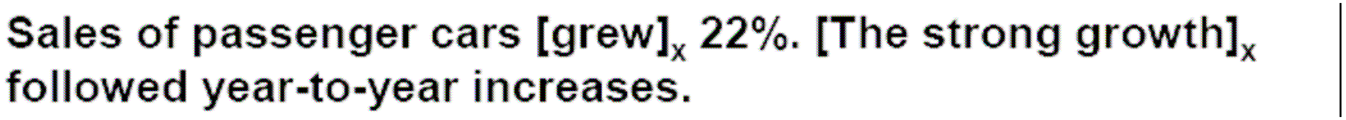
Attribution not marked as coref (unlike MUC and ACE)



[Mary] is [a linguist]

Called [Otto's Original Oat Bran Beer], [the brew] costs about $12.75 a case.

Identity only, but also references to EVENTS



Sales of passenger cars [grew]$_x$ 22%. [The strong growth]$_x$ followed year-to-year increases.

AGREEMENT ON ANAPHORA, 2

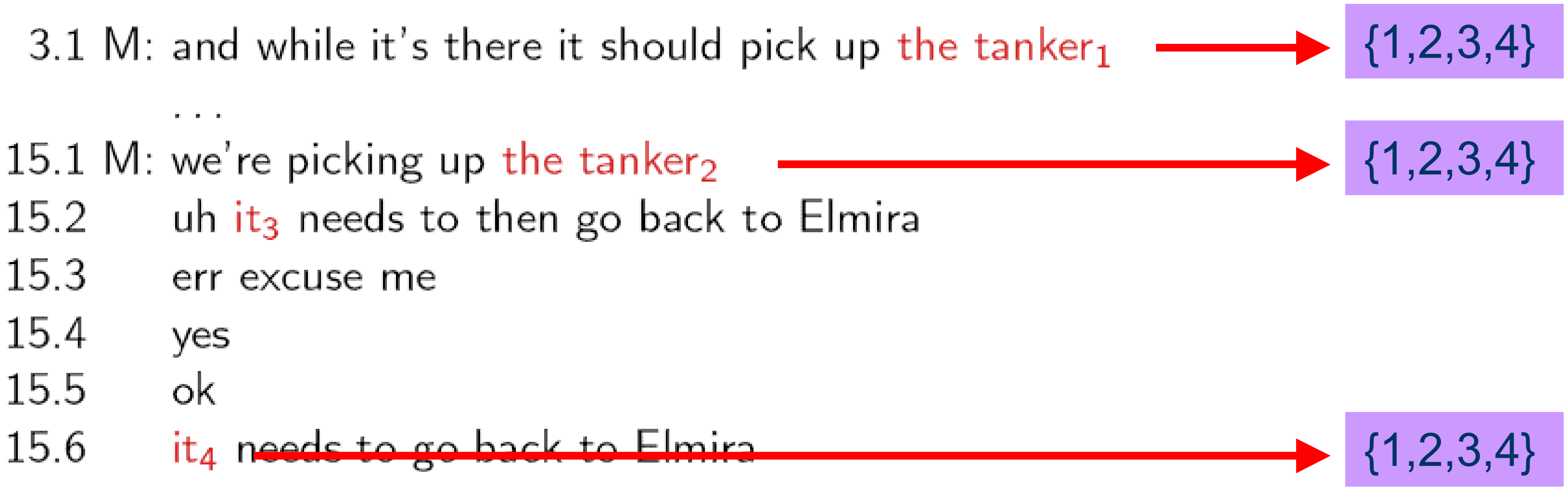
- K not appropriate for anaphora
- Not all cases of disagreement are due to a poor coding scheme: the case of ambiguity

Problem: K for anaphora

- Problem: the most obvious ‘labels’ for measuring agreement over anaphora are the anaphoric chains
- But the longer the chain, the less likely that all coders will include all mentions in it
- Need a coefficient of agreement that takes into account partial agreement

K for anaphora

The most obvious ‘label’ for computing agreement on anaphora: the chains (see e.g., Passonneau, 2004)



3.1 M: and while it's there it should pick up the tanker$_1$ → {1,2,3,4}

...

15.1 M: we're picking up the tanker$_2$ → {1,2,3,4}

15.2 uh it$_3$ needs to then go back to Elmira

15.3 err excuse me

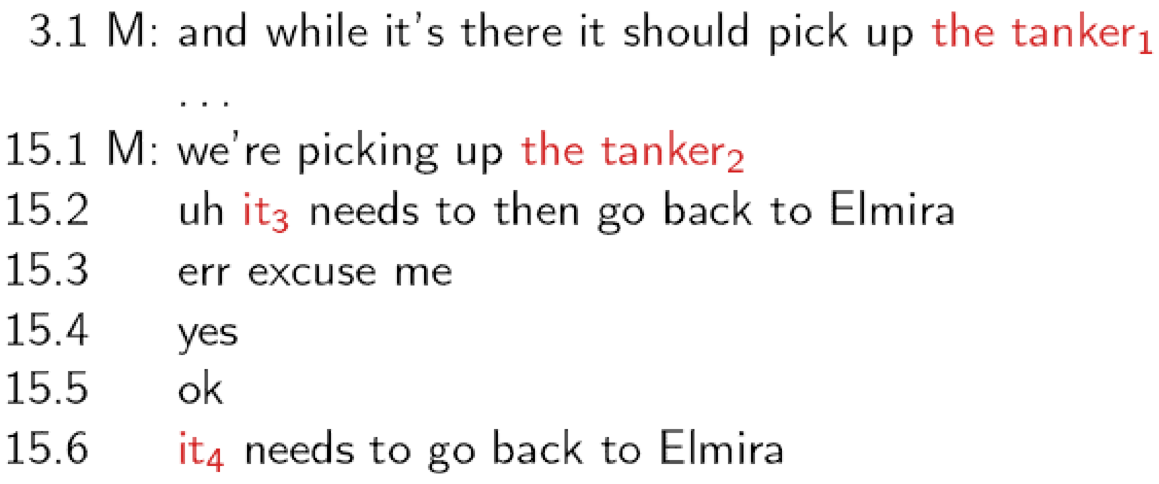
15.4 yes

15.5 ok

15.6 it$_4$ needs to go back to Elmira → {1,2,3,4}

The problem

Problem: especially in long texts, most annotators forget some mention



	A	B
3.1 M: and while it's there it should pick up the tanker$_1$	{1,2,4}	{1,2,3}
...		
15.1 M: we're picking up the tanker$_2$	{1,2,4}	{1,2,3}
15.2 uh it$_3$ needs to then go back to Elmira		
15.3 err excuse me		
15.4 yes		
15.5 ok		
15.6 it$_4$ needs to go back to Elmira	{1,2,4}	{1,2,3}

Need a coefficient that gives ‘partial credit’

From K to α

- Krippendorff’s α a more general coefficient of agreement that can also be used for non-categorical decisions

FROM K TO α

$$
\begin{aligned}
K &= \frac{P_o - P_e}{1 - P_e} \\
P_o &= 1 - \text{observed disagreement} = 1 - D_o \\
P_e &= 1 - \text{expected disagreement} = 1 - D_e \\
K &= \frac{(1 - D_o) - (1 - D_e)}{1 - (1 - D_e)} \\
&= 1 - \frac{D_o}{D_e} \\
&= \alpha
\end{aligned}
$$

FROM K TO α

	A	B	C	D
A	3	0	3	2
B	1	6	2	4
C	0	3	2	2
D	1	1	1	12

FROM K TO α

	A	B	C	D
A	3	0	3	2
B	1	6	2	4
C	0	3	2	2
D	1	1	1	12

Distance metrics in α

$$\begin{aligned}\alpha &= 1 - \frac{\text{observed disagreement}}{\text{expected disagreement}}\\ &= 1 - \frac{\text{mean disagreement per item}}{\text{mean disagreement}}\\ &= 1 - \frac{\frac{1}{\mathbf{ic}(\mathbf{c}-1)}\sum_{i\in I}\sum_{k\in K}\sum_{k'\in K}\mathbf{n}_{ik}\mathbf{n}_{ik'}\mathbf{d}_{kk'}}{\frac{1}{\mathbf{ic}(\mathbf{ic}-1)}\sum_{k\in K}\sum_{k'\in K}\mathbf{n}_{k}\mathbf{n}_{k'}\mathbf{d}_{kk'}}\end{aligned}$$

$d_{kk'}$: a task-dependent
DISTANCE METRIC

Distance metrics for anaphora

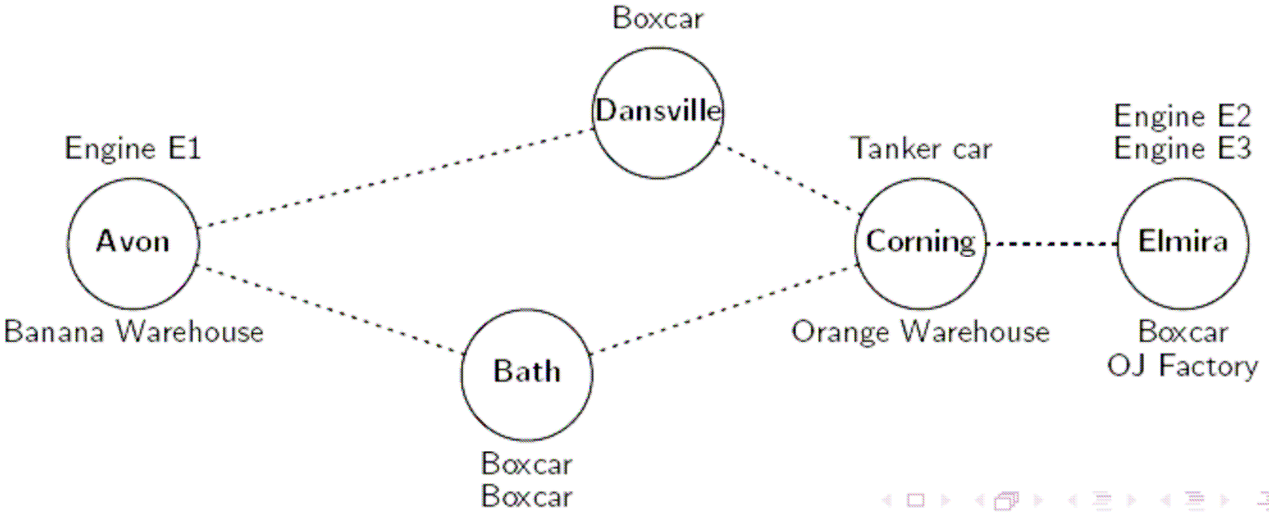
$$\mathbf{d}_{AB}^{\text{Passonneau}} = \begin{cases} 0 & \text{if } A = B \\ {}^{1}\!/_{3} & \text{if } A \subset B \text{ or } B \subset A \\ {}^{2}\!/_{3} & \text{if } A \cap B \neq \emptyset \text{, but } A \not\subset B \text{ and } B \not\subset A \\ 1 & \text{if } A \cap B = \emptyset \end{cases}$$

$$\mathbf{d}_{AB}^{\text{Jaccard}} = 1 - \frac{|A \cap B|}{|A \cup B|}$$

$$\mathbf{d}_{AB}^{\text{Dice}} = 1 - \frac{2\,|A \cap B|}{|A| + |B|}$$

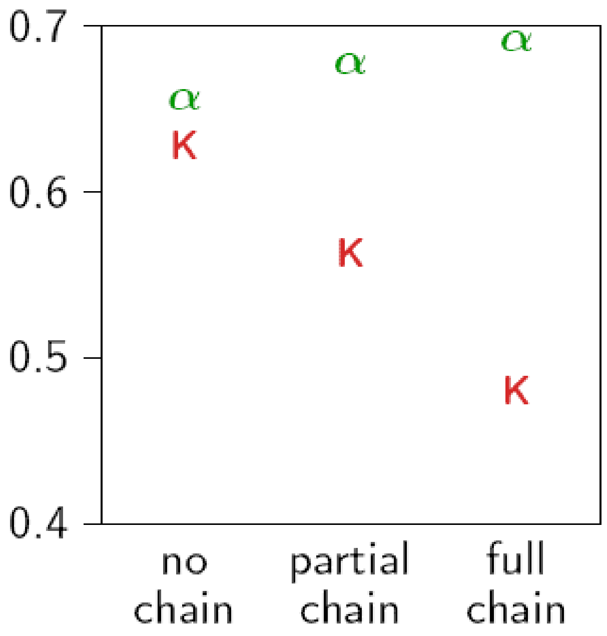
Example

- 18 naïve subjects
- Dialogue 3.2 from the TRAINS 91 corpus
- MMAX 2 annotation tool
- Map of the “TRAINS world”

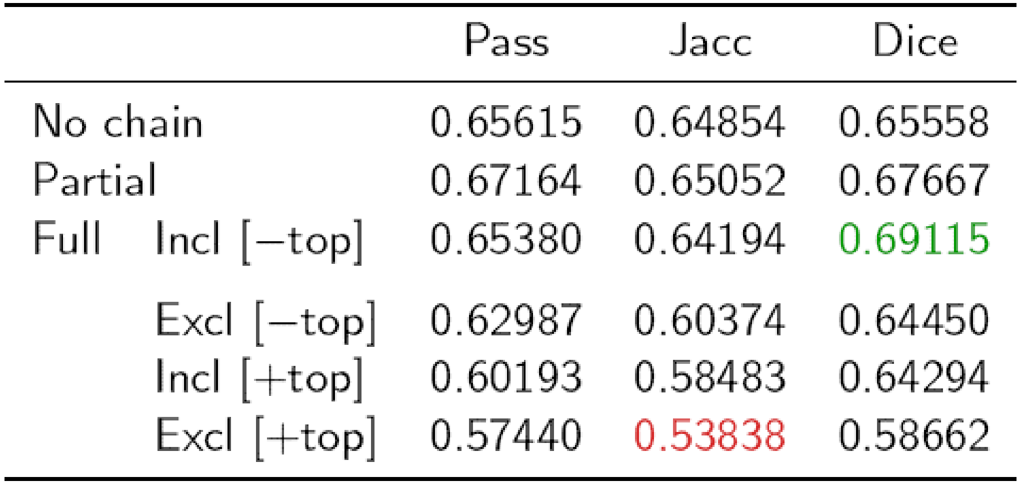


K vs α

Chain	K	α
None	0.62773	0.65558
Partial	0.56325	0.67667
Full	0.48042	0.69115



α’s dependence on distance metric



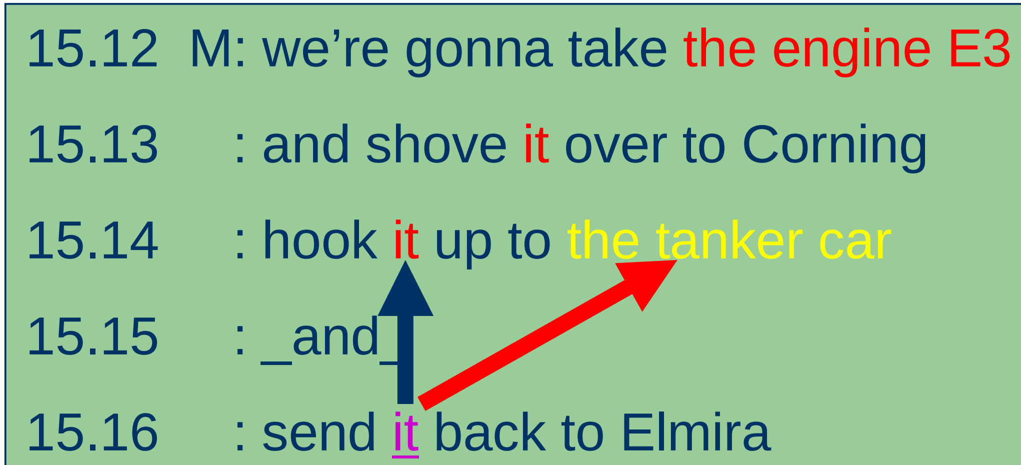
		Pass	Jacc	Dice
No chain		0.65615	0.64854	0.65558
Partial		0.67164	0.65052	0.67667
Full	Incl [−top]	0.65380	0.64194	0.69115
	Excl [−top]	0.62987	0.60374	0.64450
	Incl [+top]	0.60193	0.58483	0.64294
	Excl [+top]	0.57440	0.53838	0.58662

Caveats

- The value of α can change greatly depending on the metric you choose
- Examples:
  - ACL05
  - BRANDIAL06

AMBIGUITY

AMBIGUOUS ANAPHORIC EXPRESSIONS



15.12 M: we’re gonna take the engine E3

15.13 : and shove it over to Corning

15.14 : hook it up to the tanker car

15.15 : _and_

15.16 : send it back to Elmira

(from the TRAINS-91 dialogues collected at the University of Rochester)

Summary of results

- Virtually perfect agreement on places
- In all of the cases similar to the one in the example (about 15 in total) at least a few coders DO mark an explicit ambiguity
  - But: at least as many settle on distinct interpretations (IMPLICIT ambiguity)
- Almost half of the disagreements (18.2%) on DISCOURSE NEW / DISCOURSE OLD ambiguity, which could not be explicitly marked
- Main problem: many coders can't distinguish between AMBIGUITY and PLURALITY

An example

19.10: we need to get the bananas to Corning by 3
19.11: uh
19.12: *maybe* it 's gonna be faster if we
19.13: send E1
19.14: E1 's boxcar picks up at Dansville
19.15: instead of going back to Avon
19.16: have it go on to Corning
19.17: uh pick up the tanker get the oranges send them to Elmira
19.18: cause that 's gonna be the longest thing

Key: Full agreement One outlier Implicit Explicit

The ARRAU Annotation effort

Try it out

Conclusions: some lessons

- There is much more to context dependence that simple ‘coreference’
- Annotating context dependence is doable at least for text, but you need
  - A clear idea of the goals of the annotation
  - Some pretheoretical understanding
- Quite a few schemes now exist which have been tested in large-scale efforts
- Reliability: even ‘easy’ decisions may be quite complex
  - Identity relations: usually OK
  - Bridging relations: you have to be selective
- K not appropriate for anaphora (but α problematic as well)

Open questions

- More complex cases of bridging
- References to implicit objects (e.g.,discourse deixis): how much agreement there is among humans on the sort of antecedent?
- Ambiguity

URLs

- MATE:

  http://www.ims.uni-stuttgart.de/projekte/mate/mdag/cr/cr_1.html
- GNOME:

  http://cswww.essex.ac.uk/Research/nle/corpora/GNOME/
- ARRAU:

  http://cswww.essex.ac.uk/Research/nle/ARRAU