LOT Winter School 2005

iscourse Anaphora: Theories, Data and Applications

Bonnie Webber
School of Informatics
University of Edinburgh
January 2005

Day 3: Heuristic Approaches (continued); Corpora for Discourse Anaphora

Outline

Day 3: Heuristic Approaches to Anaphor Resolution (continued)

- Identifying and resolving definite NP anaphors
- Identifying and resolving event reference [Eckert & Strube,2000; Byron 2002; Navarretta 2004]

Day 3: Corpora for Discourse Anaphora

- Gold Standard annotation & Annotation tools
- Annotation schemata – GNOME, Penn Discourse TreeBank
- Annotated corpora – GNOME, PDTB
- Automated annotation and "noisy" gold standards [Kehler, Appelt et al 04]
- Discussion
- Summary

⇨ Heuristic algorithms for Definite NP anaphors

l from Lecture 1:

urud, 1990] reported 269/745 (36%) definite NPs in Swedish technical prose were *isolated mention* (i.e., non-anaphoric).

io & Vieira, 1998] reported 52% of definite NPs in WSJ portion of the Penn TreeBank were *discourse new* (i.e., non-anaphoric), and an additional 5% were bridging/associative anaphors (i.e., non-coreferential).

efore, one might try to *exclude* instances of non-referential and non-anaphoric te NPs at some point in the resolution process, rather than

reating them as *out-of-algorithm* errors in the evaluation;

manually excluding them right before evaluation.

?

Non-anaphoric Definite NPs

[Vieira & Poesio 2000] attempted to recognise and exclude from resolution

- **Functional definites**: Definite NPs with an ordinal or superlative modifier
  - the second woman to sail solo around the world
  - the most convenient place to stop

  or with a head such as *fact*, *belief*, *fear*, etc.
  - the belief (or fear) that one will live forever
- **Appositives & Predicate nominatives** such as
  - Glenn Cox, the president of Phillips Petroleum
  - Glenn Cox was the president of Phillips Petroleum from 1986–1990.

If resolution failed to provide an antecedent, they tried to exclude from evaluation

- Definite NPs identified as **Determined Proper Names** – e.g.
  - The Federal Trade Commision
  - The European Working Time Directive

Definites established by restrictive modifiers – e.g.

- the hotel we stayed at last summer

"Larger situation" definites – e.g.

- the sun, the pope

ess in 2-way classification of definite NPs into *discourse new* (DN) and *horic*:

recision: 60-75%
ecall: 72-85%
-score: 70-77%

ed on 1600 articles from MUC-4, and tested on 50 texts, performance of & Riloff]'s DN detector was

	R	P
Baseline (assign DN to all)	100	72.2
Syntactic Heuristics+S1+EHP+DO	81.7	82.2

Other Non-anaphoric Definite NPs

[Bean & Riloff, 1999] add two *discourse* features and a *frequency* feature, all derived from 1600 articles in the MUC-4 corpus.

- Any definite NPs found in the first sentence (S1) of a text (849 NPs)
- NPs exhibiting similar patterns to those extracted from S1s (Existential Head Pattern, EHP)

  `the N+ Government`, extracted from *the Salvadorean Government*, *the Guatemalan Government*, etc.

- additional determined proper names, identified by their appearing at least 5 times with the definite article and never with the indefinite article (Definite only, DO), such as

  the National Guard, *the FBI*, etc.

  (65 head nouns, 321 full NPs)

Discourse New Detection / Uniqueness Detection

Uryupina (2003) noticed that properties significant to *uniquely specifying* NPs

- proper names
- functional (semantic) definites
- "larger situation" definites

may differ from those significant to NPs that are *discourse new* (DN), since *uniqueness* and *DN* are independent features:

(1) [The British Navy]$_1$ is considering [a new ship]$_2$ ...
[The Navy]$_3$ would like [the new ship]$_4$ to ...

- [The British Navy]$_1$ <+DN, +*unique*>
- [a new ship]$_2$ <+DN, -*unique*>
- [The Navy]$_3$ <-DN, +*unique*>
- [the new ship]$_4$ <-DN, -*unique*>

Detectors for DN and Uniqueness

pina (2003) developed separate detectors for *DN* and *uniqueness*, both based set of 32 features:

tring-level features – e.g., capitalisation, special symbols, digits

yntactic features – e.g., part-of-speech (PoS) of NP head, type of determiner, $/s$ post-modification, c/s appositive

ontext features – e.g., distance in NPs to previous same-headed NP (if any), listance in Ss to previous same-headed NP (if any)

web-based frequency information on whole NP and its head (H)

- count("the <NP>") / count("<NP>") – rough indicator for plurals
- count("the <NP>") / count("a <NP>") – rough indicator for singulars
- count("the <H>") / count("<H>") – rough indicator for plurals
- count("the <H>") / count("a <H>") – rough indicator for singulars

Singular non-uniques occur about equally often as definites and indefinites – i.e., count("the <NP>") / count("a <NP>") $\sim$ 1.

e.g. count("the retailer")/count("a retailer") = 204551/309392 $\sim$ .65

While singular uniques occur much more often as a definite.

count("the government")/count("a government") = 5539661/1109574 $\sim$ 5

Performance of DN Detector

pina took 20 texts from the MUC-7 corpus. She trained the ML classifier *r* on each subset of 19 texts, tested it on the held out text, and then averaged sults.

nising for *precision*, the DN dectector's performance was

	Features	P	R	F
All the	All	88.5	84.3	86.3
entities	Syntax+Context	87.9	86	86.9

performance of the *uniqueness* detector was even better

Optimization	P	R	F
Best precision	95.0	83.5	88.9
Best recall	87.2	97.0	91.8
Best accuracy	87.8	96.6	92.0

Since +*unique* and -*unique* NPs are clearly separable, this suggests using a *uniqueness* detector before assessing *DN*.

To do this, Uryupina trained the *DN* detector separately on +*unique* NPs and -*unique* NPs, which improved *precision*, without a significant drop in *recall*.

	P	R	F
Uniques	85.2	68.8	76.1
Non-uniques	90.4	88.9	89.6
All	88.9	84.6	86.7

She attributes the low performance on detecting whether *uniques* are *DN* to her (incorrect) "invarient head" assumption about subsequent reference to *uniques*:

- Correct for "The British Navy", "The Navy"
- Incorrect for "Lockheed Martin Corporation", "Lockheed Martin", "Lockheed".

Nevertheless, it appears useful to determine *uniqueness* prior to attempting to detect *DN* NPs.

⇒ Heuristic approaches to event reference

l from Lecture 1 that the same forms that can be used for *coreference* onal & demonstrative pronouns, headed NPs) can also be used to refer to *act objects* (facts, events, states, propositions, speech acts, actions) associated VPs, clauses, sequences of clauses or utterances.

'm not inviting you to my party.
. *This* shouldn't offend you because ...
i. *It*'s not because I don't like you, but ...

st as it makes sense to filter out *discourse new* NPs before attempting to ve *anaphoric* ones, it also makes sense to distinguish those instances that are to refer to NP-evoked entities from those that are likely to refer to ally-evoked AOs.

Resolving Event Reference

Recall also that event reference is constrained to have an AO of the right sort for what's predicated of it.

(3) John kicked Sam. **This/That* lasted until Sam surrendered.

(4) Fred thought John kicked Sam. *This/That* lasted until he learned otherwise.

What's the problem with Example 3?

Usually hearers can find a way of interpreting a VP, clause, sequence of clauses or utterance in terms of the required type of **AO**. So antecendents of DDs are rarely far away.

√ As one is attending to instances likely to refer to **AO**s, one can simultaneously try to identify its antecedent and coerce it to the right type of **AO**.

nominal Anaphora in Spoken vs. Written Language

ence resolution algorithms designed for written text are not necessarily opriate for spoken dialogue because

heir unit of analysis is the sentence, where spoken language is often made of ragments;

he participants in a dialogue may not be focussing on the same entities at any given point in the dialogue.

pronouns in spoken dialogue have **no** antecedent. In the SwitchBoard us, [Eckert & Strube 2000] estimate that 33% lack an antecedent.

e needs a reference resolution algorithm specified in terms of units that ate that *common ground* has been achieved between the participants, ectly indicating what entities exist in the shared discourse model for quent anaphoric reference.

Frequency of Event Reference

[Eckert & Strube 2000] also report that antecedents for 22% of pronouns in SwitchBoard come from units other than NPs.

[Byron 2002] reports that of the 180 referential 3^{rd} person and demonstrative pronouns in 10 of the Trains93 dialogues, only about half have NP antecedents.

[Navarretta 2004] reports that antecedents for 15% of her corpus of Danish text and 48% of her corpus of Danish dialogues came from units other than NPs.

Here, briefly consider these three efforts to resolve both NP coreferential pronouns and pronouns referring to AOs.

Eckert & Strube Algorithm

rt & Strube's approach [1999,2000] was designed for spoken dialogue, ugh it can be adapted for written text. It identifies the *antecedent* of a oun that refers to an **AO**, but not the **AO** itself that is required by its cative context.

e 1: Examine the predicative context of the input to determine whether it is *mpatible* with the input referring to an individual, concrete entity *ompatible* – *I) or *incompatible* with the input referring to an **AO** *compatible* – *A) or compatible with both (*ambiguous*).

e 2: Resolve the input or, if unresolvable, classify it as a vague 3^{rd} person oun (*VagPro*) or vague demonstrative (*VagDem*).

ronoun: An anaphor in a position where it cannot refer to an **AO**:

Equative constructions, where the anaphor is equated with a concrete ndividual referent – e.g., *x is my car.*

Copula constructions whose predicate adjectives can only be applied to a oncrete individual referent – e.g., *x is loud, expensive, tasty.*

Arguments of verbs describing physical contact/stimulation – e.g., *break x, drink x*

Phase 1: *I and *A

Pronouns are characterised as *I or *A (or neither), depending on an (incomplete) set of approximations of preferences.

***I pronoun**: An anaphor in a position where it cannot refer to individual, concrete entities:

- *Equative* constructions, where the anaphor is equated with an **AO** e.g., *this/that/it is only a suggestion.*
- *Copula* constructions whose predicate adjectives can only be applied to **AO**s e.g., *this/that/it is false.*
- Arguments of propositional attitude verbs or of verbs that mainly take S'-complements – e.g. *assume this/that/it say this/that/it*.
- Object of *do* – i.e., *do it/this/that*.
- *Equative* constructions, where the anaphor is equated with a reason – e.g. *this/that/it <be> because/why* – or a cleft *how* construction – *this/that/it <be> how*.

Phase 2: 3^{rd} person pronoun

case PRO is *A
  if resolveInd(PRO) *Using Strube Alg*
    then classify as IPro
    else classify as VagPro
case PRO is *I
  if resolveDD(PRO)
    then classify as DDPro
    else classify as VagPro
case PRO is *ambiguous*
  if resolveInd(PRO) *Using Strube Alg*
    then classify as IPro
    else if resolveDD(PRO) **then** classify as DDPro
      else classify as VagPro

Phase 2: Demonstrative

DEM is *I
if resolveDD(DEM)
    then classify as DDDem
    else classify as VagDem
DEM is *A
if resolveInd(DEM) *Using Strube Alg*
    then classify as IDem
    else classify as VagDem
DEM is *ambiguous*
if resolveDD(DEM) ***different preference than PRO****
    then classify as DDDEM
    else if resolveInd(DEM) **then** classify as IDem
        else classify as VagDem

e 2

esolveInd, S-list({Europe and NA}, schizophrenia) → schizophrenia

esolveInd, S-list({Europe and NA}, schizophrenia, those born in mid-March, March) → schizophrenia

esolveInd, S-list(schizophrenia, the offspring of . . . , their pale-skinned *ères*, the countryside, dark-skinned immigrants, northern countries, people in cities, cities, those born in the countryside) → schizophrenia

ResolveDD → [$_S$And the offspring . . .]

Examples

Schizophrenia is mysterious. In Europe and NA it_1 is commonest in those born in mid-March. It_2 is more prevalent among people born in cities than those born in the countryside. And the offspring of dark-skinned immigrants to northern countries are 3-4 times more likely than their pale-skinned *confrères* to suffer from it_3.

$That_4$ looks like a disparate collection of facts.
[The Economist, 9 Feb 02, *Let the Sun Shine in* p.83]

Phase 1
it_1: *ambiguous*?
it_2: *ambiguous*?
it_3: *ambiguous*?
$that_4$: *ambiguous*?

S-list ranking(review): old(U_i) ≺ old(U_{i-1}) ≺ mediated(U_i) ≺ mediated(U_{i-1}) ≺ new(U_i) ≺ new(U_{i-1})

PHORA [Byron 2002]

• PHORA operates on the output of the TRAINS'93 intepreter, which associates a semantic class with each argument of each predicate and with the subject or predicate adjective and predicate nominative constructions. That acts as a constraint on referring expressions – e.g.

So *it* will get there at 3pm.
`(Arrive :theme x :dest y :time z)`
x ε Movable-Object

That will take two hours.
`(Take-time :theme x :cost y)`
x ε Event

• Using semantic class information, PHORA attempts to resolve event reference as well as identifying antecedents.

in Eckert & Strube's approach, PHORA

keeps potential referents for event anaphora spearate from those for coreferential anaphors;

ses different search orders for resolving 3^{rd} person singular and demonstrative pronouns.

- 3^{rd} sg pros try NP-evoked entities first;
- Demonstrative pros try clausally-evoked entities first.

cause dialogue rather than written text, PHORA doesn't associate a potential ent with an indefinite NP or an utterance expressing REQUEST speech acts the (containing) utterance is confirmed by the other dialogue participant.

ORA uses *referring functions* [Nunberg, 1979] to coerce plausible referents to tity that meets semantic class constraints.

Event(d): d → event
Proposition(d): d → proposition
Situation(d): d → situation

(5) Engine 1 goes to Avon to get the oranges.
That will take two hours. *vs* No, I guess *that*'s not a good idea.

	Total 3^{rd} person pros (it, they, them, t'selves)	Total demonstrative pros (that, this, those)	Out of Scope	
Raw Word count	119	147	326	
Referential pros	87	111	0 524	
Excluded pros	7	11	326	
Evaluation set (Ref-Excl)	80	100	0	
LRC baseline	52	14	0	66
+ semantic type constraint	58	19	0	77
+ considering AO referents	64	56	0	12
+ different search orders	63	67	0	13

N.B. Byron notes that *without* domain-specific semantics from TRAINS' interpreter and no other source of lexical semantic information, only 51% of test pronouns are correctly resolved.

olving Inter-sentential Coreference & Event Reference in I

rpus for [Navarretta 2004] contains both texts (computer manuals, novels,) and previously transcribed dialogues (doctor-patient, everyday ersation).

ploys a salience-ordering on NP-evoked entities that reflects focus-marking nish and common use of parallelism.

e Eckert & Strube and Byron, keeps potential referents for event and erential anaphora separate and uses different search strategies for nstrative and 3^{rd} person pronouns. But strategies also differ for weak and g pronouns.

cause problematic cases are manually excluded and intra-sentential pronouns solved manually, results are not really comparable with those of Byron and rt & Strube.

⇒ Corpora for Discourse Anaphora

A *Gold Standard*, indicating correct labels, is needed for all evaluation. (Labels can be PoS-tags, syntactic units, word senses, predicate-argument sets, anaphor-antecedent pairs, coreference sets, . . .)

This Gold Standard is also often used in training corpus-based approaches to *coreference resolution* and *anaphora resolution*.

A Gold Standard usually produced manually, usually at significant cost, so it worth making them *reusable*.

main forms of Gold Standard annotation:

n-line annotation, where labels (and possibly associated features) are inserted nto the source text;

tand-off annotation, where each set of labels (and associated features) exists n a separate file, keyed to an index of the source text – e.g.

- character offset from start of file
- word token offset from start of file
- hierarchical offset from start of file, e.g., (<P>, <S>, token)

ntly accepted that *stand-off annotation* is more generally reusable.

al annotation systems available for creating *stand-off annotation*: MMax2 pean Media Lab, Heidelberg), WordFreak (Institute for Research in itive Science, University of Pennsylvania), NITE

Annotation: Coreference Chains

Purpose: To identify all expressions (*markables*) that denote/describe the same entity.

Called *Identity of reference* in the MUC co-reference task definition. However, expressions are not necessarily those standardly taken to refer – e.g.

Expression: "union leaders"
"union" marked and indexed as denoting entity **a**
"union leaders" marked and indexed as denoting entity **b**
$\rightarrow$ ((union)$_{\mathbf{a}}$ leaders)$_{\mathbf{b}}$

equence of *markables* that denote/describe the same thing is called a *erence chain*.

. . ( )$_{\mathbf{a1}}$ . . .
. . . ( )$_{\mathbf{a2}}$ . . .
. . . ( )$_{\mathbf{a3}}$ . . .

e case of coreference resolution, the Gold Standard is manually annotated coreference chains of *markables*.

e guidelines for annotating markables:

The only *markables* are nouns, noun phrases (NPs) – including Named Entities – and pronouns.

A substring of a Named Entity is **not** markable, hence can't participate in a coreference chain:

Equitable of Iowa has its company headquarters in Ames, Iowa.

In a report issued January 5, 1995, the program manager said that there would be no new funds this year.

- A prenominal modifier is a markable only if it is coreferential either with a Named Entity or with the syntactic head of an NP.

  He was accused of money laundering and (drug)$_k$ trafficking. However, the trade in (drugs)$_k$ . . .

- The noun that heads an NP is not a separate markable. It is markable only as part of the entire NP.

  (The rate)$_{e6}$, which was (6 percent)$_{e6}$, was higher than that offered by any other bank.

"That" is not separately markable as being corefential with "the rate", as it heads the NP "that offered by any other bank".

MUC Coreference Annotation

inal Text

"This doesn't surprise me at all," said Trish Neusch, who farms land near
he plant 17 miles northeast of Amarillo and 7 miles from the airport .
"We are out there and see these planes all the time."

ference annotation

```
s doesn't surprise <COREF ID="MARKABLE_33">me</COREF> at all," said
EF ID="MARKABLE_32" TYPE="IDENT" REF="MARKABLE_33">Trish Neusch</COREF>,
farms <COREF ID="MARKABLE_40" MIN="land">land near
EF ID="MARKABLE_34" TYPE="IDENT" REF="MARKABLE_20">the plant</COREF></COREF>
les northeast of
EF ID="MARKABLE_35" TYPE="IDENT" REF="MARKABLE_36">Amarillo</COREF> and
es from <COREF ID="MARKABLE_37" TYPE="IDENT" REF="38">the airport</COREF>.
re out
```



```
<COREF ID="MARKABLE_39" TYPE="IDENT" REF="MARKABLE_40"> there</COREF> and see
<COREF ID="MARKABLE_41" TYPE="IDENT" REF="MARKABLE_42" MIN="planes">these planes<
all the time."
</p>
```


ss Exercise: Annotating Markables and Coreference

1:

Eastern Air)$_{a1}$ Proposes (Date for Talks on ((Pay)$_{c1}$-Cut)$_{d1}$ Plan)$_{b1}$

Eastern Airlines)$_{a2}$ executives notified (union)$_{e1}$ leaders that the carrier
vishes to discuss selective ((wage)$_{c2}$ reductions)$_{d2}$ on (Feb. 3)$_{b2}$.
(Union)$_{e2}$ representatives who could be reached)$_{f1}$ said (they)$_{f2}$ hadn't
lecided whether (they)$_{f3}$ would respond. By proposing (a meeting
late)$_{b3}$, (Eastern)$_{a3}$ moved one step closer toward reopening current
igh-cost contract agreements with ((its)$_{a4}$ unions)$_{e3}$.

Text 2:

(Ocean Drilling & Exploration Co.)$_{a1}$ will sell ((its)$_{a2}$ contract-drilling business)$_{b1}$, and took a $50.9 million loss from discontinued operations in (the third quarter)$_{c1}$ because of the planned sale . (The New Orleans oil and gas exploration and diving operations company)$_{a3}$ added that (it)$_{a4}$ doesn't expect any further adverse financial impact from the restructuring. . . . In (the third quarter)$_{c2}$, (the company , which is 61%-owned by Murphy Oil Corporation of Arkansas)$_{a5}$, had (a net loss of ($46.9 million)$_{d1}$, or (91 cents a share)$_{d2}$)$_{e1}$ . . . . It has long been rumored that (Ocean Drilling)$_{a6}$ would sell (the unit)$_{b1}$ to concentrate on (its)$_{a7}$ core oil and gas business.

Penn Discourse TreeBank (PDTB)

ttp://www.cis.upenn.edu/˜pdtb

Stand-off annotation of discourse connectives and their arguments in the Penn WSJ corpus (annotation over raw text)

- Done with WordFreak
- Identifying connectives and their arguments
- Characterising the sense(s) of each connective and its arguments.

Exposes a clearly defined and relatively easily identifiable level of discourse tructure whose usefulness is not limited to one particular theory of discourse.

ntegrated with the Penn TreeBank (PTB) and PropBank

Aimed at supporting linguistic discovery and the development of robust lgorithms for discourse processing.

- Annotation of four types of connectives and their arguments
  - Explicit connectives (∼20k tokens): coordinating and subordinating conjunctions, subordinators, discourse adverbials
  - Implicit connectives (∼10k tokens) between sentences in the same paragraph with no explicit connective between them.
  - **NOT** pragmatic discourse markers such as "now", "well", "so".
- Annotation of the "senses" of connectives. (**N.B.** Since all are binary, one can annotate their arguments *without* having to annotate sense.)
- Annotation done by 2–4 annotators
- Expected date of completion: November 2005
- Extensive guide to annotation: http://www.cis.upenn.edu/˜pdtb/manual/pdtb-tutorial.pdf

Annotatable arguments in PDTB

ctically, an arg must be a clause, sentence, sentence sequence, nalisation, discourse deictic (*this*, *that*), VP (in coordinated VPs) – i.e., thing interpretable as an *abstract object*.

a. Even though critical, [it was just the kind of attention they were seeking.] **So** [they fired back at the Goldman Sachs objections in their own economics letter, "The BMC Report."]

b. Buyers can look forward to double-digit annual returns if [they are right]. But they will have disappointing returns or even losses if [interest rates rise] **instead**.

c. But some investors have used such filings [to boost the value of their stock holdings], which – without buying more stock – [they] **then** [sold].

Inter-annotator Reliability: Argument Annotation

- 10 explicit connectives – 2717 tokens
- Independent assessment of agreement for each argument (ARG1 and ARG2).
- Exact match: 90.2% overall (92.4% on subordinating conjunctions, 71.8% on adverbials)
- Partial overlap: 94.5% overall

When argument annotations overlapped but didn't match exactly, it was due to differential inclusion of:

- a clause's governing verb
- a dependent clause at the periphery of an argument
- a parenthetical in the middle of an argument – e.g.

(7) Bankers said [warrants for Hong Kong stocks are attractive] *because* $[_{1,2}$ they give foreign investors$]_1$, wary of volatility in the colony's stock market, $[_1$an opportunity to buy shares without taking too great a risk.$]_{1,2}$

nitial Experiments on PDTB [Prasad et al 2004]

To substantiate the theoretical claim of D-LTAG that some/many/all? discourse adverbials are anaphoric, having a different patterns of arguments han structural connectives (coordinate and subordinate conjunctions and ubordinators).

To examine the order in which arguments commonly appear with different tructural connectives.

Discourse Adverbials: Arg1 Location

Arg1 appears in:

- the same sentence as the adverbial (SS): main clause, complement clause, subordinate clause, relative clause, VP conjunction.
- the (immediately) previous sentence (PS)
- the (immediately) previous multi-sentence span (PP)
- one or more non-contiguous sentences (NC).

SS	PS	PP	NC	Total
41 (17.9%)	144 (62.9%)	12 (5.2%)	33 (14.4%)	229

NC location of Arg1 is only compatible with the idea that discourse adverbials get their arguments *anaphorically* rather than syntactically.

Arg1: Distribution by Adverbial

Adv	SS	PS	PP	NC	Total
Nevertheless	9.7%	54.8%	9.7%	25.8%	31
Otherwise	11.1%	77.8%	5.6%	5.6%	18
As a result	4.8%	69.8%	7.9%	19%	63
Therefore	55%	35%	5%	5%	20
Instead	22.7%	63.9%	2.1%	11.3%	97
Total	41	144	12	33	229

ourse adverbials pattern differently from coordinate & subordinate nctions **and** from one another.

Subordinate Conjunctions: Relative Order of Arguments

Relevant for NLG: placement of connectives [Walker, Prasad & Stent 2003]

2408 tokens: *when, because, even though, although, so that*

- Arg1 < Arg2
  Mr. Lothson of Paine Webber said the company's sales pace has been picking up largely because the effect of unfavorable exchange rates has been easing.
- Arg2 < Arg1
  Because they want a truck that is "Texas tough", the commercial concludes, "Texans drive Chevy".

Relative Order of Arguments: Results

Conn	Arg1-Arg2	Arg2-Arg1	Total
When	54%	46%	1010
Because	90%	10%	915
Even though	75%	25%	103
Although	37%	63%	347
So that	100%	0%	33
total	1606 (67%)	802 (33%)	2408

ally-based annotation can reveal distinctions that may be hidden with
act RST relation annotation.

n the RST corpus, *even though* and *although* occur in *concessive* relations.

Williams & Reiter (2003) found that 77% of *concessives* in the corpus were in
Arg2-Arg1 order and took it as the preferred pattern for NLG.

This ignores the very different patterning of *even though* and *although*.

Summary

- If a form has multiple functions (DN vs. anaphoric, coreference vs. event reference), makes sense for algorithm to try to distinguish them before or during further processing, since the different functions are likely to have different properties.
- Reusable corpora are expensive, but still worth the time and funds to create them.
- Since the more data, the better, both for evaluation and training, it is worth spending some effort thinking about how to reduce the manual effort needed to create Gold Standard corpora.