LOT Winter School 2005

Discourse Anaphora: Theories, Data and Applications

Bonnie Webber
School of Informatics
University of Edinburgh
January 2005

Day 1: Overview, Definitions and Descriptions

Outline

Day 1: Overview, Definitions and Descriptions

- Overview of course
- Definitions of major concepts
- Range of discourse anaphora
- Constraints and preferences in anaphor resolution
- Distribution of discourse anaphora in text
- Discussion
- Summary

Day 2: Heuristic Approaches to Anaphor Resolution

- Baseline algorithms for resolving coreferential pronouns
- Heuristic approaches to resolving coreferential pronouns: Centering
- Interlude: Performance evaluation
- Discussion
- Summary

Day 3: Heuristic Approaches to Anaphor Resolution (continued)

- Heuristic algorithms for identifying and resolving definite NP anaphors
- Heuristic algorithms for identifying and resolving event reference

Day 3: Corpora for Discourse Anaphora

- Recording Gold Standard annotation – in-line vs. stand-off; markables, features, chains
- Annotation tools – MMax2, WordFreak
- Annotation schemata – GNOME, Penn Discourse TreeBank
- Annotated corpora – GNOME, PDTB
- Automated annotation and “noisy” gold standards
- Discussion
- Summary

Day 4: Corpus-based Approaches to Anaphor Resolution

- Machine-learning for Coreference
- Machine-learning for Event Reference
- Machine-learning for Comparative Anaphora
- Discussion of Features used in ML approaches
- Summary

Day 5: Applications and Speculations

- Coreference Resolution and Information Retrieval
- Coreference Resolution and Information Extraction
- Coreference Resolution and Question Answering
- Exploiting anaphoric relations other than coreference (e.g. Comparative Anaphora and Ontology Construction)
- Discussion
- Summary

$\Rightarrow$ Definitions: Coreference/Anaphora

Coreference: Two expressions α_1 and α_2 *corefer* if and only if Referent(α_1) = Referent(α_2).

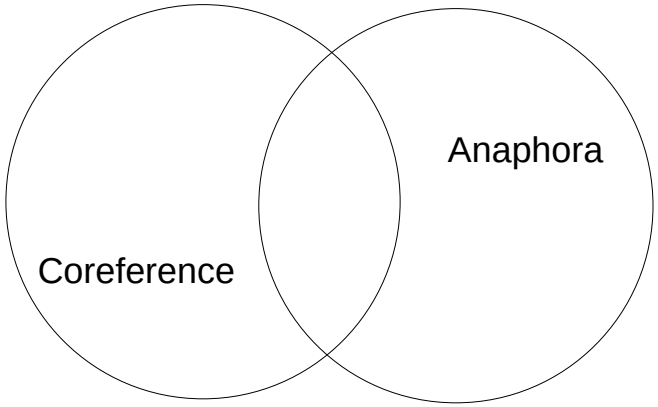
The expressions can be in the same text or different texts, in the same language or different languages.

Anaphora: An expression α_1 is in an *anaphoric relation* with expression α_2 if and only if the interpretation of α_1 depends on α_2.

As so defined, the relationship holds *within* a text.

[van Deemter & Kibble 2001] show how the MUC coreference task obscures the difference between the two.

Relationship between Anaphora and Coreference



Some expressions are both coreferential and anaphoric.

(1) A bus had to divert to the local hospital when one of the passengers had a heart attack. *It* got to *the hospital* in time and *the man*'s life was saved.

Some expressions are coreferential but not anaphoric.

(2) *Alberto R. Gonzales*, the White House counsel, intervened directly with Justice Department lawyers to obtain a legal ruling on the extent of the president's authority to permit extreme interrogation practices in the name of national security.

Mr. Gonzales's role in seeking a legal opinion on the definition of torture . . .

Some expressions are anaphoric but not coreferential.

- An anaphoric expression may not refer.

  (3) Every TV network reported *its* profits.

- Or its interpretation may depend in a way other than coreference.

  (4) A bus had to divert to the local hospital when one of *the passengers* had a heart attack. . . .

Some expressions are neither.

Definition: Predication

[van Deemter & Kibble 2001] also show how the MUC coreference task obscures the difference between *predication* and *coreference*.

Predication is a relation at the syntactic/semantic interface. A predicative NP is interpreted as saying that its description holds of some syntactically related NP.

(5) Alberto R. Gonzales, *the White House counsel*, intervened directly with Justice Department lawyers ...

(6) In Texas, as *Governor Bush's legal advisor*, Mr. Gonzales wrote briefs on condemned prisoners' appeals for clemency that were notoriously sloppy.

(7) Mr. Gonzales was *Governor Bush's legal advisor*.

Since predicative NPs are non-referential, they don't corefer.

While both coreference and predication are distinct from anaphora, they can nevertheless be useful in resolving anaphora, so shouldn't be ignored.

Definition: Presupposition

> "There is more literature on presupposition than an almost any other topic in pragmatics (except perhaps speech acts) . . . ."
> [Stephen Levinson, **Pragmatics**. Cambridge University Press, 1983, p. 167]

Strawson (1952) used the term **presupposition** to refer to the relation between

(8) The King of France is wise.

and

(9) There is a present king of France.

as a constraint on the use of referring expressions.

So informally, **presupposition** involves an expression *presupposing* the truth of a proposition if that proposition must be true for the expression to make sense.

Other linguistic expressions enumerated by Kartunnen and recorded by Levinson (pp. 181-184) that “trigger” presuppositions include:

- Factive verbs: “realise”, “regret”

  (10) Mary was sorry that Fred left.

  presupposes: Fred left

- Change of state verbs: “stop”, “start”

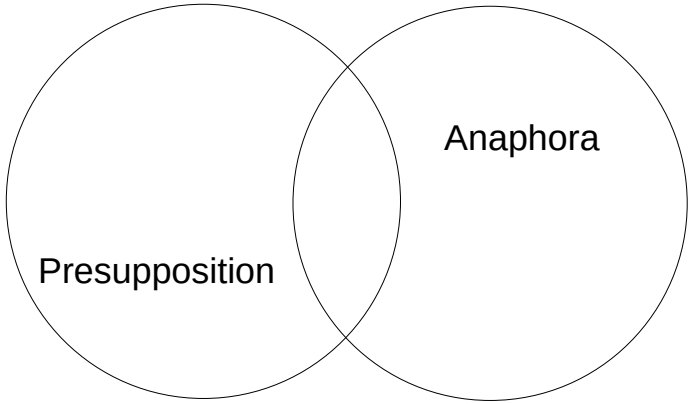
  (11) Fred closed the door at 5pm.

  presupposes: The door was open (just) before 5pm.

- Iteratives: “again”
- Temporal clauses linked by “when”, “before”, “after”
- Comparatives (including “other”)
- Wh-words: “why”, “when”, “how”

Issue: Rob van der Sandt and Bar Geurts have argued that *presupposition* should be considered *anaphoric* based on the fact *inter alia* that both may have to be *accommodated* into the discourse context.

Relationship between Presupposition and Anaphora



Although some expressions are both anaphoric and presuppositional

(12) April 2001 unemployment figures for the UK were published last week. *The figures* showed unemployment to be significantly lower than expected.
presupposes: There is a (unique) set of figures.
anaphoric antecedent: April 2001 unemployment figures for the UK

[Bosch 2001] points out that there are expressions that carry presuppositions without being anaphoric

(13) *A bachelor whom my cousin was dating* was doing a PhD in Informatics.
presupposes: The person my cousin was dating was male and an adult.

(14) The *April 2001 unemployment figures for the UK* were significantly lower than expected.

presupposes: There is a unique set of figures that is the April 2001 unemployment figures for the UK.

And expressions that are anaphoric without carrying presuppositions – e.g., *epithet NPs*

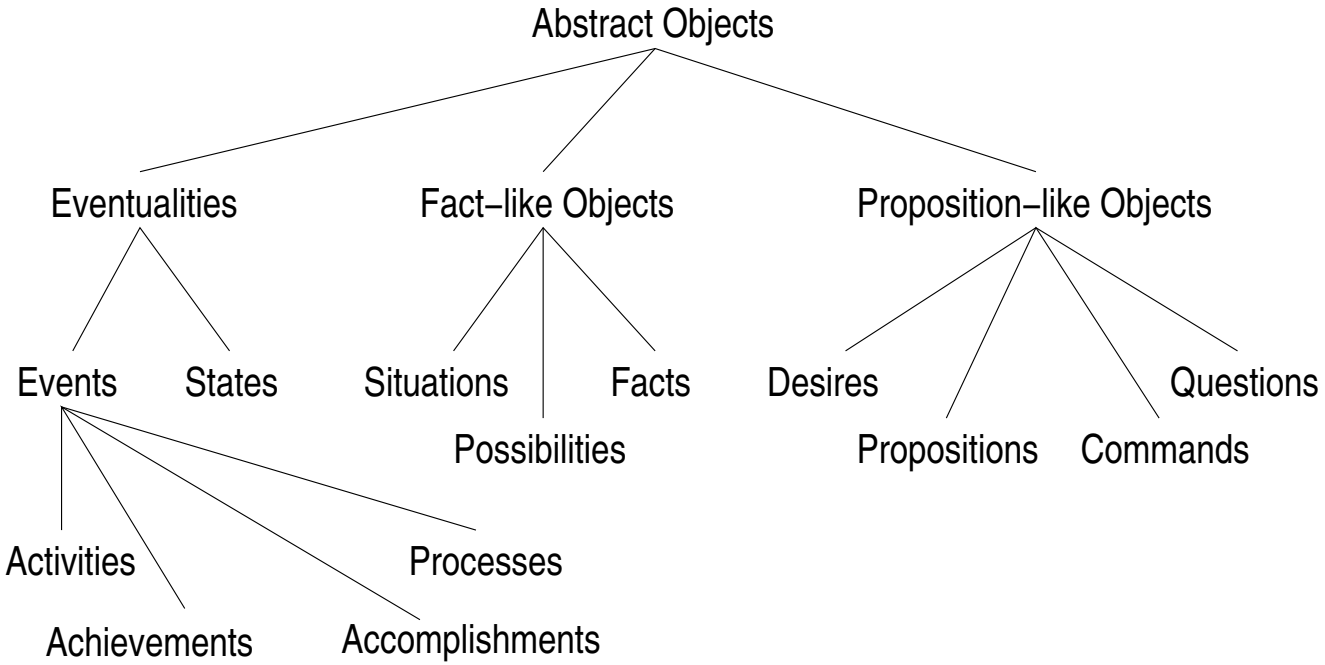
(15) When I arrived at Jones' office, *the old grocer* greeted me with the bill.

N.B. With prosodic marking – "the old GROcer" – the expression is no longer *coreferential* with Jones and *non-presuppositional*, but rather *non-coreferential* with Jones and *presuppositional*. Contra [Bosch 2001], it is *anaphoric* in both cases.

So we will not consider *presupposition* in these lectures on *discourse anaphora*.

Definition: Abstract Object

Every clause is interpreted as an *Abstract Object* (AO) of a contextually appropriate type.



[Asher, Nicholas. **Reference to Abstract Objects in Discourse**. Kluwer, 1993]

Common to call reference to AOs *event reference*.

⇒ Range of discourse anaphora

Important to stress from the start that the mapping between form and function in most/all languages is not 1:1.

One function can be served by multiple forms – e.g. marking *contrast*

(16) *Prosodic marking*:
Mary likes bananas. JOHN likes ORANGES.

(17) *Lexical marking*:
Mary likes bananas. On the other hand, john likes oranges.

One form can serve multiple functions – e.g. *it-clefts*

(18) *Marking new, often contrastive information*:
Did John wash the dishes? No, it was Mary who washed the dishes.

(19) *Presenting new information without implying that the hearer doesn't already know it*
It was just about 50 years that Henry Ford gave us the weekend.

Pronominal Anaphora

- **form**: pronouns – 3*rd* person, demonstrative, zero.
- **function**: coreference, binding, "type" reference, "event" reference

Coreference, 3*rd* person pronoun

(20) Evolution has laid down multiple systems for protecting us against hunger; *it* has not given us much to protect against plenty.

Binding, 3*rd* person pronoun

(21) Every TV network reported *its* profits.

Type reference, 3*rd* person pronoun

(22) The lizard's tail fell off. But it grew back within a month.

Event reference, 3rd person pronoun

(23) Sindona, the onetime Vatican financial adviser with reported links to the Mafia, died on March 22, 1986, at age 65, reportedly after drinking cyanide-laced coffee in an Italian prison. It happened four days after he was sentenced to life in prison for ordering a 1979 murder. [Penn TreeBank WSJ_1266]

Event reference, demonstrative pronoun

(24) Segal, however, had his own problems with women: he had been trying to keep his mariage of seven years from falling apart; when *that* became impossible, . . .

(25) Using microscopes and lasers and ultrasound, he removes tumors that are intertwined with children's brain stems and spinal cords. There is only the most minute visual difference between the tumors and normal tissue. Operations can last 12 hours or more. The tiniest slip can kill, paralyze or leave a child mentally retarded. *This* is the easy part of his job. [New York Times, 8/11/90]

Associative reference, demonstrative pronoun

(26) First, multiply your height in metres by itself, then invert *that* and multiply it by your weight in kilograms.

Headed NP anaphora

- **form**: headed definite, indefinite and demonstrative NPs.
- **function**: coreference, “associative” reference, event reference

Coreference, headed definite NP

(27) Meridia has been the subject of a $50 million marketing campaign in the US, aimed at doctors and their weight-conscious middle-aged female patients. *The drug*, generically known as sibutramine, ...

Associative reference, headed definite NP

(28) Meridia has been the subject of a $50 million marketing campaign in the US, aimed at doctors and their weight-conscious middle-aged female patients. *The ads* appear 20 times a day on TV.

Associative reference, headed indefinite NP

(29) A bus had to divert to the local hospital when *a passenger* had a heart attack.

Coreference, headed demonstrative NP

(30) The proposed system has many variables, including what attributes are considered in a match and how matching is done on those attributes.

Non-anaphoric headed demonstrative NP

(31) The Penn Discourse TreeBank annotates a subset of the Penn WSJ corpus with *those discourse relations that can be taken to hold between pairs of discourse spans identified in the corpus*.

Comparative NPs

- **form**: NPs modified by “other”, “such”, comparative adjectives.
- **function**: comparative anaphora, comparative associative anaphora

Comparative anaphora, “other” NP

(32) Will Quinlan had not inherited a damaged retinoblastoma suppressor gene and, therefore, faced no more risk than *other children*.

Comparative anaphora, comparative adjective NP

(33) Corgis are very high-strung and nervous. *Larger dogs* have a more placid disposition.

Comparative associative anaphora, “other” NP

(34) Sue lifted the receiver as Tom darted to *the other phone*

.

Discourse adverbials

- **form**: discourse adverbs and prepositional phrases
- **function**: additional relation with salient *Abstract Object* (AO)

(35) If the light is red, stop. *Otherwise* you'll get a traffic ticket.
relation: if not (negative conditional)
salient AO: you stop

(36) If the light is red, stop. *Otherwise* just go straight on.
relation: if not (negative conditional)
salient AO: the light is red

(37) Going to the beach sounded boring. *Instead*, John went to the zoo.
relation: alternative to
salient AO: going to the beach

(38) Compounding the trouble to Japanese investors, mortgage securities pay interest monthly, since most mortgages require homeowners to make monthly payments. But Japanese institutional investors are used to quarterly or semiannual payments on their investments, so the monthly cash flow posed administrative problems. *As a result*, Japanese investors steered clear of the mortgage securities.

relation: result

salient AO: the monthly cash flow posing administrative problems

Other Discourse Anaphors

Temporal anaphora

(39) The FDA *turned down* an application by a subsidiary of AHP to market a more powerful form of fenfluramine, Redux. They *were concerned* by reports that a hundred people in Europe *had discovered* a rare, incurable and fatal respiratory disease after taking it. AHP and their allies *lobbied* for a second hearing.

VP anaphora

(40) a. A Middle East peace agreement needs to be negotiated. A peace agreement for Sri Lanka *does* ϕ too.

b. A Middle East peace agreement needs to be negotiated. Colin Powell offered to *do it*.

c. As an imperial statute, the British North Americ Act could be amended only by the British Parliament, which *did so* on several occasions.

⇒ Hard Constraints on Anaphor Resolution

Constraint 1. An entity of the right sort must be available for the anaphor.

Hearer Status: Whether the speaker believes entity$_i$ to be known (*hearer old*) or unknown (*hearer new*) to the hearer at the time of reference.

Indefinite NPs introduce *hearer new* entities. (May still be anaphoric.)

(41) I bought *a book* at the bookstore yesterday.

(42) A bus had to divert to the local hospital when *a passenger* had a heart attack.

Un-elaborated proper names introduce entities as *hearer old*.

(43) It was nice to hear *Alberto Gonzales* tell the Senate Judiciary Committee that he doesn't approve of torture.

Elaborated proper names introduce entities as *hearer new*.

(44) Alberto R. Gonzales, *the White House counsel*, intervened directly with Justice Department lawyers . . .

Discourse Status: Whether entity$_i$ has (*discourse old*) or hasn't (*discourse new*) been introduced into the discourse at the time of reference.

The interpretation of any anaphor depends on a *discourse old* entity. Pronouns can only be used for a subset of *discourse old* entities, which Gundel calls *in focus*.

There may also be a preference on pronoun antecedents, depending on whether they were introduced as *hearer old* or *hearer new*. (cf. discussion of [Strube 98; Strube & Hahn 99] in Lecture 2)

Constraint 2. If a pronominal anaphor is coreferential, the entity it refers to must satisfy gender/number agreement. (Constraint on the referent, not on the antecedent, though usually checked on the antecedent.)

(45) a. I saw John's new Acura. *He* bought *it* at the Acura dealership. *They* were having their year-end sale. (**coreference**)
he = ?
it = ?
they = ?

b. John buys a new Acura every year. Last year it_1 was red. This year, it_2 is silver. (**type reference**)
it_1 = ?
it_2 = ?

c. John buys a new Acura every year. *It*/*They* is/are (always) red. (**type reference**)
it = ?
they = ?

Constraint 3. Also if an anaphor is co-referential, it must satisfy *intra-sentential binding constraints*:

(46) John bought him$_i$ a car. (him$_i$ $\neq$ John)

He$_i$ bought John a car. (he$_i$ $\neq$ John)

He$_i$ said that he$_j$ bought John a car.

(He$_i$ $\neq$ John; he$_j$ $\neq$ John)

Preferences in Anaphor Resolution

Approaches to co-reference resolution vary in what factors they appeal to when choosing between alternative possibilites. (No approach uses them all.)

1. Predicate-argument preferences

(47) a. John crashed his car into a tree. *It* was travelling at 80mph.
it = ?

b. John crashed his car into a tree. He had planted *it* only 5 years before.
it = ?

c. John crashed his car into a tree. *It* happened when it was really foggy.
it = ?

2. Preferences from general world knowledge

(48) a. John parked his car in the garage. *It* was incredibly messy, with styrofoam cups strewn around everywhere.
it = ?

b. John parked his car in the garage. *It* was incredibly messy, with old car parts strewn around everywhere.
it = ?

3. Recency preferences

(49) a. At the zoo, John visited the armadillos and Sam visited the otters. *They* were in the water, eating fish.
they = ?

b. At the zoo, John visited the armadillos and Sam visited the otters. *He* wanted to see animals he wasn't familiar with.
he = ?

4. Preferences from grammatical role: subj > obj > indirect obj > other roles

(50) John bought a book from Fred. *He* needed it for class.

he = ?

5. Preferences based on repetition: An entity that has already been made salient can be taken to continue salient through repeated mention.

(51) John needed a car to get to this new job. *He*$_1$ decided *he*$_2$ wanted something sporty. Bill accompanied *him*$_3$ to the Acura dealership, where *he*$_4$ bought an Integra.

he$_1$ = he$_2$ = him$_3$ = John

he$_4$ = ?

N.B. Often recency, role and repetition are lumped together under a notion of *salience*, or *in focus*.

6. Preference for parallelism: If clauses can be viewed as parallel by resolving a referring form in a particular way, that may be preferred.

(52) Mary likes Sue's brother, and Harry likes *her* sister.
her = ?

7. Preferences based on causality: In English, the subject of most verbs is the implicit source of their *causality*. For a small class, the implicit source of causality fills a different grammatical role. This can affect resolution preferences.

(53) a. John telephoned Bill. *He* apologised for losing the assignment.
he = ?

b. John admired Bill. *He* could juggle six bananas.
he = ?

N.B. This is not the same as predicate-argument preferences.

8. Preferences based on Point of View:

(54) "The trouble with Herbie," I remember complaining to Tim that night, "is that *he*$_1$ gives the impression in the House that *he*$_2$'s not on top of *his*$_3$ brief. I don't know if *he*$_4$ is, and I don't suppose you do either, but the impression is what counts."

"*He*$_5$'s still feeling *his* way," said Tim. *He*$_6$ jumped off my desk and did an imitation of Morrison . . . .

he$_1$ = he$_2$ = his$_3$ = he$_4$ = he$_5$ = Herbie
he$_6$ = Tim

[Robert Barnard, *A Scandal in Belgravia*, 1990.]

9. Situational preferences

Prefer an antecedent "present" in the situation.

(55) When Mary drove up to the house, Fred came out and opened *the door* for her.

(56) When Mary drove up to the house, Fred was inside, so she knocked on *door*.

10. Predicate-Argument Preferences (event anaphora)

Prefer an antecedent whose AO interpretation is appropriate to the predication.

(57) Hey, management has reconsidered its position. They've promoted Fred to second vice president.

a. *That*'s a lie. (*assertion*)

b. *That*'s false. (*proposition*)

c. *That*'s a funny way to describe the situation. (*description*)

d. When did *that* happen? (*event*)

(58) Over the course of a year, John grew over 9 inches in height. So he was now 6 foot 3 inches tall. After *that*, he took up basketball.

(Need an *event* as referent for *that*.)

Constraints/Preferences on Resolving *Instead*

Background: “Instead” comes in two forms:

1. with an “of” PP modifier

   (59) John ate an apple *instead of* a pear.

   (60) John spent the afternoon at the zoo *instead of* at the museum.

   Here the interpretation of *instead* can be derived compositionally from the syntactic-semantic interface:

   - Arg1 = chosen alternative w.r.t. predication, deriving from the modified NP (“an apple”, “at the zoo”)
   - Arg2 = salient but unchosen alternative w.r.t. that predication, derived from the “of” PP (“a pear”, “at the museum”).

3. as a discourse adverbial

   (61) *Chrysler officials resisted* cutting output. *Instead*, they slapped $1000 cash rebates on vehicles.

   Here the interpretation of *instead* requires both anaphor resolution and compositional derivation at the syntactic-semantic interface:

   - Arg1 = chosen alternative w.r.t. predication, deriving from the matrix clause (“they slapped $1000 cash rebates on vehicles”)
   - Arg2 provided through *anaphor resolution*.

Constraints/Preferences on Resolving *Instead*

Constraint: *Salient* from the previous discourse must be something interpretable as the unchosen alternative.

(62) a. John ate an apple. #*Instead* he wanted to eat a pear.

b. John wanted to eat a pear. *Instead* he ate an apple.

Preference: The salient alternative should be an appropriate alternative to Arg1.

(63) a. But it is wrong to think that any one design will dominate in future, says Mr Nakaizumi.

b1. *Instead*, different types of users will want different styles, depending on whether they maily use the devices for voice calls, text messaging, music or games.

b2. *Instead*, one should allow for a variety of designs.

What suggests “unchosen alternatives”?

- **Presence of negation**

  (64) John *didn’t* sleep. Instead, he wrote code. (**Verbal neg**)

  (65) *No one* slept. Instead, everyone wrote code. (**Subj neg**)

  (66) John ate *none of* his spinach. Instead, he fed it to his frog. (**Obj neg**)

- **Presence of a monotone-decreasing quantifier**

  (67) *Few* students like to do homework. Instead, they would rather party.

  (68) Students *seldom* sleep in class. Instead, they take notes assiduously.

- **Presence of a modal auxiliary** (expressing need, ability or necessity: not future)

  (69) You [*should*, *must*, *could*] exercise more. Instead you sit like a couch potato.

  (70) You [*shall*, *will*] exercise more. #Instead you sit like a couch potato.

- **Embedding in a higher clause expressing a propositional attitude**

  (71) *Paine Webber considered* recommending specific stocks. *Instead*, it just urged its clients to stay in the market.

  (72) *Chrysler officials resisted* cutting output. *Instead*, they slapped $1000 cash rebates on vehicles.

  N.B. This can't be a *factive predicate*, since factives presuppose the truth of their embedded clause:

  (73) John regrets that Fred eats meat. #Instead Fred likes tofu.

- **Presence of verbs whose own semantics suggest unchosen alternatives**

  (74) John *doubted* Mary's resolve. *Instead*, he thought she would just resign.

  (75) NBC is contemplating *getting out of* the cartoon business. *Instead*, it may "counter-program" with shows for an audience that is virtually ignored in that time period: adults.

  (76) Investors have *lost* their enthusiasm for the stock market. *Instead*, they are buying government bonds.

  (77) But respectability still *eludes* Italy's politics. *Instead*, it has the phenomenon of Mr. Belusconi.

$\Rightarrow$ Would be useful to know what verbs have this property (or these properties).

- **Presence of other lexico-syntactic elements that trigger unchosen alternatives**

  (78) The tension was evident on Wednesday evening during Mr. Nixon's final banquet toast, normally an opportunity for reciting platitudes about eternal friendship. *Instead*, Mr. Nixon reminded his host, Chinese President Yang Shangkun, that Americans haven't forgiven China's leaders for the military assault of June 3-4 that killed hundreds, and perhaps thousands, of demonstrators.

  "Normally" and "opportunity" can both trigger alternatives:

  (79) Normally, we eat pasta on Tuesday. *Instead*, tonight we're having fish.

  (80) John had the opportunity to buy a used car. *Instead*, he bought a scooter.

What else?

Determining the relative importance of salience factors

Languages may differ in how they weight factors relevant to *salience*. If salience is required for pronoun use, [Prasad & Strube 2000] offer an interesting methodology for assessing the strength of factors in a given language.

1. Collect a large set of utterance pairs U_{i-1} and U_i that satisfy the following three conditions:
   (a) U_i realizes two and only two of the entities from U_{i-1}.
   (b) In U_i, only one of the NPs realising these entities is pronominalised.
   (c) The pronoun in U_i is ambiguous (for gender and number) between the two entities in U_{i-1}.
2. Categorize the linguistic properties of the NPs (e.g., grammatical function, word order, hearer/discourse status).
3. Compare subgroupings by pairs of factors (e.g., comparing salience of subjects and direct objects, direct objects and indirect objects, etc.)

P&S's analysis of salience factors in Hindi

Frequencies for Relative Salience of Grammatical Functions

Ranking	Number	Total	Frequency (%)
Subject > Direct Object	144	149	96
Subject > Indirect Object	50	57	87
Subject > PP Object	128	128	100
Direct Object > Indirect/PP Object	22	22	100
Subject/Object > Adjunct	96	110	87
Possessor > Head	22	22	100
Subject > Possessor of Direct Object	50	50	50
Indirect Object > Possessor of Subject	22	22	100
Total	534	560	95

Similar analyses for word order and Information Status showed no major effects.

⇒ Distribution of Pronouns in Text

[Vicedo & Ferrández 2000] show that for English text

- Pronouns occur with different frequencies in different types of text;
- Different pronouns occur with different frequencies in the same type of text.

Text Collection	LAT	TIME	FT	MED	CACM	FR	Cranfield
Pronoun type							
he,she,they	38.59%	31.2%	26.2%	15.07%	8.59%	13.31%	6.54%
his,her,their	25.84%	35.01%	20.52%	21.46%	15.69%	20.7%	10.35%
it, its	26.92%	22.42%	46.68%	57.41%	67.61%	61.06%	79.76%
Pronouns in sentences							
with 0 pronouns	44.8%	51.37%	64.04%	77.84%	79.06%	84.92%	90.95%
with 1 pronoun	30.4%	29.46%	23.07%	15.02%	17.54%	11.64%	8.1%
with 2 pronouns	14.94%	12.26%	8.54%	4.75%	2.79%	2.57%	0.85%
with 2+ pronouns	9.56%	6.9%	4.34%	2.39%	0.6%	0.88%	0.09%

Distribution of NPs in Text

Earlier, [Fraurud 1990] showed that, for Swedish non-fictional text (brochures, newpaper articles, text books, debate books)

- Different kinds of NPs occur with different frequencies.
- Only a fraction of NPs evoke entities that anchor subsequent coreference.
- Most definite NPs were not coreferential and not anaphoric.

	Full IndefNPs	Full DefNPs	Other NPs	Row total
Initial mention (count)	115	89	126	330
(row %)	34.8	27.0	38.2	100.0
(column %)	9.4	11.9	6.6	–
Isolated mention (count)	929	365	585	1879
(row %)	49.4	19.4.	31.1	100.0
(column %)	75.9	49.0	30.7	–
Subsequent mention (count)	101	269	606	976
(row %)	10.3	27.8	62.1	100.0
(column %)	8.3	36.1	31.8	–
Other (count)	79	2	591	692
(row %)	11.0	3.2	85.4	100.0
(column %)	6.5	3.0	31.0	–
Column total	1224	745	1908	3877

“Other NPs” include NPs with possessive determiner, pronominal NPs, “totality” pronouns (*all*, *every*), NPs with an elliptical head, coordinate NPs, and a variety of other NPs.

“Other occurences” of NPs (692) include predicate nominatives.

Of 269 subsequent mention def NPs, 114 (42.4%) refer to an entity introduced by an indefinite NP $\Rightarrow$ anaphoric coreferential def NPs.

The rest, 155/269 (57.6%) were coreferential with earlier (initial mention) def NP, often identical in form $\Rightarrow$ corefential without being anaphoric.

Type of initial mention NP	Count	%	Count	%
Indefinite NP	114	42.4		
Definite NP	155	57.6		
Indentical defNP			53	19.7
Non-identical simple defNP			27	10.0
Non-identical complex defNP			35	13.0
Other defNP			40	14.9
Total	269	100.0		

Poesio & Vieira (1998): Analysis of definite NPs in the WSJ portion of PTB.

- 52% *discourse new* (Fraurud's *initial* or *isolated* mention);
- 15% *bridging* reference.

The subset of *bridging* reference that constitutes *associative* reference (e.g., bus/passenger) would be considered *initial* or *isolated* mention in Fraurud's terms. Ones with different heads would be considered *subsequent* mention.

Summary – Take-away Points

- Worth distinguishing *coreference* from *anaphora*.
- With Anaphora, as with the rest of language, form:function $\neq$ 1:1.
- There are both constraints and preferences on anaphor resolution.
- Preferences may vary from language to language. Will see that algorithms vary in what/how preferences are used.
- The frequency of different types of anaphora varies across genres.