

Back to Part II

4. Beyond coreference: genericity, bridging and other classes of information status

4.1 Genericity and token/type

Some schemes (e.g., ARRAU) require coders to mark anaphoric reference to generic entities in cases of multiple reference to kinds, such as *the homeless population* in (4.1.1).

(4.1) Your comments implied we had discovered that the principal cause of homelessness is to be found in the large numbers of mentally ill and substance-abusing people in the homeless population. [...] The study shows that nearly 40% of the homeless population is made up of women and children and that only 25% of the homeless exhibits some combination of drug, alcohol and mental problems.

Other schemes (e.g., OntoNotes) require coders not to do that, but this causes problems (see, e.g., Nedoluzhko's statement). In ARRAU, generic reference is also marked in cases like (4.1.2) in which it has often been proposed that the 'antecedent' is in the scope of a generic operator; and (4.2.1.3), where the 'generic' element is a nominal modifier.

(4.1.2) A workshop needs to be planned carefully. Otherwise it may turn in a disaster.

(4.1.3) Extra virgin oil consumption in Italy is falling. Extra virgin oil has become too expensive.

In order to reduce work, nominal modifiers are however only supposed to be treated as markable when subsequently referred to –but this creates serious problems for anaphoric resolvers.

Challenges for the guidelines which types of generic anaphors should be marked?

A more general problem raised by generic expressions is the lack of a universally accepted theory of genericity.

Challenge for linguistic theory Is there by now an established theory of genericity?

From the point of view of information structure, more specific questions raised by generics include:

* Are discourse-new generic expressions on a par with other discourse-new expressions or are they rather *hearer-old* because they refer to a general concept (a 'kind') readily available in the lexicon?

* Are discourse-old generic expressions a case of *coreference* or not?

A specific issue with "Discourse-old generics" was raised in the Switchboard annotation. They kept on coming up as a difficult decision for annotators, and they were clearly not well covered in the annotation scheme. A solution could be to place a few cases on a continuum of increasing degree of referentiality:

1. **discourse-new generics** as in *dogs are such nice animals* where no dog(s) was mentioned earlier so there is no antecedent and the generic (type) is a discourse-new entity: $\emptyset \rightarrow$ type

2. **discourse-old generics** -- there are two different kinds, depending on antecedent
- two generic mentions, so the second mention can be conceived as anaphoric, though still on a type rather than referential level: type → type
 - a generic mention after a referential mention, as in *I loved my dog. Dogs are such nice animals.* This is a type of bridging reference with an anaphoric type mention (the generic) whose antecedent is referential (thus a token): token → type

1. anaphoric, referential (but not co-referential) mentions. This would be for example the case of "other" or set-bridging: token_i → token_j [MP: **This point needs elaboration**]

Challenge for annotation schemes Decide how to classify them. They are not new of course, they are akin to generics, but have an additional link to something previously mentioned. Would it make sense to mark them all as different kinds of generics (but the last kind)?

Challenge for annotation tools Possibly support associative link to antecedent, though it's not properly a coreferential link -- so allow for links to be labelled. (General issue for bridging)

Challenge for systems Detect that subsequent mentions of the same string are NOT coreferential. Distinguish type- from token-coreference?

In the RefLex system, the distinctions between discourse-old and discourse-new generics, coreferential items and non-coreferential, specific repetitions of the same concept are captured by assigning distinct referential information status (discourse referents) and lexical information status (lexemes/word senses concepts), as in (4.1.4)-(4.1.8).

(4.1.4) Smith was happy. The polls showed a majority for the politician[r-given, l-new]. (epithet, i.e. coreferring expression, non-generic, new lexical material)

(4.1.5) A car [r-new, l-new] was waiting in front of the hotel. I could see a woman in the car [r-given, l-given]. (non-generic items, normal coreference)

(4.1.6) A cat [r-generic, l-new] is often kept as a pet. Moreover, a cat [r-generic, l-given] is quite independent. (generic items, generic "coreference")

(4.1.7) After the holidays, John arrived in a new car [r-new, l-new], and also Harry had bought a new car [r-new, l-given]. (non-generic items, no coreference but conceptual repetition)

(4.1.8) This is a lion. A lion [r-generic, l-given] is a predator. (first entity: "token", second entity "type", no coreference but conceptual repetition)

4.2 Deixis / situational accessibility

With references that are both deictic and anaphoric, some sort of decision has to be made whether to mark anaphoric links. First and second person pronouns for instance are always readily available and can occur without a textual antecedent. Third person pronouns, by contrast, usually require a textual antecedent. In (4.2.1), therefore, a decision has to be made whether the aspect of *coreference* should be given priority or the fact that deictic pronouns usually obtain their referent from the situational (or utterance) context.

(4.2.1) I came home late and realised that I had left the lights on.

In dialogue data, we additionally encounter demonstratives (gestural deixis), which -- depending on the language -- may take the form of third person pronouns.

In PROIEL, overt properties of the NP is used to decide between deixis (ACC-SIT) and anaphora (OLD). Consider example (4.2.2) from Haug.

(4.2.2) Dad to Håvard: “Here is your sandwich.”

Håvard to dad: “I don’t want this sandwich, I want the sandwich you already ate.”

By saying (yelling) ‘this sandwich’, my son instructs me to resolve the reference by following his finger, not by scanning our joint discourse, which would yield the same result. In PROIEL, the overt demonstrative would make us go for ACC-sit, not OLD.

4.3 Delimiting information status classes

Identifying discourse-new information seems to be relatively easy. However there are various views on the issue how to single out discourse-new information which is given/old/known in some other respect. Prince (1992), Nissim et al. (2004) and Götze et al. (2007) distinguish between (*hearer-*)*new* information and *hearer-old* (*mediated-general/accessible-general*) information. This distinction, however, requires assumptions about the expected audience and may depend on the temporal distance between writing and reading because knowledge about certain individuals may fade away, see (4.3.1)

(4.3.1) Da saß Katherina Reiche nun in der ARD-Talkrunde von Günther Jauch und igelte sich in der Vergangenheit ein.

Baumann and Riester (2012) argue against using hearer-knowledge as a (toplevel) criterion for delimiting information status. Instead they propose to distinguish between definite discourse-new expressions (like proper names) (*r-unused*) and discourse-new indefinites (*r-new*).

It is unclear though whether it is ok to include (in-)definiteness (that according to some is a morpho-syntactic feature) as a defining feature for (semantic) information status categories. (Should morpho-syntactic marking be integrated as a separate annotation level?) In RefLex, e.g., a referring expression is only marked as 'new' if it is indefinite.

- Advantages: general logical correspondence between newness and (specific) indefiniteness; clear rule for annotators
- Disadvantages: conflation of annotation levels; how to deal with counterexamples?

Given (or accessible) indefinites

(4.3.2) A bird is sitting in the tree. It has just lost a feather.

(A *feather* is a part of a *bird* and thus not totally new. It would count as mediated (Nissim et al.) or ACC-inf (Götze et al.) Corresponding RefLex annotation: r-new + l-accessible)

In Götze et al. (2007), ACC-gen is assigned if "the referent [is] assumed to be inferable from assumed world knowledge". But our access to assumptions about world knowledge goes through the overt linguistic material, and I guess (4.3.3) would count as NEW, but (4.3.4) as old, even if the referent is the same. At least that is how we proceeded.

(4.3.3) a. A man called Thorbjørn Jagland gave me the Nobel peace prize today.

(4.3.4) The Secretary-General of the Council of Europe gave me the Nobel peace prize today.

On the other hand, if we want to use our IS annotations to study how the linguistic realization of noun phrases

is influenced by IS, this can lead to circularity. The GNOME and PROIEL groups had some success in annotating complex properties (such as topicality) by breaking it down to component parts that are easier to annotate. Would it be possible to apply the same technique here? We would have coreference edges labelled with the type of coreference (full, partial, bridging (with subtypes?)) and then annotation that would let us retrieve the relevant properties of the NPs (demonstrative in (4.3.2), indefinite with a relative clause in (4.3.3) and definite with a genitive dependent in (4.3.4)). Then we could try to synthesize an IS tag from these annotations.

General knowledge still remains the real bottleneck, as we know. So I really would like to take the chance to brainstorm on concrete ways to deal with this issue. Where do we get it from? At what stage of processing do we access knowledge? How much connection is needed to have a link rather than a new mention? How do we know what knowledge two or more people actually share?

Challenge for annotation schemes General knowledge enters the issue of annotation schemes in terms of what is considered to support and justify mediated and old information. The issue for bridging is rather obvious, but also there's also the issue of what's mutually shared between speaker/writer and hearer/reader. Who decides the extent of shared knowledge? How does this change also according to different domains and genres?

Challenge for annotation tools Nothing much different than other issues, unless we want to include information on where the information comes from to support a certain decision, or what this information is (this is related to the previous issue of "bridging anchors").

Challenge for systems Well, obviously knowing when additional knowledge is required, accessing such knowledge and finding/using the right portion of it.

4.4 Bridging

Bridging references cause all sorts of problems both for annotation and for interpretation, ranging from providing clear criteria for distinguishing between discourse new and mediated entities to spotting all bridging references. It is also the one hard problem on which there has been substantial work in the literature on anaphora and information structure.

4.4.1 Annotation

The schemes that have achieved reasonable results in terms of reliability / recall have done so by restricting the range of bridging references to annotate (see, e.g., the schemes used in GNOME or the Prague Dependency Treebank).

Challenges for the guidelines What is the state of the art in annotating bridging references? What is the broader range of bridging references that we can annotate reliably these days?

Challenges for the resolution systems what is the state of the art in resolving bridging references?

4.4.2 Anchors

Inter-annotator agreement in recognising a bridging anaphor can reach acceptable values provided that the task is suitably restricted ($k > .6$) [2, 1]. However, agreement on *what serves as an anchor* is far from trivial. There can be more than one to choose from for an annotator, or there can be none exactly identifiable.

Challenge for annotation schemes Major challenge, especially regarding instructions for annotators, as more than one anchor might be identifiable. Would it make sense to mark more than one? Should there be specific

rules and/or tests to decide which anchor should be selected (e.g. choose a noun over a verb, if equally likely, or choose closest always, for example). The usual issue of what counts as bridging obviously still holds, linking the anaphor just adds to the task's complexity.

Challenge for annotation tools Possibly supporting multiple links -- from one anaphor to multiple antecedents? Is it possible/useful to annotate more than one anchor in support of a bridging anaphor? In any case links should be labelled.

Challenge for systems Well, it's obviously hard, at least as hard as full NP anaphora resolution. Systems should have access to the general background knowledge which makes the link natural and possible for humans, and be able to select the right portion of it. The kind and structure of knowledge also depends on how bridging links are defined in the annotation scheme (i.e. in the theoretical model.) Strategies in feature selection should differ according to what is allowed/preferred to feature as anchor (restrictions on grammatical type might be enforced, for instance -- such as "always prefer an NP antecedent over a VP where both valid.")

4.4.3 'Situational' bridging and semantic role labelling / frames

In so-called situational or scenario-based bridging (e.g., *IBM --- the vice-president*) a bridging reference refers to frame roles with a clear connection to semantic role labelling. Yet --- with a few exceptions such as Anette Frank's recent work -- the two communities operate independently of each other. It might be good to collaborate. Semantic role labelling can also shed insights on annotation of bridging when there are competing possible antecedents where they fulfil a different role of the bridging anaphor.

[I think that the connection between bridging, semantic roles (and Pustejovsky's qualia) was quite vivid in the old days (e.g. Bos et al. 1995) but has since then been neglected.]

4.4.4 Resolving bridging references

The resolution of anaphora beyond pure coreference is in a bad state. Even when simplifying the task, the performance on tasks such as bridging resolution is rather low, both regarding bridging recognition and antecedent selection. In contrast to coreference resolution, currently it is still necessary to use gold information from several levels. This needs to be overcome.

Markert, Hou and Strube have made some progress on both fronts. Particular problems that they found are

1. the prevailing pipeline model for recognition and antecedent selection might not be appropriate
2. context-dependent bridging where the world knowledge is not available in context-independent knowledge-bases
3. Long-distance bridging where fixed window-sizes for antecedent selection fails and one has to content with a wide range of competing candidates.
4. Just as in coreference, the pairwise mention-mention or mention-entity models is deficient as it does not allow for global consistency in resolution

Another problem is the interaction between coreference and bridging resolution, or more generally the joint modelling of all anaphora types.

The lack of big, reliably annotated data hamper the development of statistical models both for bridging anaphora recognition and bridging antecedent selection.

Bridging anaphora recognition is even worse than bridging antecedent selection. Morphosyntactic features don't work, and not much was gained from more knowledge based features. It usually depends on specific

context.

Challenge for the systems What kind of knowledge do we need to recognize bridging anaphora? Is it possible to make this knowledge explicit so the machine can learn it? Can we automatically acquire lexical resources necessary for some kinds of relation detection (e.g. from Wikipedia), or do we need to rely on hand-crafted (and possibly incomplete) resources such as WordNet?

Finally, there is the question of the impact of varying the granularity of the annotation scheme. Which tasks work better with simple binary classifications, and for which tasks is a wide range of possible labels better?

4.5 Ambiguity

Mentions can be ambiguous in a number of respects, ranging from the semantic question of whether they are referring or semantically vacuous, as in the *it* in (4.5.1), to the information structure question of whether they refer to a discourse new or discourse old entity (as in (4.5.2), where ‘your cream’ in the title could be referring either to Dermovate Cream (product) or to the specific instance of the product the customer holds in her hands, either the reference to Dermovate Cream could either corefer with the first mention or be a bridging reference) to the coreference question of what is the actual antecedent (as in (6.1.3), where three distinct areas could be the antecedent of the reference in the last utterance).

(4.5.1) "I beg your pardon!" said the Mouse, frowning, but very politely: "Did you speak?"

"Not I!" said the Lory hastily.

"I thought you did," said the Mouse. "--I proceed. Edwin and Morcar, the earls of Mercia and Northumbria, declared for him: and even Stigand, the patriotic archbishop of Canterbury, found it advisable--"

"Found WHAT?" said the Duck.

"Found IT," the Mouse replied rather crossly: "of course you know what 'it' means." (Alice in Wonderland)

(4.5.2) What is in your cream

Dermovate Cream is one of a group of medicines called topical steroids.

"Topical" means they are put on the skin. Topical steroids reduce the redness and itchiness of certain skin problems. (Gnome corpus)

(4.5.3) About 160 workers at a factory that made paper for the Kent filters were exposed to asbestos in the 1950s. Areas of the factory were particularly dusty where the crocidolite was used. Workers dumped large burlap sacks of the imported material into a huge bin, poured in cotton and acetate fibers and mechanically mixed the dry fibers in a process used to make filters. Workers described "clouds of blue dust" that hung over parts of the factory, even though exhaust fans ventilated the area. (Poesio & Vieira, 1998)

Challenges for the guidelines Artstein & Poesio found that ambiguity was best identified not by asking coders to code ambiguous expressions (explicit ambiguity coding) but by comparing the annotations produced by different coders (implicit ambiguity). More than 10 coders are however typically required for this, which is too costly for most projects.

A related ambiguity problem is found in a cross-linguistic setting. The PROIEL group annotated texts in both Greek and Latin, and they transferred the annotation to Armenian and Old Church Slavic (and they intend to transfer it to Gothic as well). These languages differ in various ways, the most relevant being the absence of a

deifinite article in the languages other than Greek. Since we have the same story (the New Testament) in all these languages, we do in a way expect the information structure to be 'the same'. After all, it is the same story, about the same referents. On the other hand, if we take the form of the linguistic expression in to account, we might expect there to be vast differences because of the lack of the article. Here is an example from the Vulgate:

(4.5.4)

quasi	ad	latronem	existis	cum	gladiis	et	fustibus
as	to	robber.SG.M.ACC	go_out.2.PL.PRF.ACT	with	swords.PL.M.ABL	ans	sticks.PL.M.ABL

'Be ye come out, as against a thief, with swords and staves?' (Vulg. LUKE 22.52)

gladiis is ambiguous between just "swords" (which would be NONSPEC in PROIEL) and "your swords" (ACC-inf (inferrable)? after all, these are soldiers). And if it is sufficient for ACC-gen that the referent is present in the situation, we could go for that. In the Bible, we can simply use the Greek as a reference. In this case, the Greek has no definite article, so ACC-inf is out. But of course we get the same problem in native Latin texts. The lack of determiners in Latin leads to all sorts of problems. For example, how does anaphoric reference really work in a language with no articles? Is an NP anaphoric/presuppositional just in case its presupposition is verified? Does it make sense to distinguish between accommodation, general world knowledge and plain NEW when there is no article? Consider the opening of Tacitus' Annals:

(4.5.5)

Urbem	Romam	a	principio	reges	habuere
city.ACC	Rome.ACC	from	beginning	kings.NOM	have.3.PL.PFV

'The city of Rome was from the beginning ruled by (the/some/Ø) kings.'

The definite article may sound unnatural in translation, but then we have a more distant relationship to the early Roman kings than the average Roman reader.

4.6 Quasi-coreference, metonymy

Many cases of disagreements between coders are ones in which the relationship between two entities is near to identity. Examples include

(4.6.1) On homecoming night Postville feels like Hometown, USA, but a look around this town of 2,000 shows it's become a miniature Ellis Island [...] For those who prefer the old Postville, Mayor John Hyman has a simple answer. (Recasens)

Notice that (4.6.1) is not always a case of ambiguity: in this example at least it's pretty clear from a semantic perspective that the old Postville does not refer to the individual Postville or to the stage current Postville, but to an earlier stage of the individual. However in practice when Recasens and colleagues attempted to ask coders to classify cases of near identity according to their semantic interpretation they had very bad agreement.

Challenges for the guidelines Recasens found that the best results at identifying quasi-coreference were obtained by simply comparing disagreements rather than asking coders to try to spot them.

Challenges for annotation tools: Near-identity relations create a chain of decreasing identity in that mention A

is nearly identical with mention B, mention B is nearly identical with mention C, but mentions A and C are not identical in any sense. Tools based on links are more appropriate for capturing this than tools that treat entities as equivalence classes and give the same ID to all coreferent mentions.

Challenge for systems Near-identity cases, which are borderline cases of coreference, introduce noise and adversely affect the performance of automatic systems. Near-identity annotations can be used to train coreference resolution systems that are able to discriminate clear identity of coreference from near-identity. A system that focuses on clear identity can learn a more accurate--and higher-performing--coreference resolution model. The automatic detection of near identity goes hand in hand with metonymy detection, bridging relations, etc and requires more sophisticated features.

A related issue is that of metonymy, illustrated by examples like (4.6.2). In this case the White House is clearly a reference to the person who lives in the building, ie the president, or possibly his staff.

(4.6.2) Rep. John Dingell, an important sponsor of President Bush's clean-air bill, plans to unveil a surprise proposal that would break with the White House on a centerpiece issue ...

4.7 Abstract anaphora

The majority of schemes require coders not to mark references to abstract entities as in (4.7.1). (Exceptions: AnCora, ARRAU, Dipper & Zinsmeister, Kolhatkar & Hirst, Navarreta, PDT) In most cases, only event anaphora is marked, as in (4.7.2). (E.g., OntoNotes.)

(4.7.1) There was nothing so VERY remarkable in that; nor did Alice think it so VERY much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually TOOK A WATCH OUT OF ITS WAISTCOAT-POCKET, and looked at it, and then hurried on, Alice started to her feet, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and fortunately was just in time to see it pop down a large rabbit-hole under the hedge.

(4.7.2) The car spun out of control and hit the guardrail. That happened at 3pm.

Challenges for the guidelines Abstract anaphora raises a number of issues for the guidelines, starting from identifying a span to be associated with the reference. With the exception of event anaphora, where the verbal complex could approximate the span, coders generally find it hard to agree on what the span is (Artstein & Poesio, 2006).

Challenges for the Annotation Tools Arguably, properly annotating abstract anaphora requires a different type of annotation tool.

Challenges for the resolvers Resolving abstract anaphora is extremely hard and only few systems attempted it (Eckert & Strube, Byron, Kolhatkar, ...)

4.8 For how long is a 'discourse old' entity still old?

The Phrase Detectives corpus contains many 'short' stories that are not in fact that short: they can be over twenty pages. In some of these stories, an entity is mentioned at the beginning and not again until the end. In RefLex, such entities are called 'r-given-displaced'.

Challenge for the theory How far away may the distance be between two occurrences to still count as

'r-given(-displaced)'? How to deal with topic changes in the intervening stretch (decreases the givenness of the second occurrence) versus elaborations of the topic under discussion (increases the givenness of the second occurrence)?

4.9 Beyond nominal coreference: e.g., Ellipsis

A proper study of coherence in text (e.g., for summarization or segmentation) will require going back to the beginning (UCREL) and start annotating ellipsis as well. We are only aware of a few efforts in this direction (Hardt, Bos)

Go to Part IV

RAIS/DocumentPartIII (zuletzt geändert am 2013-04-18 15:45:30 durch ArndtRiester)

- [IMS home](#)
- [wiki-support\(at\)ims](mailto:wiki-support(at)ims)
- [Legal Notice](#)