

Conditional Anaphora

Henk Zeevat

ILLC

University of Amsterdam

Nieuwe Doelenstraat 15

1012 VB Amsterdam

henk.zeevat@uva.nl

Abstract

An alternative route for pronoun resolution is explored in which not the accessibility relation of discourse representation theory is taken as a starting point, but the idea of a single information state that supplies antecedents for pronouns. The single information state is obtained by a non-monotonic relation of overlaying. The approach is applied to a number of outstanding puzzles with pronouns where it gives a simple and uniform treatment.

The restrictions on accessibility of antecedents for anaphora formed one of the attractive sides of early DRT but—as was quite clear already at the time—is too restrictive. (?) gives many difficult examples, especially with plurals, but there is a host of other types of counterexample, like the modal subordination cases, the anaphora discussed in Asher, the Geach sentence (?), the examples in (?), the paycheck sentences and others.

In the basic DRT fragment, accessibility can be described as existence in the local context of interpretation of the pronoun. This local context can be described by recursion: it is the local DRS merged together with any

higher DRS. One can quibble here about explicit existence as a discourse referent, or inferable existence. But that may well be conflating two kinds of restrictions on anaphoric relations: good antecedents do not merely exist in the local context, they also need to be maximally salient. And existents that are only inferable are almost per definition not salient enough. A realistic notion of pronominal antecedents combines the two: the antecedent is a salient local existent. So I take it that it is not formal presence of the discourse referent that explains why the tenth marble is not an antecedent, but a lack of salience¹.

If one considers extensions of the basic fragment to modals, corrections, or attitudes the local contexts need no longer be consistent. And this is a problem for the account I advocate. In inconsistent local contexts anything exists, in particular also the things that are not accessible for pronoun resolution but that are salient. The solution to this problem in DRT is technical: one defines accessibility not by local contexts but by geometrical configuration, as is done in certain kinds of syntax. The problem is however to explain the restrictions, not the descriptive definition.

The explanation above of accessibility in the basic fragment is not arbitrary. It follows

¹Salience is way too crude a concept as a cursory glance at the study of pronoun resolution tells us, but it is good enough in the context of this paper.

a view of Ewan Klein (p.c.) on the development of context dependent semantics. First, in the work of Montague shortciteMontague and (?), one had lists of contextual parameters functioning much like Tarski's variable assignments in his definition of satisfaction. The revolution which can be attributed to (?; ?; ?) and citeHeim is that this set of parameters can be replaced by an information state and that the utterance itself partly determines what the information state for the next sentence or the subordinated material is. There are some adaptations necessary, like developing a theory of deixis (e.g. one can have a pointed information state that indicates the current utterance as one of its discourse referents), but this is an interesting view of what happened in the dynamic revolution. In this view, the possible discourse antecedents must be discourse referents of the local context of the pronoun.

Another problem arises when one considers an updated version of Montague's scheme:

$$M \models \varphi[c, c']$$

In the new theory c is the incoming context and $c' = c[\varphi]$, the update of c with φ . But what about M ? Contexts that evolve in conversation are the works of humans and it cannot be excluded that imperfect knowledge will lead to error with respect to the state of affairs in which φ was uttered. The answer should be that it does not matter as long as there is a model on which φ , c and c' are true, i.e. they are all consistent and consistent with each other. Particular utterances may be false from an external point of view, but the conversational partners may not have noticed or have chosen to ignore the divergence. Only if the conversation gets trapped into inconsistencies, it cannot be a *bona fide* context of interpretation anymore and guide processes such as pronouns and presupposition resolution or disambiguation.

So the revolution seems to have led to an improvement in our understanding of what a

context is, of how the context influences interpretation and of how the communication itself changes the context. But apart from that it still follows the scheme of Montague. The context must be one simple object, e.g. a (pointed) information state and not some technical concoction out of information states, accessible discourse referents and whatever else.

So how about the extensions to modals, attitudes and corrections? The proposal of this paper is very simple. One should not merge the higher DRSs with the local one but only add so much of the higher DRSs as one can without becoming inconsistent. An operation doing this was invented by (?) (his satisfiable incrementation) which is a very cautious one: anything that could be inconsistent with anything is omitted. One can perhaps do better, but let's not worry about the refinements. This is Gazdar's definition applied to DRSs.

$$K \cup ! K_1 = K \cup \{A \in K_1 : \neg \exists K_2 \subseteq K \cup K_1 : K_2 \text{ is consistent and } K_2 \cup \{A\} \text{ is inconsistent}\}$$

It is not necessary that $\cup!$ constructs the information state. One can keep the information states separate and merely use $\cup!$ to compute the information state that is coded by two information states or a sequence of information states. One then has one information state (the context of interpretation) and a sequence of information states that can be used to determine where the incoming new information goes (foregrounded material to the first element, accommodations towards the tail) and the result of the complete updates. So I assume that $\cup!$ merely gives the extension of a syntactic operation **over** that connects information states.

Local contexts can now be defined for corrections, modals and belief contexts in the following way.

Corrections

A: Bill ate the cake.

B: No, it was John who ate the cake.

1 over K where K is the old DRS and 1 the

empty DRS.

1 is updated by “It was John who ate the cake” using 1 over K or intermediate stages $K1$ over K as the context of interpretation (e.g. for the resolution of the presupposition “ x ate the cake” that may well be resolved to part of the other speaker’s utterance: Bill ate the cake). The updating process leads to a state $K2$ over $K3$ and its denotation $K2! \cup K3$ is the result of the correction.

Beliefs: John believes that S .

$bel(j, K)$ over K where K is the old DRS, $bel(x, K)$ the DRS representing the beliefs that x has according to K . S updates $bel(x, K)$ with the indicated context of interpretation that may be changed by accommodations. The result of the update $K1$ may be entered into the (possibly changed) K as a condition $belief(j, K1 \setminus bel(x, K))$.

Modals: It might have been that S .

1 is updated by S using 1 over K as the context of interpretation and the result can be stored under an appropriate modal operator.

counterfactuals: if A , would B

A uses 1 over K as context of interpretation, and B the result of updating 1 over K with A . If it is to be stored (rather than just checked?), this could be done by some suitable new syntax.

In corrections, initially the whole background is visible but correcting material will soon hide the corrected material. In beliefs, other beliefs that are in conflict with the common ground may hide parts of the common ground.

The proposal here leads to contexts of interpretation that make antecedents available that are not themselves identified by the local context: locally they may not be the object that is called “Bill”, or the speaker’s brother. The local contexts must identify these objects in their own ways, and the only claim seems to be that if the embedded context were true, the referents of the pronoun would be the same as the referents of the antecedents in the embed-

ding context. In that sense, the modals, corrections, beliefs and suggestions can be said to be about the referents of their antecedents, if any. This is an approximation of the sense in which these modals and attitudes can be said to be about the objects that the embedding context is about. A lot can be said about this view of “quantifying in”, but this is not the place. Suffice it to say that it conforms with a view in which the belief subject has her own mode of presentation and in which there is no criterion for when a mode of presentation is good enough for supporting “de re belief”.

1a. John will come tonight.

1b. No, he is ill.

1c. Bill thinks that he will not

1d. He may be ill.

1e. If he forgot, he will not.

But the proposal can be extended. It is not problematic to take salient non-entailed material in the context (material that was denied, material that was merely reported or suggested) and add it in exactly the same way to the embedding context, i.e. the local material is added independently of its truth and can obliterate material from the context in which it is embedded. This allows pronominal reference to objects that the context itself is not committed to.

It gives an analysis of intentional anaphora. These are the famous cases like the Geach sentence or the examples provided by Edelman. In all cases we, the conversationalists, know that there are no witches, that nobody had an accident and that Smith and Jones just had an accident. In (2b.), Harry arranged a fake accident by pushing a car against a tree and spraying tomato ketchup in the grass next to the driver’s seat. John is the first who notices the car and reaches the conclusion that somebody had an accident. Mary arrives when John has left the scene, notices the ketchup and reaches her conclusion that the driver is wounded.

In (2c.) and (2d.) the two detectives are investigating the putative murders of Smith and Jones but have reached different conclusions. Arsky thinks that two different murderers were involved but Barsky thinks that only one murderer was responsible.

2a. Hob believes that a witch killed his pig and Nob believes that she poisoned his well.

2b. John thinks that someone had an accident and Mary thinks he was wounded.

2c. Arsky thinks that someone killed Smith and Barsky thinks that he killed Jones too.

2d. Barsky thinks that someone killed Jones and Arsky thinks that he killed Smith (too).

On the current proposal, the second clauses get interpreted in a context of interpretation that overlays the content of the belief that provides the antecedent over the embedding context. This provides an antecedent for the pronoun in the second clause, i.e. on the one hand, it licenses the speaker to use the pronoun because the intended referent is a highly activated member of the context for this part of the generation process and, on the other hand, it allows the interpreter to interpret the pronoun as standing for this highly activated member of the local context of interpretation.

The claim the speaker has to defend is that the amended context (the embedding context plus what Hob, John, Arsky or Barsky believes) fixes the referent of the pronoun in the second conjuncts: if the amended context were true, it would identify the object Nob, Mary, Barsky or Arsky has their belief about. This is true for (3b.) and (3c.) and false for (3d.) under the background that Edelberg provides. (in d. the first conjunct can be true because somebody killed Jones and nobody else and Arsky does not believe of that guy that he killed Smith). In (3a.) the counterfactual is true under any of the possible explanations that philosophers have provided: Hob telling Nob about his belief, a story in the newspaper, a rumour in the village etc. All these involve communication of some kind, but as Edelberg

showed this is not essential. The real explanation is that a counterfactual is true: if the suggested material were true, the pronoun would refer to the same object as the antecedent.

Apparently the context that was computed for the antecedent belief is still available when the second belief report comes along and provides all that is needed for pronominal reference, independently of any need for that context in the update process for the second belief report.

It is also not hard to state the truth-conditional contribution of the anaphoric relation, provided there will be a day when a good truth conditional semantics for counterfactuals is available: if both beliefs were true, the referent of the antecedent would be the same as the antecedent of the pronoun. This is what is predicted by the complicated account in (?), but that account lacks a satisfactory intuitive foundation.

The account in this paper is not just another philosophical theory. It is a cognitive science hypothesis. In constructing a context of interpretation for beliefs, by overlaying a local context on top of the embedding context, a picture is formed of what it would be like if the belief were true. The pronoun picks up a referent from the picture. The interpretation of that referent is conditioned by the counterfactual obtained by counterfactually assuming the truth of the antecedent belief. The operation of overlaying one context with another is a natural ingredient of the semantics of counterfactuals. To the extent that the Ramsey test is not by itself the correct story about counterfactuals, it has a reflex on the account of this paper: Gazdar's operation is not a final truth about overlaying either. Any improvement should apply to both overlaying and counterfactuals.

The account straightforwardly extends to modal subordination and is not in conflict with the existing accounts of that, which however cannot deal with cases with two different

operators, like (3).

3. A wolf might come in. Bill thinks it is prowling about in the neighbourhood.

The counterfactual is: if a wolf would come in it would be the one that Bill thinks is prowling about. This forces specificity on both Bill's belief and the speaker's modal, but not existence. Both may come from an unfounded rumour, as in the Hob-Nob example.

More interesting are the like the paycheck sentence and Landman's famous (4).

4. If a farmer owns a donkey, he beats it. If he owns a horse, he treats it well.

Assume the second clause is interpreted with respect to a context of interpretation that still has: "a farmer owns a donkey" on top. Overlaying "he owns a horse" now functions as a correction and we can take the result of the correction as giving the content of the condition. This may be the same mechanism that operates in contrast under parallelism, as in the paycheck cases and squares well with the observation that both kinds of anaphora are limited to parallelism.

I started from the observation that discourse representation theory has been forced to give up one of the intuitions on which it is founded: a context of interpretation that is an information state and which supplies information to the interpretation process. I have explored a way to keep this information in and gave up instead monotonicity. It should not be controversial to assume that human cognition is able to do corrections and consider counterfactual states of affairs. It should therefore not be a surprise that one gains a new perspective on some quite old puzzles.

References

- Walter Edelberg. 1992. Intentional identity and the attitudes. *Linguistics and Philosophy*, 15:561–598.
- Gerald Gazdar. 1979. *Pragmatics: Implicature, Presupposition and Logical Form*. Academic Press, New York.
- Peter Geach. 1962. *Reference and Generality*. Cornell University Press, Ithaca, New York.
- Hans Kamp. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen, and Martin Stokhof, editors, *Formal Methods in the Study of Language, Part 1*, volume 135, pages 277–322. Mathematical Centre Tracts, Amsterdam. Reprinted in Jeroen Groenendijk, Theo Janssen and Martin Stokhof (eds), 1984, *Truth, Interpretation, and Information; Selected Papers from the Third Amsterdam Colloquium*, Foris, Dordrecht, pp. 1–41.
- David Kaplan. 1989. Demonstratives. In J. Almog, J. Perry, and H. Wettstein, editors, *Themes from Kaplan*, volume 135, pages 481–566. Oxford University Press, New York.
- Lauri Karttunen. 1976. Discourse referents. In James McCawley, editor, *Syntax and Semantics 2: Notes From the Linguistic Underground*, pages 363–385. Academic Press, New York.
- Robert Stalnaker. 1979. Assertion. In Peter Cole, editor, *Syntax and Semantics*, volume 9. Academic Press, London.
- Henk Zeevat. 2001. Demonstratives in discourse. *Journal of Semantics*, 16:279–314.