Anaphora Resolution and Minimal Models

Ariel Cohen

Ben-Gurion University, Israel
arikc@bgu.ac.il

Abstract

Some anaphora resolution algorithms are based on model builders, and use the fact that they generate minimal models: only those elements that are necessary are postulated to exist in the model. In this way, such systems have the desirable property that if anaphora can be resolved to a linguistically available antecedent, this resolution applies, and only if there is no suitable antecedent, a deictic reading is generated.

In this paper I formalize the entailments that follow from such anaphora resolution algorithms. In particular, I will suggest a simple, linguistically motivated, underspecified representation for anaphora—DRT, and place the burden of the resolution of anaphora and its consequences on an independently motivated logic for default reasoning—Default Logic.

1 Introduction

Consider a simple case of ambiguous anaphoric reference:

(1) I had gone to see John before I visited Bill and Mary. He doesn't want to speak with her.

What can we say about the resolution of the anaphora? The pronoun *her* probably refers to Mary; the pronoun *he* is ambiguous between John and Bill, but most likely refers to John. And either pronoun (or both) may be used deictically, referring to some other individual that is not denoted by a linguistic antecedent. What we would like is a system that allows us to represent all these options, pick those we consider plausible, and draw some inferences even in the absence of a clear resolution.

Intuitively, the deictic interpretation is dispreferred; we will assume it only if there is no suitable linguistic antecedent. An elegant explanation of this fact can be provided by anaphora resolution algorithms that use *domain building* techniques (e.g., [1,10,11]). Model builders receive as input a set of propositions, and produce a model for them if such exists. Typically, the models so generated are minimal, i.e. models whose domain is only as large as it needs to be. Thus, if the referent of the pronoun can be identified with a linguistic antecedent, no additional elements need to be postulated. Only if this turns out to be impossible, will an additional element be added to the model, resulting in the deictic reading.

The goal of this paper is not to propose new algorithms, but to formalize the idea of using minimal models to resolve (pronominal) anaphora, and the conclusions that can be drawn by such a system in case the anaphora is not resolved. [1]

2 An Underspecified Representation for Anaphora

As the discourse in (1) exemplifies, anaphora is often ambiguous. Moreover, the deictic possibility always exists, so it is always possible, in principle, that what we had identified as the antecedent of a pronoun actually is not, and the pronoun is used deictically. In the case of (1), since we have two pronouns, one with three possible interpretations (John, Bill, or the deictic use) and the other with two (Mary or deictic), we will have six potential interpretations. We need to be able to represent the ambiguity, but still draw inferences as best we can on the basis of what we know. This calls for some sort of underspecified representation, and some inference mechanism to derive conclusions from it.

Many special formalisms have been proposed, whose sole purpose is to allow efficient representation of and reasoning with underspecification. I will not, however, go down this road, for several reasons. A formalism that is not independently motivated on linguistic grounds, and whose sole justification is to represent underspecification, may work in a practical system, but its explanatory adequacy from a linguistic point of view would be dubious. To give one example, recall that deictic readings of a pronoun are always possible, and this is the case across languages. Why is this? Why don't we have languages where pronouns are restricted to linguistic antecedents only, and deictic readings are indicated only by, say, demonstratives? A formalism that is only geared toward underspecification would be quite adequate if pronouns could only refer to linguistic antecedents, and it is hard to see why it would necessitate the availability of deictic readings. It is, of course, preferable to have the possibility of deictic readings follow directly from the representation, thus explaining the puzzle.

Furthermore, a nonstandard representation will typically require nonstandard inference methods, especially tailored for the representation. [2] Again, these inference methods would not be independently justified, unlike rules of common-sense inference that must, in one way or another, be used in order to understand natural language.

[1] While this paper only deals with pronominal anaphora, the approach may be extended to handle definite descriptions—see [6,7]) for an account based on model building.

[2] Though see [15], who uses a nonstandard representation of anaphora, but applies Default Logic to generate its perceived readings.

An additional reason for keeping the representation as simple and as close to standard linguistic representations as possible is the fact that it is not likely to be replaced by a fully specified representation during the interpretation process. Normally, one uses an underspecified representation in the hope that, in the fullness of time, or as the need arises, it will be fully specified. In this sense, an underspecified representation is only a "temporary measure." However, because it is always possible to interpret pronouns deictically, we can never fully specify the representation. The possibility always exists that we will receive later some information that will force us to interpret the pronoun deictically and undo our previous resolution. Hence, the representation of anaphora cannot be treated as a temporary measure, and must be as close as possible to the fully motivated representation.

In this paper I am going to suggest that we don't need to look far for a representation and its associated inference method. A standard, linguistically motivated representation, without special machinery for underspecification, will do.[3] For concreteness, I choose Discourse Representation Theory ([9]). Thus, for example, the discourse in (1) will be represented by the following DRS (here and elsewhere, ignoring tense and the possibility of a collective reading of the conjunction):

(2)

$x\ y\ z\ u\ v$
John(x)
Bill(y)
Mary(z)
go-to-see(I,x)
visit(I,y)
visit(I,z)
male(u)
female(v)
¬want-to-speak(u,v)

Note that this DRS does not resolve the anaphora. In this representation, u and v are subject to existential closure, and all we know is that *some* antecedent exists. So, in effect, the DRS (2) is an underspecified representation, containing all the possible ways of resolving the anaphora. Any specific resolution of the anaphora results in the addition of equalities identifying the referents of the pronouns. For example:

[3] Of course, it may be the case that some sort of special underspecified representation is needed for other reasons, e.g., to represent scope ambiguities. All I claim is that such special representations are not necessitated by the need to represent anaphora.

(3)

$x\ y\ z\ u\ v$
John(x)
Bill(y)
Mary(z)
go-to-see(I,x)
visit(I,y)
visit(I,z)
male(u)
female(v)
¬want-to-speak(u,v)
u=x
v=z

The problem of anaphora resolution now becomes the problem of inferring the necessary equalities from the representation. Of course, DRT places some constraints on acceptable antecedents—they have to be accessible. Accessibility constraints can be modeled simply as inequalities between all inaccessible pairs of discourse referents. Additional constraints come from our world knowledge. For example, if we know that *her* must refer to a female individual, and that John is not female, we know that John cannot be a suitable antecedent.

3 Default Logic

Inferring the equalities identifying pronoun with antecedent must be nonmonotonic: we may later find that our anaphora resolution was wrong, and revise it. Some form of nonmonotonic reasoning, attempting to derive consistent conclusions from an incomplete description of the world, is independently necessary for any kind of system that attempts to draw inferences from natural language texts. Thus, instead of devising a special form of inference mechanism for our underspecified representation, I will use well studied and independently motivated mechanisms for nonmonotonic reasoning (cf. [12,15]).

Specifically, I choose Default Logic ([17]). Default Logic is one of the most widely used nonmonotonic formalisms and may be the only one that has a clearly useful contribution to the wider field of computer science through logic programming and database theory.

A *default theory* is a pair (D, A), where D is a set of defaults and A is a set of first-order sentences (axioms). Defaults are expressions of the form

$$(4) \qquad \frac{\alpha(x) : \beta_1(x), \ldots, \beta_m(x)}{\gamma(x)},$$

where $\alpha(x)$, $\beta_1(x), \ldots, \beta_m(x)$, $m \geq 1$, and $\gamma(x)$ are formulas of first-order logic whose free variables are among $x = x_1, \ldots, x_n$. A default is *closed* if none of $\alpha, \beta_1, \ldots, \beta_m$, and γ contains a free variable. Otherwise it is *open*.

Roughly speaking, the intuitive meaning of a default is as follows. For every n-tuple of objects $t = t_1, \ldots, t_n$, if $\alpha(t)$ is believed, and the $\beta_i(t)$s are consistent with one's beliefs, then one is permitted to deduce $\gamma(t)$.

Crucial to the interpretation of Default Logic is the notion of an *extension*. Roughly speaking, an extension of a default theory is a set of statements containing all the logical entailments of the theory, plus as many of the default inferences as can be consistently believed. Sometimes a default theory has more than one extension, as in the well known *Nixon diamond*. Suppose we have the following set of defaults:

$$(5) \qquad \left\{ \frac{\mathbf{Quaker}(x) : \mathbf{pacifist}(x)}{\mathbf{pacifist}(x)}, \frac{\mathbf{Republican}(x) : \neg\mathbf{pacifist}(x)}{\neg\mathbf{pacifist}(x)} \right\}.$$

If Nixon is both a Quaker and a Republican, in one extension he will be pacifist, and in another he won't be. So, is Nixon a pacifist or isn't he?

When faced with multiple extensions, there are two general strategies we can use to decide which conclusions to accept: skeptical or credulous reasoning. Skeptical reasoning means taking only what is true in all extensions. In the case of the Nixon diamond, we will believe neither that Nixon is a pacifist, nor that he is not a pacifist. Credulous reasoning means picking one extension, based on whatever principles one deems appropriate, and accepting its conclusions. This means we will pick one extension, perhaps using our knowledge of Nixon's statements and actions, and based on this extension, conclude whether he is a pacifist or not.

4 Equality by Default

4.1 A default rule for equality

Resolving anaphora means generating an equality between two discourse referents. I suggest that we will generate such an equality by default: we assume that two elements are equal if they cannot be proved to be different. The idea underlying this notion has been proposed, though not formalized, in [2]. Charniak's approach is further explored in [5], and formalized more fully in [3,4].

The idea of equality by default can be implemented in Default Logic very simply, by adding the following default:

$$(6) \qquad \frac{: x = y}{x = y}$$

This default rule means that whenever it is consistent to assume that two elements are the same, conclude that they are.[4] What does it mean to say that it is consistent to assume $x = y$? It means that it not known that $x \neq y$. From the axioms of equality it follows that this is equivalent to saying that there is no property ϕ s.t. we know $\phi(x)$ but we also know $\neg\phi(y)$.

4.2 Minimality of models

In order to explain what it means for the models of our theory to be minimal, we will need some definitions. In particular, since (6) is an open default, we need to provide a semantic definition of extensions of open default theories. Since model builders generate what are, in essence, Herbrand models, it seems natural to assume that the theory domain is a Herbrand universe (cf. [14, Chapter 1, §3]). Fortunately, such a definition has already been proposed ([13,8]), and I will follow it here.

Suppose we have a first order language $\mathcal{L}$, and we augment it with a set of new constants, b, calling the resulting language $\mathcal{L}_b$. The set of all closed terms of the language $\mathcal{L}_b$ is called the *Herbrand universe* of $\mathcal{L}_b$ and is denoted $\boldsymbol{T}_{\mathcal{L}_b}$.

A *Herbrand b-interpretation* is a set of closed atomic formulas of $\mathcal{L}_b$.

Let w be a Herbrand b-interpretation and let φ be a closed formula over $\mathcal{L}_b$. We say that w *satisfies* φ, denoted $w \models \varphi$, if the following holds:

(i) If φ is an atomic formula, then $w \models \varphi$ if and only if $\varphi \in w$;
(ii) $w \models \varphi \rightarrow \psi$ if and only if $w \not\models \varphi$ or $w \models \psi$;
(iii) $w \models \neg\varphi$ if and only if $w \not\models \varphi$; and
(iv) $w \models \forall x \varphi(x)$ if and only if for each $t \in \boldsymbol{T}_{\mathcal{L}_b}$, $w \models \varphi(t)$.

For a Herbrand b-interpretation w, the $\mathcal{L}_b$-*theory* of w, denoted $\boldsymbol{Th}_{\mathcal{L}_b}(w)$, is the set of all closed formulas of $\mathcal{L}_b$ satisfied by w. For a set of Herbrand b-interpretations W, the $\mathcal{L}_b$-*theory* of W, denoted $\boldsymbol{Th}_{\mathcal{L}_b}(W)$, is the set of all closed formulas of $\mathcal{L}_b$ satisfied by all elements of W.

Let E be a set of closed formulas over $\mathcal{L}_b$. We say that w is a *Herbrand b-model* of E, denoted by $w \models E$, if $E \subseteq \boldsymbol{Th}_{\mathcal{L}_b}(w)$.

Extensions of open default theories are then defined as follows:

Definition 1 *(cf. [8, Definition 27]) Let b be a set of new constant symbols and let (D, A) be a default theory. For any set of Herbrand b-interpretations W let $\Delta^b_{(D,A)}(W)$ be the largest set V of Herbrand b-models of A that satisfies the following condition.*

For any default $\dfrac{\alpha(x) : \beta_1(x), \beta_2(x), \ldots, \beta_m(x)}{\gamma(x)} \in D$ and any tuple t of elements of $\boldsymbol{T}_{\mathcal{L}_b}$ if $V \models \alpha(t)$ and $W \not\models \neg\beta_i(t)$, $i = 1, 2, \ldots, m$, then $V \models \gamma(t)$.

A set of sentences E is called a b-extension *for (D, A) if $E = \boldsymbol{Th}_{\mathcal{L}_b}(W)$*

[4] Note that this is, in a sense, the opposite of the Unique Name Assumption ([16]). The uniqueness of names can still be ensured, by following standard DRT practice and defining appropriate external anchors.

for some fixpoint W *of* $\Delta^b_{(D,A)}$.

It has been shown ([4]) that if E is a b-extension for the default theory $\left(\left\{\frac{:x=y}{x=y}\right\}, A\right)$, and w is a Herbrand b-model of E, then w is *minimal.* That is to say, there is no Herbrand b-model w' of E such that

$$(7) \qquad \{\langle t_1, t_2\rangle : w \models t_1 = t_2\} \subset \{\langle t_1, t_2\rangle : w' \models t_1 = t_2\}.$$

In other words, the proposed default theory minimizes the number of different elements in the models, as desired.

4.3 Deictic interpretations

It turns out that using Herbrand models has a consequence that is particularly important for our purposes. Note that the new elements introduced in b, by being new, are equal by default to any term. In particular, they are are equal by default to any pronoun; this is the reason why deictic interpretations of pronouns are always possible. Hence, we have a logical explanation for a linguistic phenomenon—the universal availability of deictic readings.

Note that this theory allows deictic readings, but only as a last resort, when no other readings are possible. Given the discourse in (1), we have a good reason to believe that *her* refers to Mary, i.e. $v = z$. It is true that we have in the Herbrand model additional new terms, but this does not negate the minimality of the model. Since these terms are new, nothing is known about them and consequently it is consistent to assume that, for any such new term n_i, $v = n_i$. It is also consistent to hold the conjunction of all these beliefs, namely the belief that $n_1 = n_2 = n_3 = \cdots = v$. So, the model is, indeed, minimal; the addition of new constants does not mean that they denote additional entities. Thus, we capture the intuition that deictic readings are dispreferred, and are only available when no suitable antecedent is available. Note that if we didn't have this requirement of minimality, deictic readings would be on an equal footing with anaphoric readings.

If necessary, however, we *can* get a deictic interpretation, i.e. equate the pronoun with an element that is different from all other discourse referents. This happens when no possible antecedent is available, i.e. for every discourse referent t other than v, we know, or can deduce $v \neq t$. Then, we will have an extension where for some new term n_i, $v = n_i$. By the axioms of equality, n_i will not be equal to any of the other discourse referents, hence the domain will not be minimal. Of course, we may have an extension where the new terms are be equal to other terms, but none will be equal to v; but this extension will not constitute resolution of the anaphora, and will therefore be ruled out.

5 Inference

Let us see the kinds of inferences that this theory gives rise to. First, note that, although we are quite liberal in our assumption of equality, we can still rule out inappropriate antecedents. Recall that antecedents that are not accessible, in the DRT sense, will be explicitly stated to be different from the pronoun. Hence, obviously, it will not be consistent to assume that they are, so such equalities will arise in any extension.

We can also rule out antecedents that are semantically incompatible. For example, if we know that $\mathbf{male}(u)$ but $\neg\mathbf{male}(z)$, we cannot assume $u = z$; this is because if u is male and z is female, they have to be different, by the axioms of equality.

But suppose we have two acceptable antecedents for a pronoun u: in our example, it is possible that $u = x$ (John), but it is also possible that $u = y$ (Bill). If we know that they are different people, we know $x \neq y$, so it is impossible to believe both $u = x$ and $u = y$. We will therefore have two extensions: in one of them, the pronoun is equated with John, and in the other, with Bill.

How do we deal with these extensions? If we prefer one antecedent over the other, for reasons of pragmatic plausibility or salience, we apply credulous reasoning and pick the appropriate extension. In this extension, the pronoun will be equated with the chosen antecedent; hence, by the nature of equality, all properties of the antecedent will also hold of the pronoun.

At other times, however, the anaphora may be genuinely ambiguous, and we may have no reason to prefer one reading over the other. In this case, it makes sense to apply skeptical reasoning, and accept only what is true in all candidate extensions.

Consider, for example, the following discourse:

(8) John met Bill at the ice cream parlor. He was upset.

In this case, the pronoun may be equated with either John or Bill, and there are no good grounds, without further context, to decide between them. Yet, we do know something about the antecedent of the pronoun: he was at the ice cream parlor. We know this because we know that both John and Bill were there, and the pronoun refers to one of them. Skeptical reasoning will, indeed, give us precisely this result, since in both extensions, the pronoun has the properties that its antecedent has.

But now suppose that one possible antecedent has a property than the other one lacks:

(9) John walked along the sidewalk and saw Bill inside the ice cream parlor. He was upset.

In this case, Bill has the property of being inside the ice cream parlor, but John does not. Thus, in one extension, the pronoun will have this property, and in another—its negation. If we have no reason to prefer one extension over the other, we will apply skeptical reasoning, and will not conclude of the referent either that he is or that he is not inside the ice cream parlor. This appears intuitively correct.

Now suppose we know that some property holds of one potential antecedent, but we don't know whether it holds of another:

(10) While eating ice cream, John saw Bill at the ice cream parlor. He was upset.

We know that John was eating ice cream, but we do not know whether Bill was eating ice cream too or not. In this case, intuitively, we cannot conclude about the antecedent of the pronoun that he was eating ice cream, although this is consistent with him being either John or Bill. Indeed ,the proposed system conforms with this judgment. This is because in one extension, the one where the pronoun is associated with John, the property of eating ice cream is predicated of the discourse referent corresponding to the pronoun. But in the other extension, neither this property nor its negation will be so predicated. So, in this extension it will not be true that "he" is eating ice cream, hence skeptical reasoning will not license this inference.

Note that I have ignored here the addition of the new terms. The reason is simple: since they are new, they do not make a difference to the inference patterns discussed above. Consider, for example, the inference associated with (8) again. Suppose we have a new term n_i. So long as it is possible to find at least one antecedent to the pronoun, a model for the deictic reading, i.e. where the pronoun is equated with n_i but with no other element, will not be minimal, hence it will not be the model of any extension. In every extension, then, the pronoun u will be equated with some discourse referent x. Now, suppose $n_i = x$. In this case, by the axioms of equality, n_i will also have the property of being at the ice cream parlor, hence skeptical reasoning will still conclude that the pronoun has this property. Alternatively, suppose $n_i \neq x$ (perhaps because it is associated deictically with another pronoun). Now, it follows that $n_i \neq u$, so whether or not n_i was at the ice cream parlor should have no effect on whether "he" was.

6 Conclusion

I have proposed a theory of the representation of anaphora, based on the assumption that if two elements cannot be proved to be different, then they can be assumed to be equal. This assumption is implemented using a standard linguistic representation (DRT) and a standard default reasoning system (Default Logic), and this requires no special mechanisms for representation or inference. Yet this conceptually simple theory appears to produce exactly the sort of inferences regarding anaphora that are intuitively desirable.

References

[1] Baumgartner, P. and M. Kühn, *Abducing coreference by model construction*, Journal of Language and Computation **1** (2000), pp. 175–190.

[2] Charniak, E., *Motivation analysis, abductive unification and nonmonotonic equality*, Artificial Intelligence **34** (1988), pp. 275–295.

[3] Cohen, A., M. Kaminski and J. A. Makowsky, *Indistinguishability by default*, in: S. Artemov, H. Barringer, A. S. d'Avila Garcez, L. C. Lamb and J. Woods, editors, *We Will Show Them: Essays in Honour of Dov Gabbay*, College Publications, 2005 pp. 415–428.

[4] Cohen, A., M. Kaminski and J. A. Makowsky, *Applying default logic to anaphora, vagueness, and uncertain reasoning* (in preparation).

[5] Cohen, A. and J. A. Makowsky, *Two approaches to nonmonotonic equality*, Technical Report CIS-9317, Technion—Israel Institute of Technology (1993).

[6] Gardent, C. and K. Konrad, *Definites and the proper treatment of rabbits*, in: C. Monz and M. de Rijke, editors, *Proceedings of the 1st Workshop on Inference in Computational Semantics, ICOS-1*, 1999, pp. 53–69.

[7] Gardent, C. and K. Konrad, *Interpreting definites using model generation*, Journal of Logic, Language and Information **1** (2000), pp. 193–209.

[8] Kaminski, M., *A comparative study of open default theories*, Artificial Intelligence **77** (1995), pp. 285–319.

[9] Kamp, H. and U. Reyle, "From Discourse to Logic," Kluwer Academic Publishers, Dordrecht, 1993.

[10] Kohlhase, M., *Model generation for discourse representation theory*, in: W. Horn, editor, *Proceedings of the 14th European Conference on Artifical Intelligence*, 2000, pp. 441–445.

[11] Konrad, K., "Model Generation for Natural Language Interpretation and Analysis," Ph.D. thesis, University of Saarlandes, Saarbrücken (2000).

[12] Lascarides, A. and N. Asher, *Temporal interpretation, discourse relations and common sense entailments*, Linguistics and Philosophy **16** (1993), pp. 437–493.

[13] Lifschitz, V., *On open defaults*, in: J. Lloyd, editor, *Computational Logic: Symposium Proceedings* (1990), pp. 80–95.

[14] Lloyd, J., "Foundation of logic programming, second extended edition," Springer–Verlag, Berlin, 1993.

[15] Poesio, M., *Semantic ambiguity and perceived ambiguity*, in: K. van Deemter and S. Peters, editors, *Semantic Ambiguity and Underspecification*, CSLI, Stanford, 1996 pp. 159–201.

[16] Reiter, R., *Equality and domain closure in first order databases*, Journal of the ACM **27** (1980), pp. 235–249.

[17] Reiter, R., *A logic for default reasoning*, Artificial Intelligence **13** (1980), pp. 81–132.