Anaphora Resolution by Default

Ariel Cohen
Dept. of Foreign Literatures and Linguistics
Ben-Gurion University of the Negev
arikc@bgu.ac.il

Abstract

The resolution of anaphora is dependent on a number of factors discussed in the literature: syntactic parallelism, topicality, etc. A system that attempts to resolve anaphora will typically require the interaction of many of these factors. In addition, there must be a principle that simply says that the system needs to look for an antecedent. Without such a principle, if none of the factors recommend a clear winner, the system will be left without an antecedent. This principle should work in such a way that, if there is exactly one good candidate antecedent, the system will choose it; if there are more than one, the system will still attempt to identify one, or, at least, draw some inferences about the likely antecedent; and, in case there is no candidate, the system will produce an accommodated or deictic reading.

Many systems embody some version of the this principle procedurally, as part of the workings of their algorithm. However, because it is not explicitly formalized, it is hard to draw firm conclusions about what the system would do in any given case. In this paper I define a general principle of Equality by Default, formalize it in Default Logic, and demonstrate that it produces the desired behavior. Since all other factors can also be formalized in Default Logic, the principle does not need to be left implicit in the algorithm, and can be integrated seamlessly into the rest of the explicit rules affecting anaphora resolution.

1 Factors Determining Anaphora Resolution

Identifying the antecedent of an anaphoric trigger (a pronoun, definite DP, etc.) depends on an interaction of many factors: syntactic (e.g. Binding Theory), semantic (e.g. selectional restrictions), and pragmatic (e.g. Centering Theory). Some of these factors, such as selectional restrictions and

syntactic binding requirements rule out certain antecedents, while other factors, e.g. topicality, suggest that a certain antecedent should be chosen.

Most, perhaps all of these factors are defeasible. Consider, for example, the following discourse, from Asher (1984):

(1) The Vice-President entered the President's office. *He* was nervous and clutching his briefcase. After all, he couldn't fire the Vice-President without making trouble for himself with the chairman of the board.

The pronoun in the second sentence has two potential antecedents: *the Vice-President* or *the President*. Clearly *the Vice-President* is preferred: it has the same syntactic position (subject) as the pronoun, and it is more salient. However, by the time the third sentence is processed, it is clear that this choice is wrong, and the intended antecedent is, in fact, *the President*.

Such examples abound; and they indicate that (all?) anaphora resolution factors are best thought of as defaults, which may be overridden. It is therefore attractive to model anaphora resolution as a system of prioritized defaults (Lascarides and Asher, 1993; Mitkov, 1995; Poesio, 1996).

In this paper I will not attempt to identify these factors or their relative strengths. What I do wish to point out is that formalizing all these factors is not enough. To see why, consider the discourse in (1) again. The antecedent that is eventually chosen, *the President*, is not suggested by any of the well known factors discussed in the literature: it is neither topical, nor a subject, nor does it have a parallel syntactic position, etc. This antecedent is simply chosen as a last resort, since the other potential candidate is ruled out. This "last resort" rule must be defined somehow, for, without it, no antecedent whatsoever would be chosen. Indeed, in the linguistics literature, such a rule has been proposed (Williams, 1997, p. 603):

(2) *Don't Overlook Anaphoric Possibilities* (DOAP)
Opportunities to anaphorize text must be seized.

Essentially, this rule says simply that, when we encounter a trigger , we must try to find an antecedent. If we find an antecedent that is suggested by some rule, so much the better; but even a dispreferred antecedent is better than no antecedent at all. Linguists have used DOAP in systems of prioritized defaults for anaphora resolution (Hendriks and de Hoop, 2001).

However, while factors such as syntactic parallelism or selectional restrictions are, at least conceptually, easy to implement, it is not clear how to formalize DOAP in such a way that it could be implemented. My proposal is an attempt to provide such a formalization, which, in combination

with other factors, has the potential to bring about a fully explicit system of anaphora resolution.

Of course, in practice almost all anaphora resolution algorithm obey DOAP, in the sense that they always attempt to find (at least) one antecedent, even if the anaphora is ambiguous. However, if DOAP is not defined explicitly in the object level of the logic, but is left to a metalevel description, it is hard to be clear on, let alone prove, what the system will do when there is no clear choice of antecedent: which, if any, antecedent it will choose, and which inferences it will draw.

Take, for example, systems that use model building techniques, typically generate minimal models. Minimality could be seen as an implementation of DOAP: A model in which the antecedent of a referring expression is not identified is not minimal (since it has an additional entity, namely the reference of the trigger); it is therefore dispreferred, and the anaphoric reading is chosen, if possible.

However, for many of these systems, the model cannot always be relied upon to be minimal (Bry and Yahya, 2000). Even where it can, minimality of the model is not sufficient to ensure that an antecedent is chosen.

Consider, for example, the following discourse:

(3) John met Mary. *He* didn't talk to *her*.

A model builder would generate a model whose universe consists of John and Mary, and where the denotation of the predicate *talk to* is the empty set. This model satisfies the discourse in (3), and is clearly minimal, yet it says nothing about which antecedents the pronouns refer to.

An explicit formalization of DOAP should deal with cases where there is one clear antecedent, as well as with cases where there isn't. In general, when an anaphoric trigger is encountered, there are three possibilities:

(4) a. John was eating ice cream. *He* was upset.

b. John was eating ice cream. *The waitress* brought him the check.

c. John and Bill met at the ice cream parlor. *He* was upset.

One possibility is that there is exactly one appropriate antecedent, as in (4a): *John* is the only appropriate antecedent, and we would want to resolve the anaphora by identifying the pronoun with John. Sometimes, there is no appropriate antecedent in the text, as in (4b), so an antecedent must be accommodated. If the anaphoric trigger is a pronoun, whose informational content is minimal, accommodation may be impossible (van der Sandt, 1992). In this case, the pronoun will be interpreted deictically:

(5) John was eating ice cream. *She* brought him the check.

In (5) we interpret the pronoun as referring to some extralinguistic individual, identified, perhaps, by pointing.

Sometimes there may be more than one good candidate antecedent, as in (4c): there is very little reason to prefer either *John* or *Bill* as the antecedent of the pronoun. In this case, we have two choices: we can decide on some antecedent, perhaps at random, perhaps using some criterion such as recency; or, we can acknowledge that the anaphora is genuinely ambiguous. In the latter case, all is not lost, however: although we do not know which is the antecedent, we can still draw some conclusions about him. For example, we know that, whoever he is, he was at the ice cream parlor.

Before formalizing DOAP, we need to say something about how the coreference relation between trigger and antecedent is represented. A particularly attractive choice for our purposes is the representation provided by Discourse Representation Theory (Kamp and Reyle, 1993). In this system, trigger and antecedent are related simply through equality, perhaps the most heavily studied relation in the history of logic. So, by defining coreference to be equality, we get various desirable results for free: for example, by the axioms of equality, if the antecedent has a certain property, then the trigger immediately has this property too.

In this paper I propose a simple formalization of DOAP using Default Logic (Reiter, 1980b). The idea is that a pronoun and a potential antecedent are equated by default, unless this is superseded by some rule. This default rule is given low priority, so that other factors affecting anaphora resolution can rule out inadmissible antecedents, or suggest an antecedent before DOAP applies. The result it that this principle would apply only if there is no strong preference for any antecedent; but when it does, the behavior of the resulting system complies with the desiderata described above.

The rest of the paper is organized as follows. The next section contains a discussion of Default Logic, which is used to formalize DOAP. Section 3 is a description of the proposed formalization of DOAP, dubbed Equality by Default. Section 4 discusses the inferences that can be drawn using this relation, and contains examples demonstrating that they obey the desired patterns. The final section is a brief conclusion.

2 Default Logic

Since the relation between trigger and antecedent is equality, the problem of anaphora resolution becomes the problem of inferring the appropriate

equalities from the representation of the discourse. This inference must be nonmonotonic: as examples like (1) above demonstrate, we may later find that our choice of antecedent was wrong, and revise it. To formalize this inference, I will use Default Logic (Reiter, 1980b).

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

$$(6)\quad \frac{\alpha(x) : \beta_1(x), \ldots, \beta_m(x)}{\gamma(x)},$$

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 $x = x_1, \ldots, x_n$.

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; a well known example is the *Nixon diamond*. Suppose we have the following two defaults:

1. $\dfrac{\mathbf{Quaker}(x) : \mathbf{pacifist}(x)}{\mathbf{pacifist}(x)}$

2. $\dfrac{\mathbf{Republican}(x) : \neg\mathbf{pacifist}(x)}{\neg\mathbf{pacifist}(x)}$

If Nixon is both a Quaker and a Republican, in one extension he will be a 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.

Some formalizations of Default Logic include ranking of defaults. Intuitively, this means that if default d_1 outranks default d_2, then it applies first, in the sense that there is no extension of the default theory that contains the conclusion of d_2 but not the conclusion of d_1, if both are applicable.

While ranking is a very useful device, and we will use it too, it is important to emphasize that it doesn't add to the formal power of the system: for every ranked default theory, an equivalent unranked default theory can be constructed (Delgrande and Schaub, 2000).

The semantics of Default Logic can be provided by *Herbrand models* (Lifschitz, 1990; Kaminski, 1995). 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-model* is a set of closed atomic formulas of $\mathcal{L}_b$.

3 Equality by Default

Resolving anaphora means generating an equality between two discourse referents. I propose generating this 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 Charniak (1988). Charniak's approach is further explored in Cohen and Makowsky (1993), and formalized more fully in Cohen, Kaminski, and Makowsky (2005; 2006), in which its potential for anaphora resolution is noted.

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

(7) $$\frac{: x = y}{x = y}.$$

This default rule means that whenever it is consistent to assume that two elements are the same, we conclude that they are.

It might be objected that this is rather too liberal an assumption of equality, and that we allow two many elements to be equal by default. This, however, is not the case. Equality by default does not apply in isolation; any reasonable system drawing inferences from natural language will require many more defaults, some of which deal specifically with anaphora, some don't. If we give Equality by Default a low priority, so that, if other defaults can apply, they will, inappropriate equalities will be ruled out, and rather few equalities will hold. For example:

(8) a. John saw Bill. *He* greeted *him*.

b. John hates *him*.

c. John doesn't have a car. *It* is red.

d. A man came into the ice cream parlor. *She* was upset.

The most likely interpretation of (8a) is that the first pronoun refers to John, and the second one to Bill. This interpretation is brought about by a default rule that prefers antecedents that share the grammatical position of the pronoun (parallelism). In general, Equality by Default is a principle of *last resort*: it will not be invoked if other rules suggest some antecedent.

Sentence (8b) does not have an interpretation where *him* is equated with John, for syntactic reasons. In (8c), the pronoun *it* should not be equated with the discourse referent representing the indefinite *a car*, because, according to the rules of DRT, the indefinite is not accessible to the pronoun. The discourse in (8d) is an example where the pronoun cannot be associated with the antecedent because of gender mismatch: the antecedent is male, while the pronoun must refer to a female individual. If all such constraints are represented as defaults (or as exceptionless rules) that have a higher priority than Equality by Default inadmissible antecedents will be ruled out.

Note that in the definition of Equality by Default there is no requirement that one of the discourse referents be an anaphoric trigger, and the other—a potential antecedent, so it might be thought that this would generate spurious equalities between arbitrary discourse referents. While it is easy to add such a restriction, this is not really necessary. The reason is, again, that Equality by Default is only one of a number of linguistic constraints. For example:

(9) a. John talked to Bill.

b. An officer talked to a gentleman.

c. John is meeting a woman tonight. His mother told me so.

d. John went to the clinic. The doctor has a busy day.

In (9a) John cannot be equated with Bill—this is known as the Unique Names Assumption (Reiter, 1980a). Our system cannot have the Unique Names Assumption, because different terms are assumed to be equal by default, but it can mimic its effects. In DRT, names get their reference by *anchoring* them to individuals in the model, rather than by equality. If the names John and Bill are anchored to different individuals, with different properties, then they must be different and cannot be equal by default.

In (9b), *a gentleman* cannot be equated with *an officer*, because indefinites are novel (Heim, 1982), hence must be different from any previously introduced discourse referent.

In sentence (9c), *his mother* does not refer back to *a woman*. The reason is due to conversational implicature (Grice, 1975): a speaker who knows that John is meeting his mother should say so, hence we conclude that the woman is someone else.

Sentence (9d) is an example of bridging: *the doctor* is the one working at the clinic. The sentence is ambiguous between a reading where the doctor is equated with John, and another, perhaps more plausible reading, where John is a patient rather than a doctor. This reading is obtained because the notion of a clinic also introduces the notions of patients, together with the restriction that the patients are different from the doctor. According to one default conclusion, John is equated with the doctor, but according to another, he is equated with one of the patients, and is different from the doctor. Clearly, the two default conclusions are incompatible, hence we will have two extensions, one for each reading.[1]

Thus, although the assumption of equality by default appears very permissive, in fact it allows rather few elements to be equal by default. These are intended to be anaphoric triggers and their potential antecedents, when no antecedent is suggested by an anaphora resolution factor.

Like other default theories, Herbrand models can provide a semantics for Equality by Default (Cohen, Kaminski, and Makowsky, 2006). A clarification, however, is in order. Since the Herbrand universe of a language $\mathcal{L}_b$ is the set of all closed terms of $\mathcal{L}_b$, then, by definition, in a Herbrand model no two terms are identical. But in our default theory, two terms are equal by default. Is this a contradiction? The answer is no. Equality is any relation that satisfies the equality axioms, and is not necessarily identity.[2] Hence, there is no problem about two terms being *equal*, even though they are not *identical*.

4 Unresolved Anaphora

As mentioned above, there are two cases when anaphora may remain unresolved: when there is no appropriate antecedent, or when there is more than one. In the first case, the trigger needs to be interpreted as referring to an entity not provided by the linguistic content (i.e. an accommodated or deictic interpretation). In the second case, the anaphora is truly ambiguous, and this ambiguity needs to be either resolved arbitrarily, or left unresolved,

[1] For more on the treatment of ambiguity, see below.

[2] Of course, we can a have a non-Herbrand model where equality *is* identity—such models are called *normal*, see Mendelson (1997, p. 100) for details.

drawing as many inferences as possible.

4.1 No potential antecedent

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 equal by default to any anaphoric trigger; this is how accommodated and deictic readings are possible.

This theory allows accommodated and deictic readings, but only as a last resort, when no other readings are possible. More precisely, it has been shown (Cohen, Kaminski and Makowsky, 2006) that if E is an extension for Equality by Default, 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

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

Now, consider a model w of extension E where trigger u is accommodated or interpreted deictically. This means that, in w, for every x_i, a potential antecedent to u, $u \neq x_i$; and for some new element $n \in b$, $u = n$. Since w is minimal, there is no Herbrand b-model w' of E that contains all the equalities in w and adds to them. Therefore, it is not only in w, but in all models of E, that u is different from all its potential antecedents.

What this means is that there is at least one extension, i.e. at least one plausible way of reasoning from the known facts, that is inconsistent with an anaphoric reading of u. Hence, accommodated or deictic readings are only available when the anaphoric reading is implausible (or impossible).

4.2 Multiple potential antecedents

Suppose we have two acceptable antecedents for some trigger. For example, in (4c), repeated below, the pronoun may be equated with *John* or with *Bill*.

(11) John and Bill met at the ice cream parlor. *He* was upset.

If we make the standard assumption, as described above, that different names are anchored to different individuals, we know that *John* is different from *Bill*, so it is impossible to believe that the pronoun is equal to both. 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? We may decide to force a decision, which means applying credulous reasoning and picking one extension.

Note that, by the axioms of equality, once such a choice is made, any property of the antecedent becomes also a property of the trigger. For example, if we choose the extension where u, the pronoun, is equated with Bill, and if Bill is bald, it will immediately follow that u is bald.

There are cases, however, where the anaphora is genuinely ambiguous, and we may have no reason to prefer one reading over the other. But even if we decide not to resolve the anaphora, there are still inferences we can draw. In this case, it makes sense to apply skeptical reasoning, and accept only what is true in all candidate extensions.

Recall that we want to infer, even without resolving the anaphora, that whoever the pronoun refers to was at the ice cream parlor. Skeptical reasoning will, indeed, give us precisely this result, since in both extensions, the antecedent of the pronoun has this property.

But now suppose that one possible antecedent has a property than the other one lacks, or is not known to have:

(12) a. John walked along the sidewalk and saw that Bill was inside the ice cream parlor. *He* was upset.

b. John saw that Bill was eating ice cream. *He* was upset.

In (12a), Bill is inside the ice cream parlor, but John is outside. Thus, in one extension, *he* will have the property of being inside the ice cream parlor, and in the other—its negation. If we apply skeptical reasoning, we will be able to conclude nothing—this appears intuitively correct.

In (12b), we know that Bill was eating ice cream, but we do not know whether John was. Intuitively, we cannot conclude that *he* was eating ice cream, although this is consistent with the pronoun being equated with either John or Bill. Skeptical reasoning predicts this result: while in one extension the property of eating ice cream is predicated of the antecedent of the pronoun, in the other extension, neither this property nor its negation will be so predicated. Therefore, it is not true in all extensions that *he* was eating ice cream.

5 Conclusion

The proposed system provides a Default Logic formalization of DOAP, which appears to produce the intuitively correct results. It is ensured that if it is possible to find an antecedent, we do so; if more than one is a good candidate,

we can use Default Logic techniques for dealing with multiple extensions; and if there is none, we accommodate or interpret the pronoun deictically. Thus, the way in which the system will resolve a given case of anaphora, and the inferences it would draw from such a resolution, can be proved explicitly, rather than being left to be implicit in the workings of the algorithm.

References

[Asher, 1984] Asher, N. (1984). Linguistic understanding and nonmonotonic reasoning. In *Proceedings of the 1st International Workshop on Nonmonotonic Reasoning*, New Paltz.

[Bry and Yahya, 2000] Bry, F. and Yahya, A. (2000). Positive unit hyperresolution tableaux and their application to minimal model generation. *Journal of Automated Reasoning*, 25:35–82.

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

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

[Cohen et al., 2006] Cohen, A., Kaminski, M., and Makowsky, J. A. (2006). Notions of sameness by default and their application to anaphora, vagueness, and uncertain reasoning. ms., Ben-Gurion University and The Technion.

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

[Delgrande and Schaub, 2000] Delgrande, J. P. and Schaub, T. (2000). Expressing preferences in Default Logic. *Artificial Intelligence*, 123:41–87.

[Grice, 1975] Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J. L., editors, *Syntax and Semantics 3: Speech Acts*. Academic Press.

[Heim, 1982] Heim, I. (1982). *The Semantics of Definite and Indefinite NPs*. PhD thesis, University of Massachusetts at Amherst.

[Hendriks and de Hoop, 2001] Hendriks, P. and de Hoop, H. (2001). Optimality theoretic semantics. *Linguistics and Philosophy*, 24:1–32.

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

[Kamp and Reyle, 1993] Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht.

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

[Lifschitz, 1990] Lifschitz, V. (1990). On open defaults. In Lloyd, J., editor, *Computational Logic: Symposium Proceedings*, pages 80–95, Berlin. Springer–Verlag.

[Mendelson, 1997] Mendelson, E. (1997). *Introduction to mathematical logic*. Chapman and Hall, London.

[Mitkov, 1995] Mitkov, R. (1995). An uncertainty reasoning approach for anaphora resolution. In *Proceedings of the Natural Language Processing Pacific Rim Symposium (NLPRS'95)*, pages 149–154, Seoul, Korea.

[Poesio, 1996] Poesio, M. (1996). Semantic ambiguity and perceived ambiguity. In van Deemter, K. and Peters, S., editors, *Semantic Ambiguity and Underspecification*, pages 159–201. CSLI, Stanford.

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

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

[van der Sandt, 1992] van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377.

[Williams, 1997] Williams, E. (1997). Blocking and anaphora. *Linguistic Inquiry*, 28:577–628.