Making Sense of Reference to the Unfamiliar

Helen Seville and Allan Ramsay*
Centre for Computational Linguistics
UMIST, PO Box 88, Manchester M60 1QD, England
heleng/allan@ccl.umist.ac.uk

Abstract
Computational approaches to reference resolution, like Centering Theory, are best at resolving referring expressions which denote familiar referents. We demonstrate how, by taking a proof-theoretic approach to reference resolution within a Centering-type framework, we are able to make sense of referring expressions for unfamiliar referents. These include, in addition to bridging descriptions, definite descriptions like “the first man” and “the first snowdrops of Spring”. We claim that the first of these denotes a unique subset of a plural discourse antecedent. While the second has no discourse antecedent, we similarly treat it as denoting a unique subset of a familiar referent.

1. Introduction
How do referring expressions denote? According to Russell, a definite description such as “the King of France”, denotes a unique individual by virtue of its meaning. But, according to Familiarity Theory (Heim, 1983), referring expressions need not denote uniquely by virtue of their meaning as they refer to individuals made familiar by the discourse or other context. This observation plays a key role in Centering Theory (Grosz and Sidner, 1986; Grosz et al., 1995) and other computational approaches in which referring expressions are resolved by locating their antecedents in the discourse. The reference of pronouns like “he”, definite descriptions like “the woman”, and referential tenses like “had” clearly has more to do with salience in context than with uniqueness of meaning. Similarly, while names like “Mary” need not denote individuals prominent in the discourse context, they must nevertheless denote individuals familiar to conversants if they are successfully to refer. However, there is another class of referring expressions in relation to which we believe the concept of uniqueness of meaning does have an essential role to play. These include such definite descriptions as “the first man” and “the first snowdrop of Spring”, along with such variations on these as “the first three men” and “the first snowdrops of Spring”.

In implementing a system of reference resolution, we have attempted to reconcile the notions of familiarity and uniqueness. This enables us to dereference expressions like “the first snowdrop of Spring” in a unified framework alongside anaphors\(^1\), pronouns, referential tenses, names, and other definite descriptions like “the man”.

\(^1\)We use this term to distinguish reflexives like “herself” from pronouns like “he” and “him”.

(1) Two men arrived.
(2) The first man spoke.

In the case of a referring expression like “the first man”, there may be an antecedent of sorts in the discourse, but it is not the individual referred to (or indeed an individual at all). We will say that the antecedent “two men” introduces a set, and that the referring expression “the first man” denotes, by virtue of the meaning of first, a unique subset of this familiar set.

(1) Mary saw the first snowdrop of Spring.

In the case of “the first snowdrop of Spring”, there need be no explicit antecedent in the discourse. We will say that, in the same way that “Mary” denotes a familiar individual, “the snowdrops of Spring” denotes a familiar set, or
property. Again, by virtue of the meaning of first, “the first snowdrop of Spring” can be said to denote a unique subset of the familiar set. We will not claim that it denotes a unique individual, but that rather it denotes a unique subset of the specified cardinality, i.e., 1. This treatment has the advantage that it extends to plural referring expressions.

Below we outline the approach we have developed to the representation and resolution of referring expressions, before discussing in more detail its extension to deal with unfamiliar referents.

2 A Framework for Reference Resolution

Our framework for reference resolution has been implemented in the system of language understanding described in (Ramsay, 1999). The starting point for reference resolution is the logical form we obtain from parsing. For example, the following is the logical form we get for the utterance “Mary slept.”

$$\exists A : \{ A \text{ is interval } \& \text{ ends before} (\text{ref}(\lambda B(\text{speech.time}(B, 1)), A)) \}$$

$$\exists C : \{ \text{aspect(simple, } A, C) \}$$

$$\theta(C, \text{agent, ref}(\lambda D(\text{named}(D, Mary) \& \text{card}(D, 1))))$$

$$\& \text{ sleep}(C)$$

$$\& C \text{ is event}$$

We use the inference engine described in (Ramsay and Seville, 2000) to update the discourse model with a new discourse state containing the information explicitly represented in the logical form we obtain from parsing. For example, the following is the logical form we get for the utterance “Mary slept.”

Given that many referring expressions do not in themselves denote uniquely, however, we need a theory of reference resolution to enable us to obtain the appropriate (i.e., intended) referent for any referring expression. We incorporate our theory of reference resolution into the actual representation of referring expressions; for example, we label anaphors with the property “salient” and pronouns (and also referential tenses) with the property “centred”:

“himself”

$$\text{ref}(\lambda X(\text{salient}(X, \text{ref}(\lambda \text{cds}(D))) \& m(X)))$$

“she”

$$\text{ref}(\lambda X(\text{centred}(X, \text{ref}(\lambda \text{cds}(D))) \& f(X)))$$

Reference resolution relies on maintaining, as in Centering Theory, a list of forward-looking centres for each discourse state (corresponding to an utterance) in the discourse. Furthermore, for the purposes of reference resolution, the discourse states themselves are organized into a discourse tree, which is constructed automatically based on referential cues, as described in (Seville, 1999).

The nodes in such a tree correspond to discourse states. Those on the right-hand frontier are open, which essentially means that the entities mentioned in them are available to pronominal reference.

The process of reference resolution for the various referring expressions can be briefly described as follows. Anaphors, characterised as salient, are resolved to a less oblique argument of the same verb (Pollard and Sag, 1994) within the current discourse state, which is constructed

---

2Strictly speaking, it is a set which is denoted. For readability, our referring expressions confine the properties of sets and their members. In this case, the cardinality is a property of the set denoted, but the name Mary is a property of its member.

3Here ref(\lambda D(\text{cds}(D))) is a reference to the current discourse state and the properties m and f refer to male and female gender respectively.

4The tree illustrated was constructed using pronominal cues. Each discourse state was attached as a daughter of the highest node in the discourse tree to which all pronouns and referential tenses (like had) mentioned in it could be anchored.
incrementally. We also start our search for the referents of pronouns and other centred entities in the current discourse state, which is necessary if we are to resolve such referring expressions as "her" in "Mary took John with her." However, referring expressions containing the property centred are prevented from being dereferenced to salient entities, thus ensuring that the constraint of disjoint reference is met. If we fail to find the centred entity in the current discourse state, we search the previous open node and, if necessary, further open nodes in the discourse tree, in order to deal with long-distance pronominalisation. The dereferencing of other referring expressions like \(\text{ref}(\lambda D(\text{named}(D, \text{Mary}) \& \text{card}(D, 1)))\) is similar but less constrained in that we consider entities mentioned in all nodes mentioned in the discourse tree, whether open or closed, in order of recency. This means that, essentially, names and definite descriptions are dereferenced to the most recently mentioned referent which is appropriate. Unlike in the case of pronouns, we also consider Discourse State 0, which doesn’t correspond to an utterance but, rather, contains the background knowledge assumed in the model. This is how we are able to deal with the first mention of a familiar referent like Mary (assuming that the properties \(\lambda D(\text{named}(D, \text{Mary}) \& \text{card}(D, 1))\) suffice to distinguish a particular entity in Discourse State 0 from all the others).

Our approach extends naturally to cases like "the first snowdrop of Spring" because it is proof-theoretic and so able to exploit background knowledge in reference resolution. This can be illustrated, in the first instance, by examining the background knowledge which is used in updating the utterance "Mary slept." The update step for this utterance yields Discourse State 1, containing (amongst others) the following facts:

**Discourse State 1**

\[
\begin{align*}
\text{sleep}(& \#134) \\
\theta(& \#134, \text{agent}, \#94) \\
\text{ends} \_ \text{before}(& \#4(1), \#133) \\
\text{aspect}(\text{simple}, & \#133, \#134)
\end{align*}
\]

We were able to prove \(\text{named}(\#94, \text{Mary})\) and \(\text{card}(\#94, 1)\) and so dereference \(\text{ref}(\lambda D(\text{named}(D, \text{Mary}) \& \text{card}(D, 1)))\) as the following were amongst the facts contained in Discourse State 0:

**Discourse State 0**

\[
\begin{align*}
\text{female}(\#94) \\
\text{named}(\#94, \text{Mary}) \\
\text{woman}(& \#94) \\
\text{f}(& \#94) \\
\text{card}(\#94, 1) \\
\text{adult}(& \#94)
\end{align*}
\]

These were generated from the lexical meaning postulates we stipulated for "Mary", "woman", and "female":

\[
\begin{align*}
3X(\text{named}(X, \text{Mary}) \& \text{woman}(X) \& \text{card}(X, 1)) \\
\forall X(\text{woman}(X) \rightarrow \\
\quad \text{female}(X) \& (X \text{ is human} \& \text{adult}(X)) \\
\forall X(\text{female}(X) \rightarrow f(X))
\end{align*}
\]

3 Unfamiliar Referents

In this section we show how, within the framework above, we are able to make sense of a variety of referring expressions denoting unfamiliar referents. The most straightforward of these are bridging descriptions, so we start with these.

3.1 Bridging Descriptions

(1) Mary loves her mother.

In this first case, "her mother", contains a referring expression nested within it. Having dereferenced this, the knowledge that mother of is a function enables us to obtain a unique referent. Our representation of the referring expression to be dereferenced is as follows:

\[\text{"her mother"} \quad \text{ref}(\lambda B(\text{of}(B, \\
\lambda F(\text{mother}(F)), \\
\text{ref}(\lambda G(\text{salient} \_ \text{or} \_ \text{centred}(G, 1) \\
\& \ f(G)))))) \\
\& \ \text{card}(B, 1))\]

The first step involves anchoring the referring expression by dereferencing its nested referring expression for "her".\(^5\)

\(^5\)The referent for this is characterised as \(\text{salient} \_ \text{or} \_ \text{centred}\) as we allow possessive pronouns to be dereferenced as anaphors or, failing that, as pronouns.
Current Model

\begin{align*}
\text{ends\_at\_or\_after}(&\#4(1), \#135) \\
\text{aspect(simple,} &\#135, \#136) \\
\theta(&\#136, \text{agent,} \#94) \\
\text{love(} &\#136)
\end{align*}

The partially constructed current discourse state we have when we do our dereferencing is as shown. "Mary" has already been dereferenced to \#94 and this has been entered into the list of forward-looking centres for the current utterance. We are able to prove both salient(\#94) and f(\#94), and so our nested referring expression is dereferenced to this entity.

\[\text{ref} (\lambda B (\text{of}(B, \\
\lambda F(\text{mother}(F)), \\
\#94) \\
\& \text{card}(B, 1)))\]

It is then a straightforward matter to dereference the anchored referring expression, given the following facts in Discourse State 0:

Discourse State 0

\begin{align*}
\text{mother}(\#60(\#94)) \\
\text{of}(\#60(\#94), \lambda A(\text{mother}(A)), \#94) \\
f(\#60(\#94)) \\
\text{card}(\#60(\#94), 1)
\end{align*}

These derive from our meaning postulates for "mother" and "of":

\[\forall X ((X \text{ is animal}) \& \text{card}(X, 1) \rightarrow \exists Y (\text{of}(Y, \lambda Z(\text{mother}(Z)), X) \\
\& \text{card}(Y, 1) \& f(Y))) \]

\[\forall X \forall Y \forall Z (\text{of}(X, Y, Z) \rightarrow Y \cdot X)\]

Dealing with other bridging descriptions is more complicated:

1. Mary saw a house.
2. She found the door.

In order to give an analogous treatment to the referring expression "the door", we have to treat it as elliptical for an expression containing a nested referring expression, i.e., "the door of the house". In the same way that we have a meaning postulate for the relation mother of, we have one for the relation door of:

\[\forall X ((\text{house}(X) \lor \text{card}(X)) \rightarrow \exists Y (\text{of}(Y, \lambda Z(\text{door}(Z)), X) \& \text{card}(Y, 1)))\]

This means that, having used utterance (1) above to update the discourse model, we have the following amongst the facts in Discourse State 1:

Discourse State 1

\begin{align*}
\text{see}(\#138) \\
\theta(\#138, \text{agent,} \#94) \\
\theta(\#138, \text{object,} \#139) \\
\text{card}(\#139, 1) \\
\text{house}(\#139) \\
\text{ends\_before}(\#4(1), \#137) \\
\text{door}(\#46(\#139)) \\
\text{entrance}(\#46(\#139)) \\
\text{of}(\#46(\#139), \lambda A(\text{door}(A)), \#139) \\
\text{card}(\#46(\#139), 1) \\
\text{aspect(simple,} &\#137, \#138)
\end{align*}

In updating utterance (2), the bridging description which needs to be dereferenced has the following representation:

\[\text{ref} (\lambda E (\text{door}(E) \& \text{card}(E, 1)))\]

Since we cannot guarantee that there will only be a single entity in our model satisfying the properties \(\lambda E (\text{door}(E) \& \text{card}(E, 1))\), we want to ensure that the referent we obtain is either the most recently mentioned or that with the most recently mentioned antecedent, i.e., in this case, the house \#139. Our reference resolution procedure exploits the fact that the house, \#139, is explicitly represented in the forward looking centres of Discourse State 1 and that the intended referent, \#46(\#139), is clearly a function of this (its dependency having been preserved by Skolemization). In considering the potential referents for our referring expression in order of recency, we attempt to prove, not simply, for each referent, X, whether \text{door}(X) and \text{card}(X, 1), but \text{door}(Y) and \text{card}(Y, 1) where \text{Y} is a function of \text{X}. Since \#46(\#139) is a function of the antecedent \#139, we obtain the appropriate referent in this case by proving \text{door}(\#46(\#139)) and \text{card}(\#46(\#139), 1).

3.2 Superlatives

We are now in a position to describe our treatment of the superlatives discussed in the introduction. First, we consider a case in which there is a discourse antecedent of sorts:
(1) Two men arrived.
(2) The first man spoke.

Discourse State 1 contains the following facts:

Discourse State 1

\begin{align*}
\text{arrive}(&\#107) \\
\theta(&\#107, \text{agent}, \#108) \\
\text{card}(\#108, 2) \\
\text{man}(\#108) \\
\text{male}(\#108) \\
\text{m}(\#108) \\
\text{adult}(\#108) \\
\text{ends before}(\#4(1), \#106) \\
\text{aspect}(\text{simple}, \#106, \#107)
\end{align*}

Our representation of the referring expression “the first man” is as follows:

\[
\text{rcf}(\lambda B(\text{most}(B, \\
\lambda C(\text{early}(C, \lambda D(\text{man}(D))))) \\
\#108) \\
\& \text{card}(B, 1)))
\]

The nested referring expression \(\text{rcf}(\lambda E(\text{man}(E)))\) can be straightforwardly dereferenced in this case to give the anchored referring expression:

\[
\text{rcf}(\lambda B(\text{most}(B, \\
\lambda C(\text{early}(C, \lambda D(\text{man}(D))))) \\
\#108) \\
\& \text{card}(B, 1)))
\]

Dereferencing this then involves our meaning postulate for superlatives:

\[
\forall X \forall Y \forall P \exists C (\text{ordered}(P) \land \text{most}(X, P, X) \rightarrow P. Y) \\
\forall A (\text{ordered}(X. \text{early}(B, A))) \\
\forall X \forall P (\text{early}(X, P) \rightarrow P. X) \\
\forall X (\text{man}(X) \rightarrow \text{human}(X) \land \text{male}(X) \land \text{adult}(X)) \\
\forall X (\text{male}(X) \rightarrow \text{m}(X))
\]

In this way, we prove that the referring expression makes sense, i.e., denotes. However, unlike in the previous cases, we do not reference to a familiar referent. There are no existing facts in the database about the referent \#81(\text{AA}(\text{early}(\text{A}, \ldots)), 1, 2, \#108, \#108). Instead, in this case, we have to add to Discourse State 2 the facts we have proved.

Discourse state 2

\begin{align*}
\text{speech}(\#112) \\
\text{theta}(\#112, \text{agent}, \#108) \\
\text{card}(\#108, 1, 2, \#108, \#108) \\
\text{aspect}(\text{simple}, \#106, \#107) \\
\text{most}(\#81(\text{AA}(\text{early}(\text{A}, \ldots)), 1, 2, \#108, \#108), \\
\lambda C(\text{early}(C, \lambda D(\text{man}(D))))), \\
\#108) \\
\text{card}(\#81(\text{AA}(\text{early}(\text{A}, \ldots)), 1, 2, \#108, \#108), \\
1)
\end{align*}

The further facts we prove, about our referent being early, male, etc., are required if we are to be able to subsequently refer to it using referring expressions such as “he”. These are generated from a set of associated meaning postulates:

\[
\forall X \forall Y \forall P ((\text{ordered}(P) \land \text{most}(Y, P, X) \rightarrow P. Y) \\
\forall A (\text{ordered}(X. \text{early}(B, A))) \\
\forall X \forall P (\text{early}(X, P) \rightarrow P. X) \\
\forall X (\text{man}(X) \rightarrow \text{human}(X) \land \text{male}(X) \land \text{adult}(X)) \\
\forall X (\text{male}(X) \rightarrow \text{m}(X))
\]

In addition to these, we have two further meaning postulates for superlatives:

\[
\forall X \forall Y \forall V \forall Z \forall C (\text{most}(Y, P, X) \land \text{card}(Y, C) \\
\& \text{most}(Z, P, X) \land \text{card}(Z, C) \\
\rightarrow \text{card}(X, N) \land \text{card}(Y, C) \\
\rightarrow \neg \text{more}(N, C))
\]
The first of these, the uniqueness meaning postulate, states that if there are two subsets of a set which share the same cardinality and the same superlative property, such as first, then they must be regarded as identical. The second simply ensures that any unfamiliar referent which we obtain via our meaning postulates can sensibly regarded as a proper subset of its antecedent; that is, it prevents us regarding “two men” as a potential antecedent of “the first men”:

1. Two men arrived.
2. The first men spoke.

Our treatment of superlatives without discourse antecedents is similar to that above.

1. Mary saw the first snowdrops of Spring.

There is just one major difference.

\[
\text{ref} (\lambda E (\text{most}(E, \\
\lambda F (\text{early}(F, \\
\lambda G (\text{of}(G, \\
\lambda H (\text{snowdrop}(H))), \\
\text{ref}(\lambda I (\text{named}(I, \\
\text{Spring}))))))))) \\
\text{ref}(\lambda J (\text{of}(J, \\
\lambda K (\text{snowdrop}(K))), \\
\text{ref}(\lambda L (\text{named}(L, \\
\text{Spring})))))) \\
& \text{card}(E, pl))
\]

The representation we obtain for the referring expression “the first snowdrops of Spring” is shown above. Like that for “the first man”, this contains a nested referring expression:

\[
\text{ref} (\lambda J (\text{of}(J, \\
\lambda K (\text{snowdrop}(K))), \\
\text{ref}(\lambda L (\text{named}(L, \\
\text{Spring})))))) \\
& \text{card}(E, pl))
\]

The difference is that, in this case, there is no discourse antecedent for the nested referring expression. This means that, in order to anchor our referring expression by dereferencing the referring expression nested within it, we need to introduce a meaning postulate for the nested referent (and one for its nested referent, Spring):

\[
\exists X (\text{season}(X) & \text{named}(X, \text{Spring}) & \text{card}(X, 1)) \\
\exists Y (\text{snowdrop}(Y), \\
\text{ref}(\lambda Z (\text{named}(Z, \text{Spring})))) \\
& \text{card}(X, pl))
\]

These meaning postulates simply introduce into Discourse State 0 the fact that there are snowdrops of Spring, in the same way that the meaning postulate for “Mary” introduced the fact that there is a singleton set containing an individual so named.

Discourse state 0

\[
\text{season}(\#98) \\
\text{named}(\#98, \text{Spring}) \\
\text{card}(\#98, 1) \\
\text{extended}(\#98)
\]

\[
\text{snowdrop}(\#101) \\
\text{of}(\#101, \lambda A, \text{snowdrop}(A), \#98) \\
\text{n}(\#101) \\
\text{card}(\#101, pl)
\]

Given the above facts in Discourse State 0, anchoring our referring expression is straightforward.

\[
\text{ref}(\lambda E (\text{most}(E, \\
\lambda F (\text{early}(F, \\
\lambda G (\text{of}(G, \\
\lambda H (\text{snowdrop}(H))), \\
\text{ref}(\lambda Z (\text{named}(Z, \text{Spring})))))))) \\
& \#101) \\
& \text{card}(E, pl))
\]

From this point onwards, the proof that this referring expression denotes proceeds in the same way as in the previous example. Given the meaning postulates for superlatives, we are able to prove:

\[
\text{most}(\#81(\lambda A(\text{early}(A, \ldots)), pl, pl, \#101, \#101), \\
\lambda D (\text{early}(D, \\
\lambda E (\text{of}(E, \lambda F (\text{snowdrop}(F)), \#98))))) \\
& \#101) \\
\text{card}(\#81(\lambda A(\text{early}(A, \ldots)), pl, pl, \#101, \#101, pl))
\]
Again, as in the example above, the facts we have proved concern an unfamiliar referent, and so have to be added to the current discourse state.

Discourse state 1

\[ soc_1(#107) \]
\[ theta(#107, agent, #94) \]
\[ theta(#107, object, #81(), (A, ..., #98)), pl, pl, #101, #101) \]
\[ ends\_before(#4(1), #106) \]
\[ aspect(simple, #106, #107) \]
\[ most(#81(AA(...), #98))), pl, pl, #101, #101), \lambda D(early(D, \lambda E(of(E, \lambda F(snowdrop(F)), #98))), #101) \]
\[ card(#81(AA(...), #98))), pl, pl, #101, #101), \lambda D(of(D, \lambda E(snowdrop(E)), #98)) \]
\[ of(#81(lambda(A, ..., #98))), pl, pl, #101, #101), \lambda D(snowdrop(D)), #98) \]
\[ snowdrop(#81(lambda(A, ..., #98))), \ldots)) \]
\[ n(#81(lambda(A, ..., #98))), pl, pl, #101, #101) \]

4 Conclusion

We have shown how, by taking a proof-theoretic approach to reference resolution, we can extend a Centering-type framework to make sense of referring expressions for a variety of unfamiliar referents. Having made sense of such referring expressions, we add their referents to our discourse model. This is how we would normally deal with indefinites rather than definites. However, this approach makes perfect sense, given our treatment of such referring expressions as denoting unfamiliar subsets of familiar referents (regarded as sets). We claim that we are able to use definite descriptions to refer to the referents in question, despite their unfamiliarity, so long as we can prove that, by virtue of their meaning, they denote uniquely.

Having implemented our approach in a system of language understanding which already deals with a wide variety of referring expressions, we have demonstrated its practicality. It also has interesting theoretical implications, since it suggests a way in which pragmatic theories of reference resolution, like Familiarity Theory, and semantic theories, like Russell's, may be reconciled. However, it is fair to say that the success of the approach is not yet proven. This is because we have yet to show that we can deal with a set of related referring expressions within a single framework. The following example illustrates the kinds of cases we have in mind:

(1) Three men ate.
(2) Two men slept.
(3) The first men died.

Here, "first" in "the first men" is clearly performing a different, discourse-related function from that it plays in the cases we have been considering. We have yet to tackle such difficult cases but, since they seem to require reasoning about sets, we believe that our inference-based approach to reference resolution is a good place to start.

References

B. J. Grosz and C. L. Sidner. 1986. Attention, intentions, and the structure of discourse. Computational Linguistics, 12(3):175–204.
B. J. Grosz, A. K. Joshi, and S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. Computational Linguistics, 21(2):203–225.
I. Heim. 1983. File change semantics and the familiarity theory of definiteness. In R. Bauerle, C. Schwarz, and A. von Stechow, editors, Meaning, Use, and Interpretation of Language, pages 164–189. de Gruyter, Berlin.
C. Pollard and I. A. Sag. 1994. Head-Driven Phrase Structure Grammar. University of Chicago Press, London.
A. Ramsay and Helen Seville. 2000. Models and discourse models. Journal of Language and Computation, 1(2):159–174. forthcoming.
A. Ramsay. 1999. Does it make any sense? updating = consistency checking. In K. Turner, editor, The Semantics/Pragmatics Interface from Different Points of View. Elsevier Science B.V.
H. Seville. 1999. Experiments with discourse structure. In Third International Workshop on Computational Semantics, pages 233–246, Tilburg.