Comprehension of Coreferential Expressions

Peter C. Gordon
Department of Psychology
University of North Carolina
Chapel Hill, NC 27599-3270
pcg@email.unc.edu

Randall Hendrick
Department of Linguistics
University of North Carolina
Chapel Hill, NC 27599-3155
hendrick@email.unc.edu

Abstract

The ways in which the form of referring expressions interacts with the structure of language are reviewed. Evidence from a number of different methods – quantitatively analyzed judgments of acceptable coreference, reading time, and corpus frequency of different types of coreferential expressions – converges on a fairly simple description of patterns of coreference. A model is presented which integrates important aspects of Discourse Representation Theory and of Centering Theory in order to provide an account of how referential expressions are interpreted as part of constructing a discourse universe from a series of utterances.

1 Introduction

The study of coreference in generative linguistics has led to a very strong emphasis on how the hierarchical structure of sentences interacts with the form of referring expressions to constrain coreference (Chomsky, 1986; Reinhart, 1976). The resulting principles, embodied in the Binding Theory, provide rules that are of some use to researchers in natural-language processing because they provide information about disjoint reference – what an expression cannot refer to in certain circumstances. However, beyond that use theoretical work on the Binding Theory does not directly bear on questions central to language processing. Questions of how to resolve potentially ambiguous expression such as pronouns, or how meaning more generally is built up incrementally from linguistic expressions in context, are beyond its scope.

We (Gordon & Hendrick, 1997; in press) have used the basic methods of experimental psychology to take a close look at the phenomena of coreference and disjoint reference involving full expressions (names and descriptions) that have been cited in support of Principle C of the Binding Theory.[1] The results show that the interaction of form of referring expression and language structure is far simpler than it has been taken to be in the Binding Theory. Further, results on the judged acceptability of different configurations of referential expressions are consistent with the results of experiments that use reading time as an online measure of language comprehension (Gordon, Grosz & Gilliom, 1993; Gordon, Hendrick, Ledoux & Yang, 1999). Further, those results are consistent with the frequency of different types of coreferential configurations in corpora of naturally-occurring language (Ariel, 1994; Carden, 1982; van Hoek, 1997). The pattern of coreference that is observed is accounted for by a model (Gordon & Hendrick, 1998) that incorporates aspects of Centering Theory (Grosz, Joshi, & Weinstein, 1995) into Discourse Representation Theory (Kamp & Reyle, 1993).

2 Coreference and the Form of Expressions

Research in psycholinguistics, both ours and that of others, supports a fairly simple generalization concerning the ease of establishing coreference in sequences of different forms of referring expressions. Coreference is most easily established in name-pronoun sequences, less easily established in name-name sequences, and

[1] Principle C of the Binding Theory states that an r-expression cannot have a c-commanding antecedent. A constituent α is said to c-command another constituent β if the first branching node that dominates α dominates β as well.

least easily established in pronoun-name sequences. An example of these types of sequences taken from Gordon and Hendrick (1997) is shown below along with the proportion of naïve subjects (college students at the University of North Carolina) who judged that it was grammatically acceptable for the expressions in bold-face to refer to the same person.

John's roommates met **him** at the shop.	.94
John's roommates met **John** at the shop.	.37
His roommates met **John** at the shop.	.23

It should be noted that according to the Binding Theory coreference should be acceptable in all the sentences shown above; for both the name-name and pronoun-name sequence the second referring expression does not have a c-commanding antecedent and therefore should be free to corefer with the first referring expression. This is one example of consistent differences that we found between the accepted empirical generalizations of the Binding Theory and the quantitatively analyzed judgments of competent native speakers who were naïve to linguistic theory.

In Gordon and Hendrick (1997; 199x) we find that this pattern of relative acceptability between the different types of sequences is shown for categorical judgments, for ratings of grammaticality, for isolated sentences, for sentences in discourse context, and for different types of unreduced expressions including names, definite NPs and quantified NPs.[2] This pattern is also supported by on-line measures of reading time which show that under certain conditions sentences with repeated names are read more slowly than matched sentences with pronouns (Gordon, et al. 1993; Kennison & Gordon, 1997); similar patterns of this reading elevation are observed within sentences and between sentences (Gordon, Hendrick, Ledoux & Yang, 1999). The similarity of results that we observe with judgments of coreference and with on-line measures of reading time suggest that judgments of acceptable coreference reflect the ease in terms of mental processing with which a sentence can be understood. The similarity of results that we observe in the factors that influence the ease of establishing coreference within and between sentences suggests that the same mechanisms may be used for the two types of coreferential processing. The finding by Ariel (1994) that repeated-name coreference is far less common than pronominal coreference in a naturally occurring corpus suggests that our conjecture about the relative ease in terms of cognitive processes of establishing coreference for different types of sequences is reflected in how people use different forms of coreferential expressions.

For name-pronoun and name-name sequences, the pattern of acceptable coreference is modified by the structural prominence of the antecedent (first) referring expression. Here, we take structural prominence to mean the inverse of depth of syntactic embeddedness. An example, again from Gordon and Hendrick (1997), is shown below along with the proportion of subjects who judged that it was grammatically acceptable for the expressions in bold-face to refer to the same person.

John's roommates met **him** at the shop.	.94
John met **his** roommates at the shop.	.97
John's roommates met **John** at the shop.	.37
John met **John's** roommates at the shop.	.24

For name-pronoun sequences, coreference is more acceptable when the antecedent is more prominent (e.g., a subject) than when it is less prominent (a possessive), while for name-name sequences the opposite is true. Gordon and Hendrick (1997) found that this pattern holds for a number of manipulations of prominence: subject NP versus object NP, subject NP versus component NP of a pair of conjoined NPs, and subject NP versus NP in a relative clause. Gordon, et al. (1999) showed similar patterns in relative reading times for cases of both intersentential and intrasentential coreference.

Our studies of coreference have mostly focused on understanding contrasts in the comprehension of full and reduced referring expressions. The results that we have obtained in those comparisons are consistent with other theoretical and methodological approaches. Research on psychological heuristics for interpreting ambiguous pronouns has provided

[2] Our findings make English look more similar to the survey of crosslinguistic variation in pronoun-quantified NP sequences in Bresnan (1998).

support for a subject-assignment strategy, where a pronoun is preferentially interpreted as coreferential with the subject of the preceding clause (Crawley, Stevenson, & Kleinman, 1990; Fredericksen, 1981); in our framework, the subject of a sentence has the most structural prominence. Research on the success of algorithms for pronoun resolution (Lappin & Leass, 1994) shows that syntactic factors (such as being a subject, being a direct object, and not being contained within another noun phrase) contribute to the likelihood that an expression is the antecedent of a subsequent pronoun.

3 DPT: An Account of Basic Coreference

We have developed a model of coreference called Discourse Prominence Theory or DPT (Gordon & Hendrick, 1998). It adapts the formalism provided by Kamp and Reyle's (1993) Discourse Representation Theory (DRT). In this approach, construction rules (CRs) map linguistic expressions onto universes of discourse. In DPT the construction rules and representation of DRT are modified so that they can account for the basic facts of coreference described above. The principle modifications are: (1) The construction rules for proper names and definite descriptions introduce entities into the discourse model as they do in Kamp and Reyle (1993), but the construction rule for pronouns interprets pronouns as referring directly to entities in the discourse model which is not how the rule works in Kamp and Reyle (1993); (2) Discourse entities in the model are ranked in terms of prominence (an idea that derives from the set of forward-looking centers in Centering Theory, Grosz, et al. 1995) which influences the way in which coreference is established. Discourse Prominence Theory includes three construction rules for the major types of reference and coreference, those for proper names, pronouns, and equivalence as shown below. We explain these rules by showing how they account for the differences in the ease of establishing coreference in sequences of different types of referring expressions. Then we discuss how this process is influenced by syntactic prominence.

CR.PN (Construction Rule for Proper Names)

Triggering Condition:
$[_{\gamma}$ ... $[_{NP}$ $[_{PN}$ $\alpha]]$...]
Instructions
- •Introduce a new discourse referent u into the universe of the DRS, U_k.
- •Introduce a new condition $\alpha(u)$ into the condition set of the DRS.
- •Substitute u for $[_{NP}$ $[_{PN}$ $\alpha]]$.

CR.PRO (Construction Rule for Pronouns)

Triggering Condition:
$[_{NP}$ $[_{Pro}$ $\alpha]]$
Instructions
- Chose an antecedent v_j, after considering every v_i I<j such that v_i and v_j exist in the ordered set of discourse referents in the DRS and are suitable antecedents, and substitute v_j for $[_{NP}$ $[_{Pro}$ $\alpha]$] in the triggering condition.
- • If no suitable antecedent v_j is present, introduce into the universe of the DRS a new discourse referent u.
- • Substitute u for $[_{NP}$ $[_{PRO}$ $\alpha]$].

CR.EQ (Construction Rule for Equivalence)

Triggering Condition:
$[_{\gamma}$ ... y ....]
such that $\alpha(x)$ and $\alpha(y)$

Instructions:
- Identify the discourse referent x in the ordered list of discourse referents $v_1...v_n$ by checking v_i after v_j where i<j.
- Introduce the new condition x = y
- Remove the condition $\alpha(y)$

Name-Pronoun Coreference. The example below involves a sentence with a name-pronoun sequence. The occurrence of the name triggers the construction rule for proper names (CR.PN) which posits a new entity (here shown as x) in the universe of the discourse and which introduces a condition in the universe consisting of the name predicated on the entity. The occurrence of the pronoun triggers the construction rule for

pronouns (CR.Pro) which searches the discourse universe for a "suitable antecedent" (one that matches on grammatically encoded features). For the example sentence below this leads fairly directly to a discourse universe that successfully represents the coreferential interpretation of the name and the pronoun.[3]

Ex: Jane thinks she is sick.
Construction Rules
1. CR.PN
2. CR.Pro

x
Jane(x)
x thinks x is sick

Name-Name Coreference. The next example involves a sentence with a name-name sequence. The first occurrence of the name triggers the construction rule for proper names as does the second occurrence of the name. In the example, this leads to a situation where there are two entities (x and y), both with the name *Jane* predicated on them, where one entity thinks the other is sick. Thus, the construction rules triggered by the name-name sequence naturally leads to a state of disjoint reference as shown by the intermediate discourse universe shown below. Coreference is only established through the construction rule for equivalence which is triggered by the presence in the discourse universe of the same name predicated on two distinct entities. This rule introduces a condition that equates the two entities thereby establishing coreference as shown in the second discourse universe below. The fact that coreference in name-name sequences requires an additional construction rule and involves an intermediate representation with disjoint reference explains why coreference is more difficult to achieve in name-name sequences than in name-pronoun sequences.

Ex: Jane thinks Jane is sick.
Construction Rules
1. CR.PN
2. CR.PN

x y
Jane(x)
Jane (y)
x thinks y is sick

3. CR.EQ

x y
Jane(x)
Jane (y)
x = y
x thinks y is sick

Pronoun-Name Sequences. The final example involves a sentence with a pronoun-name sequence. First, the occurrence of the pronoun triggers the construction rule for pronouns. Because the discourse universe does not contain a suitable antecedent, a new entity is introduced. The occurrence of the name subsequently triggers the construction rule for proper names which introduces a second entity upon which the name is predicated. Thus, a discourse universe is created with two distinct entities but without the identifying information that could trigger the construction rule for equivalence. Because there is no direct way of establishing coreference it is comparatively difficult to establish coreference in pronoun-name sequences.

Ex: She thinks Jane is sick.
Construction Rules
1. CR.Pro
2. CR.PN

X y
Jane (y)
x thinks y is sick

Effects of Prominence. As discussed above, the syntactic prominence of the antecedent influences the ease of establishing coreference: Greater prominence facilitates coreference in name-pronoun sequences while it inhibits coreference in name-name sequences. DPT accounts for this effect by adopting the notion,

[3] The nature of the appropriate predicate representation is an important question but one that we believe can be tackled independently of the mechanisms for establishing reference and coreference.

from Centering Theory, that entities in the discourse representation are ranked in terms of prominence. Based on our psycholinguistic results, it is clear that the syntactic status of an antecedent expression is a major determinant of its rank. The two construction rules that establish coreference use this ranking in different ways. The construction rule for pronouns searches for a suitable antecedent starting with the most prominent discourse entity, thus a prominent antecedent facilitates coreference with a pronoun. In contrast, the construction rule for equivalence involves an evaluation that begins with the least prominent entity, thus having a prominent entity inhibits coreference with a repeated name.

4 Backwards Anaphora

Backwards anaphora involves coreference in pronoun-name sequences, something that we said that naïve subjects do not accept. In fact, in a number of experiments (Gordon & Hendrick, 1997) we found that naïve subjects do not accept coreference in the kind of pronoun-name sequences that played a critical role in motivating details of the construct of c-command. However, naïve subjects do accept coreference in pronoun-name sequences when the pronoun is in a fronted adjunct, as shown by the proportion of acceptable coreference judgments for the sample sentence below:

Before **she** began to sing, **Susan** stood up. .88

In Gordon and Hendrick (1997) we found that coreference was acceptable for pronouns in fronted adjuncts that were subordinate clauses or prepositional phrases. This restricted range of acceptable backwards anaphora is in fact consistent with what is observed in corpus studies. Carden (1982) found that in over 95% of the naturally-occurring instances of backwards anaphora that he observed, the pronoun was in a fronted adjunct. Carden's finding is bolstered by recent work by van Hoek (1997), which though it focuses on quantitative analysis of the position of the full expression rather than the pronoun, provides clear support for the idea that in backwards anaphora the pronoun is overwhelmingly present in a fronted adjunct.

An additional important fact about coreference and fronted adjuncts emerged in our studies of reading time in sentences with intersentential coreference (Gordon, et al. 1993). Our research has shown a consistent effect (the "repeated-name penalty") where a sentence with a repeated name is read more slowly than a matched sentence with a pronoun. However, this effect is not observed when the repetition occurs in a fronted adjunct as shown in the example below:

Susan gave Fred a pet hamster.
In **his/Fred's** opinion, she shouldn't have.
Giving a pet as a gift is somewhat of an imposition.

The second sentence is read equally fast when it contains a repeated name as when it contains a pronoun.

In Gordon and Hendrick (1998) we account for these two important facts about coreference in fronted adjuncts – that they enable backwards anaphora and that there is no reading time penalty for names compared to pronouns – by considering the semantic function of adjuncts. Adjuncts serve to semantically modify the main clause to which they are attached. Accordingly, a fronted adjunct should cause a departure from the normal incremental construction of a discourse model where linguistic expressions are added to the existing discourse model so as to elaborate its semantic content. Instead, a fronted adjunct must first be processed in relation to the clause to which it is attached. In our model, this is accomplished by a (possibly temporary) partitioning of the discourse universe that is triggered by a fronted adjunct. Because a pronoun in a just begun discourse segment cannot possibly have a referent, it is held in an un-interpreted state and therefore can be subsequently equated with a following name. The two construction rules given below detail the partitioning of the discourse segment and the possible establishment of coreference for an entity in such a segment. The (temporary) semantic partitioning of the discourse universe provides a unified account of the possibility of backwards anaphora in fronted adjuncts and the absence of a penalty for names over pronouns in a fronted adjunct.

CR.FRONTED.Adjunct

Triggering Condition:
$[_{CP}$ [XP] CP...]
Instructions:
- Begin a new DRS U_{k+1}
- Introduce a new condition $K_n = \{\}$ into U_{k+1}.
- For any $[_{PN}\ \alpha\ ]$ within XP, introduce a new discourse referent u into the universe of the DRS U_{k+1} .
- For any other $[_{NP}\ \alpha\ ]$ within XP, introduce a new discourse referent u into K_n.
- Introduce a new condition $\alpha(u)$ into U_{k+1}
- Substitute u for $[_{NP}\ \alpha\ ]$ in XP.

CR. EQ.Adjunct

Triggering Condition γ:
the condition set $\alpha(u)$ within K_i where α is a pronoun
Instructions:
Equate u with a discourse referent v that is within the universe K_{i-1} that contains K_i.

5 Discussion

The work reported here argues that allowable coreference emerges from how construction rules for interpreting different types of noun phrases interact dynamically with a structure representing the meaning of a discourse. Further, it is argued that interpretation of coreferential expressions within and between sentences is done in a uniform manner, making use of the same construction rules and principles of discourse representation. This manner favors the use of pronouns compared to proper names for reference to prominent entities in a discourse.

The model that is developed – Discourse Prominence Theory (DPT) – integrates and elaborates on three theoretical sources: Discourse Representation Theory (DRT, Kamp & Reyle, 1993), Centering Theory (Grosz, et al. 1995), and the Binding Theory (Chomsky, 1986).

DRT is adopted as a formalism because it provides explicit mechanisms for mapping linguistic input onto semantic representations. Further, syntactic forms play an important role in the characterization of the linguistic input to the DRT construction rules and syntax clearly plays an important role in coreference. Finally, DRT describes semantic interpretation as an incremental process in which the interpretation of an utterance involves a dynamic interaction of the characteristics of the utterance with the discourse universe that represents the meanings created from the earlier utterances in the discourse. Beyond its usefulness in providing a satisfactory semantics of discourse, this dynamic view of language processing makes DRT attractive as a framework for characterizing the psychological processes of language comprehension which are generally regarded as dynamic and incremental.

Centering Theory provides key theoretical notions for understanding how reference and the form of referring expressions contribute to discourse coherence. DPT directly incorporates the idea of a set of forward-looking centers (Cf) in order to explain how structural prominence affects the interpretation of pronouns and repeated full expressions. This incorporation extends the role of the Cf so that it plays a role in the interpretation of referential expressions within an utterance as well as between utterances in a discourse segment. The idea of a backward-looking center (Cb) is not directly incorporated into the representation of discourse within DPT. Instead, the characteristics of the Cb emerge from how the model integrates an utterance into the current discourse universe and thereby linking it to the current discourse segment. The preference for realizing the Cb as a pronoun occurs because the construction rule for pronouns (CR.Pro) interprets a pronoun as referring to an entity in the current discourse representation and therefore forces integration of the utterance into the current discourse representation. Absent a clear cue for integration, such as a pronoun, there is always the possibility that the utterance is the beginning of a new discourse segment and should not be integrated into the current discourse universe. The existence of only a single Cb in an utterance occurs because only a single pronominal reference is needed in order to force the integration of an entire utterance into the current discourse universe.

The Binding Theory contributes to DPT through its emphasis on how the syntactic characteristics of the antecedent influence patterns of coreference. While this is the case, it is

obvious that the work that we have reviewed is not consistent empirically with the generalizations accepted by the binding theorists. Coreference phenomena appear to be much simpler than the binding theorists have taken them to be, and accordingly the effects of syntactic factors on allowable coreference syntactic effects emerge from the nature of the mechanisms posited in DPT for achieving coreferential interpretations.

The discrepancies that we observed between the empirical characterizations of the Binding Theory and those obtained using more systematic experimental techniques deserve some comment given the current emphasis in computational linguistics on appropriate empirical methods. We (Gordon & Hendrick, 1997) found that there were striking differences between what was seen as acceptable coreference in the linguistics literature and what emerged from quantitative analysis of judgments that were systematically obtained from subjects who were naive to syntactic theory. Further, the view of coreference that emerged from the judgments of naïve subjects was consistent with online measures of language comprehension and with analysis of corpus data. We believe that this consistency points to the usefulness of considering the kinds of intuitive judgments that have traditionally been used in trying to develop theories of linguistic competence. However, we also believe that it demonstrates the need for careful methodology in collecting intuitive judgments. Because of substantial variation in judgments about coreference, and because of the centrality of the theoretical issues at stake, claims about the favored or acceptable interpretations of linguistic expressions in discourse need to be supported by quantitatively analyzed data obtained from subjects who are naive to the hypotheses being investigated. Data on such intuitive judgments can only provide a stable, cumulative basis for theoretical development if they are collected with such methods.

6 Acknowledgement

Preparation of this report was supported by NSF grants SBR-9807028 and IIS-9811129.

References

Ariel, M. (1994). Interpreting anaphoric expressions: a cognitive versus a pragmatic approach**. Journal of Linguistics, 30**, 3-42.

Bresnan, J. (1998). Morphology competes with syntax: Explaining typological variation in weak crossover effects. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, and D. Pesetsky, (Eds.), **Is the Best Good Enough? Optimality and competition in syntax.** Pages 59-92. Cambridge, MA: MIT Press.

Carden, G. (1982). Backwards anaphora and discourse context. **Journal of Linguistics**, **18**, 361-387.

Chomsky, N. (1986). **Knowledge of Language: Its Nature, Origin and Use**. New York. Praeger.

Crawley, R.A., Stevenson, R.J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. **Journal of Psycholinguistic Research**, **19**, 245-264.

Frederiksen, J.R. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. **Discourse Processes**, **4**, 323-347.

Gordon, P.C., Hendrick, R., Ledoux, K., & Yang, C.L.. (1999). Processing of reference and the structure of language: An analysis of complex noun phrases. **Language and Cognitive Processes**.

Gordon, P.C., & Hendrick, R. (in press). Non definite NP anaphora: A reappraisal. **Proceedings of the Chicago Linguistics Society**. Chicago, IL: University of Chicago.

Gordon, P.C., & Hendrick, R. (1997). Intuitive knowledge of linguistic co-reference. **Cognition, 62**, 325-370.

Gordon, P.C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. **Cognitive Science, 22,** 389-424**.**

Gordon, P.C., Grosz, B.J., & Gilliom, L.A. (1993). Pronouns, names, and the centering of attention in discourse. **Cognitive Science**, **17**, 311-347.

Grosz, B.J., Joshi, A.K., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse, **Computational Linguistics, 21**, 203-226.

Kamp, H., & Reyle, U. (1993). **From Discourse to Logic. Introduction to Model Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory**. Dordrecht: Kluwer Academic Publishers.

Kennison, S.M., & Gordon, P.C. (1997). Comprehending referential expressions during reading: Evidence from eye tracking. **Discourse Processes**, **24**, 229-252.

Lappin, S. & Leass, H. (1994). An algorithm for pronominal anaphora resolution. **Computational Linguistics**, **20**, 535-561.

Reinhart, T. (1976). **The Syntactic Domain of Anaphora**. MIT.

Van Hoek, K. (1997). **Anaphora and conceptual structure**. Chicago, IL: University of Chicago Press.