Rational Coordinated Anaphora Theory

Rebecca Nesson, Floris Roelofsen and Barbara Grosz

TR-01-08

Computer Science Group
Harvard University
Cambridge, Massachusetts

Rational Coordinated Anaphora Theory

Rebecca Nesson and Floris
Roelofsen and Barbara Grosz

Rational Coordinated Anaphora theory is a novel explanatory theory that predicts how speakers generate anaphoric referring expressions in discourse, how hearers interpret them, and how all conversational participants coordinate their strategies to promote clear communication while minimizing effort. Its main premise is that conversational participants are, and expect each other to be rational. This paper presents Rational Coordinated Anaphora theory in detail and then contrasts it with Centering theory, demonstrating the reason for the robustness of Centering theory's main premise, Rule 1, and explaining some of Centering theory's previously puzzling limitations.

1. Introduction

The theory we develop in this paper is concerned with the generation and the interpretation of anaphoric referring expressions—expressions that are used to refer to entities that have been mentioned in the discourse previously. We refer to this theory as RATIONAL COORDINATED ANAPHORA theory (RCA theory), because its main premise is that conversational participants are, and assume one another to be rational. Speakers and hearers minimize their communicative effort and assume each other to do so. This premise lies at the basis of many contemporary attempts to explain language use and other aspects of human behavior (cf. Zipf, 1949; Grice, 1975). RCA theory applies it in a novel way.

RCA theory describes what it means to behave rationally in the specific case of generating and interpreting anaphoric referring expressions. At a high-level, in the case of interpretation, the theory holds that it is rational to take a referring expression to refer to its most easily accessible potential referent unless the resulting interpretation is inconsistent with world knowledge or highly implausible in the particular context in which the expression is used. In such cases the expression is taken to refer to its next most easily accessible potential referent iteratively, until a plausible interpretation is obtained. In the case of generation, RCA theory holds that rational speakers whose only purpose is to specify a certain entity will use the most easily producible available expression that is expected to be resolved to the intended referent. Rational speakers who have additional purposes besides specifying a certain referent will use a specially marked expression both to refer and to signal the hearer about the existence and content of their additional purposes.

RCA theory is both a *descriptive* theory, in that it describes the generation and interpretation of anaphoric expressions, and an *explanatory* theory, in that it provides an explanation of why people generate and interpret anaphoric expressions as they do. The explanatory aspect of RCA theory sheds new light on previously proposed theories and provides novel explanations of some intriguing linguistic phenomena that have been considered in the light of those theories.

RCA theory can be used to explain the general robustness, but also some important limitations of the main generalization provided by CENTERING theory (Grosz *et al.*, 1995), a widely-cited account of anaphoric reference. In particular, RCA theory provides novel and arguably more adequate explanations of why some anaphoric expressions are perceived as anomalous in certain contexts and why some utterances containing anaphoric expressions are difficult for hearers to process. RCA theory also makes a novel prediction that whenever an utterance involves anaphoric reference to several discourse referents, the most salient discourse referent is preferably referred to first. This prediction will be motivated in detail.

For simplicity, we will concentrate in this paper on names, descriptions, and third person singular anaphoric pronouns in written English discourse. Demonstratives, stressed anaphoric expressions (which may occur in spoken discourse), and zero pronouns (which may occur in languages other than English) are left out of consideration. We do expect that stress and zero pronouns can be incorporated into our theory in a natural way and that the resulting framework will be applicable cross-linguistically to spoken as well as written discourse.

The remainder of this paper is organized as follows. RCA theory is presented in section 2. Some applications are discussed in section 3. Section 4 concludes with a recapitulation and some directions for future work.

2. Rational Coordinated Anaphora Theory

The main premise of RCA theory is that speakers and hearers behave rationally in generating and interpreting anaphoric expressions. To make this rationality claim more precise, it is important to recognize that speakers and hearers may have several purposes in using and interpreting anaphoric expressions.

In this paper we address the case in which the *speaker*'s primary (but not necessarily sole) purpose in using an anaphoric expression is to specify a certain referent.[1] Speakers may sometimes have several additional purposes in using anaphoric expressions, as illustrated in section 2.2. RCA theory predicts how speakers generate anaphoric expressions that convey these additional purposes.

The *hearer*'s primary purpose is to identify the intended referent of the anaphoric expression. Apart from identifying the intended referent, which we will refer to as *resolving* the anaphoric expression, the hearer must also recognize if the speaker had

[1] We leave aside non-referential uses of anaphoric expressions. All uses of the term "anaphoric expression" should therefore be read as "anaphoric referring expression". We also leave aside uses of referring expressions to introduce new discourse entities. RCA theory predicates that the default resolution of an anaphoric expression will be to a previously mentioned discourse entity rather than a new entity where possible, thus obviating the need for a separate account of how to determine which is intended.

any additional purposes in using the expression, and if so, must identify what those purposes were.

We can now state the RCA theory rationality claim more precisely. This statement will be illustrated and further refined in the remainder of this section.

For *hearers* to be rational is to minimize their effort in resolving anaphoric expressions. This means that they normally take an anaphoric expression to refer to the most easily accessible potential referent from among the set of all potential referents.

To state what it means for *speakers* to be rational, we distinguish between speakers whose only purpose is to specify a certain referent, and speakers who have additional purposes. We take these two cases one at a time.

Rational speakers whose only purpose is to specify a certain referent minimize their effort in producing an anaphoric expression, but only as long as effective communication is secured. This means that they use the most efficient—the most easily producible—available expression that can be expected to be resolved as intended. Rational speakers minimize their generative effort conditionally: they expect the hearer's resolution process to be rational, and use an expression only if its resolution can indeed be expected to yield the intended referent. Thus, the speaker's and hearer's rationality are mutually constraining.

If an expression cannot be expected to be resolved as intended in a certain context, we will say that the expression is *blocked* in that context. The most efficient available expression that is not blocked (for anaphoric reference to some referent R in some context C) will be called the *optimal* expression (for anaphoric reference to R in C), and any other expression that is not blocked will be called a *marked* expression (for anaphoric reference to R in C). We can now rephrase more succinctly what we stated above about rational speakers whose only purpose is to specify a certain referent: they will use the *optimal* expression.

In contrast, rational speakers who have *additional* purposes will use a *marked* expression to convey the existence and content of their additional purpose to the hearer. Rational speakers with additional purposes will not use just any marked expression, but rather the most efficient marked expression that can be used to achieve their purposes.

Which expressions are *available* for anaphoric reference to a given entity in a certain context, and which entities are the *potential* referents of a given anaphoric expression in a certain context is partly determined by syntactic constraints. Several such constraints have been discussed in the literature (cf. Chomsky, 1981; Reinhart, 1983; Büring, 2005; Roelofsen, 2007).

One constraint, which is referred to as *Agreement*, is that the form of an anaphoric expression must agree with its intended referent in number, person, and natural gender. A second constraint, which is usually referred to as *Condition B*, says that a pronoun which occurs as the argument of a certain predicate cannot corefer with other arguments of that predicate. Other grammatical constraints, as well as more refined versions of Condition B have been widely discussed in the literature. We will not explicate these here, as this would require much additional terminology and is not necessary for the exposition of our theory. Roelofsen (2007) provides an overview and further references.

In the remainder of this section we describe the predictions made by RCA theory about a range of examples familiar from the literature on the generation and interpretation of anaphoric expressions. We focus on the explication of the details of RCA theory and its adequacy for handling these cases. We defer contrasting RCA theory with other theories of anaphoric reference until Section 3. In that section we demonstrate that RCA theory correctly predicts the facts in cases where other theories do not.

2.1 Resolving an Anaphoric Expression

To resolve an anaphoric expression is to select its intended referent from among a set of *potential referents*: those entities that fit the expression's descriptive content and are not excluded by grammatical constraints. To see how grammatical constraints may exclude an entity from being a potential referent, consider the potential referents of "him" in (1-c).

(1) a. My dog has been quite obstreperous lately.
b. I took him to the groomer yesterday.
c. The mangy old beast hates him.

Before any grammatical constraints apply, the set of potential referents includes all of the previously mentioned discourse entities: the speaker, the groomer, and the dog. Agreement dictates that the "him" in (1-c) cannot corefer with the "I" in (1-b). Therefore, the referent of "I", i.e. the speaker, is not a potential referent of "him" in (1-c). Condition B dictates that the "him" in (1-c) cannot corefer with its co-argument "the mangy old beast". Thus, the referent of "the mangy old beast", i.e. the speaker's dog, is not a potential referent of "him" either.

According to RCA theory, a hearer selects the referent of an anaphoric expression from among the potential referents so as to minimize the effort spent in doing so. That is, she will always take an anaphoric expression to refer to its most easily accessible potential referent. RCA theory follows the literature in assuming that ease of accessibility is strictly correlated with a hearer's focus of attention: one potential referent is more easily accessible than another just in case the former is more salient—more in focus—than the latter. Thus, according to RCA theory, anaphoric expressions are normally taken to refer to their most salient potential referent. Several accounts of what makes an entity more or less salient have been proposed in the literature. Any of these may be adopted to obtain a fully predictive theory, and RCA theory does not commit to any particular account. Our analysis of concrete examples is consistent with the theories of salience that can be found in the literature. RCA theory explicitly adopts the assumption that entities which have already been mentioned are always more salient than entities that have not been mentioned yet. Thus, whenever possible, referential expressions are taken to refer to entities that have already been mentioned in the discourse rather than entities being introduced for the first time. In example (1) above, the only remaining previously mentioned entity is the groomer, thus "him" in (1-c) resolves to the groomer. As an additional example, "he" in (2-b) is taken to refer to John, who has already been mentioned in (2-a), rather than some other entity that has not yet been mentioned in the discourse.

(2) a. John has been acting quite odd lately.
b. He called up Mike yesterday.
c. He wanted to meet him urgently.

Utterance (2-c) provides a somewhat more involved illustration of the resolution of anaphoric expressions. The pronoun "he" in (2-c) could be taken to refer to John, to Mike or to someone who has not been mentioned yet. A rational hearer will take the expression to refer to John, because John is the most salient, and therefore the most easily accessible discourse entity after (2-b).

Next, consider the pronoun "him" in (2-c). Mike is the only potential referent that has already been mentioned (notice that, by Condition B, John is not a potential referent here).[2] But the expression could also be taken to refer to someone who has not been mentioned yet. A rational hearer will take the expression to refer to its most salient potential referent: Mike.

RCA theory's default, salience-driven resolution of an anaphoric expression may be *revised* in case the resulting interpretation is not consistent with world knowledge or highly implausible in the given context. In this case, the expression is taken to refer to its next most salient potential referent, until a plausible interpretation is obtained. This is illustrated by fragment (3), a variant of fragment (1):

(3) a. My dog has been quite obstreperous lately.
b. I took him to the groomer yesterday.
c. He doesn't like the sound of his barking.

The pronoun "he" in (3-c) is first taken to refer to the speaker's dog, which is its most salient potential referent after (3-b). But this default resolution is revised after the rest of (3-c) has been processed: eventually, "he" is taken to refer to the groomer instead of the speaker's dog. Although the hearer eventually successfully resolves the anaphoric expressions in this example, the hearer's increased effort is directly related to the speaker's violation of her expectations. Perhaps more surprising than the hearer's accommodation of the speaker in this example is that a speaker might produce this discourse fragment in violation of the rules. One possible explanation for this would be a differing perception between the speaker and the hearer regarding which entity, the groomer or the dog, is more salient. If the difference in salience were starker between

[2] This assumes that "he" is resolved prior or at least simultaneously to "him". See Section 3.4 for additional discussion of this assumption and its predictive consequences.

the two entities RCA theory would predict that speakers would not produce this kind of "garden path".

2.2 Specifying a Referent

A speaker's primary purpose in using an anaphoric expression is to specify a certain referent. How does a speaker who has no additional purposes, i.e. whose only purpose is to specify a certain referent, decide which anaphoric expression to use? According to RCA theory, speakers minimize their generative effort as long as effective communication is secured. This means that they will use the most *efficient* expression that can be expected to be resolved as intended. Similar to the treatment of salience, RCA theory adopts no particular theory of efficiency. However, we note that both the *brevity* and *frequency* of expressions are widely thought to contribute to efficiency of production. RCA theory expects, we believe uncontroversially, that any reasonable theory of efficiency will produce a hierarchy of anaphoric expressions in which pronouns are generally more efficient than names and names are generally more efficient than descriptions.

However, the most efficient anaphoric expression may be blocked—it may not be resolved as intended. A rational speaker will only enhance efficiency as long as effective communication is secured. Thus, a rational speaker whose only purpose is to specify a certain entity will use the most efficient expression that is not blocked: the optimal expression.[3]

Let us illustrate this by means of an example. Consider fragment (4).

(4) a. Susan has been very generous lately.

b. She gave Betsy a bottle of wine yesterday.

[3] Blocking will be discussed in more detail in section 2.4.

c. Betsy said it was delicious.

In (4-b), the pronoun “she” is used to refer to Susan, because it is more efficient than any other expression that would be resolved as intended. In (4-c), the name “Betsy” is used to refer to Betsy, because the more efficient pronoun “she” is blocked: it could be taken to refer to Susan, who is at least as salient as Betsy in the given context.

2.3 Communicating Additional Purposes

A speaker may have various additional purposes in using an anaphoric expression besides specifying a certain referent. For example, he may want to provide new information about that referent, or he may want to indicate the beginning of a new discourse segment. This section addresses how a speaker who has such additional purposes decides which anaphoric expression to use, and how a hearer recognizes which additional purposes a speaker had in using an anaphoric expression.

In section 2.2, we concluded that a rational speaker whose only purpose is to specify a certain referent will use the optimal expression. RCA theory proposes that a speaker who has additional purposes will use a marked expression, and his use of a marked expression will signal to the hearer that he must have had additional purposes. This is what Horn (1984) called the *division of pragmatic labour*: “unmarked forms tend to be used for unmarked situations and marked forms for marked situations” (Horn, 1984, p. 26).[4] The use of a marked expression by the speaker imposes a burden on the hearer that is symmetrical to the conditional minimization the speaker performs in choosing anaphoric expressions. When a speaker uses a marked expression the hearer must determine the optimal expression first in order to determine that the expression used

[4] This idea is also closely related to a suggestion made by Grice (1975), namely that speakers sometimes “flout” his *Cooperativity Principle* in order to indicate that some non-standard interpretation is intended.

is marked. The hearer then seeks an additional purpose on the part of the speaker to justify the speaker's choice of the non-optimal expression.

The following examples demonstrate the use of marked anaphoric expressions to communicate additional purposes. First consider example (1), which is repeated below. The optimal expression for anaphoric reference to the speaker's dog in (1-c) is the pronoun “he”, but the marked expression “the mangy old beast” is used instead to present new information about the dog. “The mangy old beast” uniquely identifies the dog, but does so in a non-optimal way, thus it is an appropriate choice by the speaker to communicate an additional purpose. It offers information about the state of the dog and the speaker's disposition toward the dog.

(1) a. My dog has been quite obstreperous lately.
b. I took him to the groomer yesterday.
c. The mangy old beast hates him.

Next consider fragment (5).[5] The optimal expression for anaphoric reference to McEwan in the beginning of the second paragraph is the pronoun “he”. However, the marked expression “McEwan” is used instead to indicate the beginning of a new discourse segment. Here ‘"McEwan" uniquely identifies the referent, but in a non-optimal way, making it a suitable marked expression to communicate an additional purpose. It provides no additional information about the referent, so the reader is likely to understand it as a signal of the beginning of a new discourse segment.

(5) Ian McEwan was born on 21 June 1948 in Aldershot, England. He studied at the University of Sussex, where he received a BA degree . . . He has been writing

[5] This fragment is excerpted from `http://www.ianmcewan.com`.

> ever since.
>
> McEwan's works have earned him worldwide critical acclaim. Among them are the Somerset Maugham Award in 1976 for his first collection of short stories First Love, . . . , Last Rites; and the Santiago Prize for the European Novel (2004).

Presenting new information about the referent and indicating the beginning of a new discourse segment are only two of the possible purposes that speakers may have in using marked anaphoric expressions. Further research is needed for a complete specification of such purposes.

Rational speakers with additional purposes will not use just any marked expression, but rather the most efficient expression that can be expected to achieve their purposes. Thus, to provide new information about a referent, a rational speaker will normally use an expression which conveys exactly *that*, and no other, irrelevant information. Similarly, to indicate the beginning of a new discourse segment, a rational speaker will normally use the most efficient available expression that is *not* the optimal one. Just as pronouns are usually optimal, names are often the most efficient non-optimal expressions for anaphoric reference to a certain entity. This explains the empirical observation that proper names are often used instead of pronouns at the beginning of a new discourse segment (cf. Asher *et al.*, 2006).

2.4 Blocking

When the most efficient expression for a given referent is blocked, the speaker must use a less efficient expression instead. A case in point is fragment (4), which is repeated below.

(4) a. Susan has been very generous lately.

b. She gave Betsy a bottle of wine yesterday.

c. Betsy said it was delicious.

In (4-c), the name “Betsy” is used to refer to Betsy, because the more efficient pronoun “she” is blocked: it could be taken to refer to Susan rather than Betsy, because Susan is at least as salient as Betsy in the given context.

The general principle at work here is that a rational, cooperative speaker uses an expression E for a given referent R in a context C only if the resolution of E in C can indeed be expected to yield R. This general principle is sometimes called the principle of *recoverability* (cf. Beaver, 2004).

The use of blocked expressions usually yields an incoherent discourse, or even an unintended interpretation. Consider the following examples:

(6) a. Sarah and Beth saw a movie yesterday.

b. She said it was great.

(7) a. Susan has been talking to John Smith and John Brown.

b. She definitely prefers John at this point.

In (6), there is insufficient difference in salience between Sarah and Beth for the hearer to determine the intended referent of “she”. Similarly in (7), there is insufficient difference in salience between John Smith and John Brown for the hearer to determine the intended referent of “John” in (7-b). To avoid these ambiguities, the speaker should have used more informative anaphoric expressions instead—e.g., “Sarah” in (6-b) and “John Brown” in (7-b).

However, the ambiguity that results from the use of blocked anaphoric expressions does not always lead to incoherence. To see this, consider the following examples:

(8) a. My dog has been getting quite obstreperous lately.
 b. I took him to the groomer yesterday.
 c. He hates him.
 d. In fact, he tried to bite him last month.

(9) a. My dog has been getting quite obstreperous lately.
 b. I took him to the groomer yesterday.
 c. He hates him.
 d. In fact, he always tries to schedule his appointments when the other groomer is on duty.

(8-c) and (9-c) are ambiguous: “he” can be taken to refer to the dog and “him” to the groomer, or the other way around. The ambiguity is resolved in (8-d) and (9-d): (8) conveys that the dog hates the groomer, whereas (9) conveys that the groomer hates the dog. Strikingly, hearers report that the ambiguities that arise in (8-c) and (9-c) do not make these discourse fragments incoherent. It appears that hearers are willing to carry the ambiguity forward until it is resolved, or possibly even to let it remain unresolved, perhaps concluding that both dog and groomer hate each other.

In the next discourses the ambiguity is between an event and an object:

(10) a. I spent all day yesterday baking a molten chocolate cake.
 b. It was great.

(11) a. I spent all day yesterday baking a molten chocolate cake.

b. It was great.

c. But the cake didn’t turn out so well.

(12) a. I spent all day yesterday baking a molten chocolate cake.

b. It was great.

c. But I don’t think it was worth the time and effort.

Similar to the previous example, “it” in (10-b) is ambiguous. It could refer to the event of baking the cake or to the cake itself. In spite of this ambiguity, which may or may not be recognized, hearers find the discourse fragment entirely coherent. It is possible to stop the discourse after (10-b) or to continue with either of the alternative final utterances given in (11-c) and (12-c).

Finally, in the following discourses (adapted from Grosz *et al.*, 1995) the ambiguity is between the specific person holding an office and the characteristics of the office itself:[6]

(13) a. The Vice President of the United States is also the President of the Senate.

b. He is the President’s key man in negotiations with Congress.

c. As Ambassador to China, he handled many tricky negotiations.

(14) a. The Vice President of the United States is also the President of the Senate.

b. He is the President’s key man in negotiations with Congress.

c. He is required to be at least 35 years old.

In (13-b) and (14-b) the referent of “he” is either the current holder of the office of Vice President of the United States or a more general statement about the characteristics of whoever holds the office of the Vice President of the United States. The alternative final

[6] Cf. Donnellan’s (1966) distinction between the referential and attributive use of definite descriptions.

utterances (13-c) and (14-c) draw out the contrast between the two. Although it is clear that the two ways of resolving the reference lead to different meanings, both (13-b) and (14-b) are coherent.

RCA theory provides an explanation for the contrast between the unacceptable ambiguity of examples (6) and (7) and the innocuous ambiguity of examples (8)–(12). Both the hearer and the speaker know and accept that their salience rankings over the discourse entities may differ slightly in cases in which two entities are of similar salience in the discourse. However, in the two unacceptable examples, the previous mention of the two potential referents of the anaphoric expression in a coordinated structure makes it clear to both the speaker and the hearer that there is no substantial salience difference between the two entities. The hearer therefore determines that it is not rational for the speaker to have a salience ranking that clearly favors one entity over the other or to assume that the hearer will have such a salience ranking.

In contrast, in the acceptable ambiguity examples the salience rankings are not as clearly fixed. Consider (8) as an illustration. In (8) the hearer and speaker may have the same salience ranking over the dog or the groomer or they may have different rankings. If their rankings are the same then the discourse will proceed without any perceived ambiguity or anomaly. If their rankings differ, the speaker may surprise the hearer by producing (8-d) or (9-d). However, the hearer can conclude that the cause of the anomaly is the result of a difference in salience ranking and correct both the interpretation of the anaphoric expression and her salience ranking. RCA theory predicts that the processing time for the hearer will increase in this situation but psycholinguistic studies are necessary to confirm this prediction. If no disambiguating utterance is ever produced, the speaker and hearer may continue in the discourse with differing interpretations of its meaning. This may occur either when the difference in interpretation is of no consequence to the speaker or when the speaker and hearer are simply unaware that the misinterpretation has occurred because there is no indication of the mismatch between their salience rankings.

Speakers do sometimes *intentionally* use blocked expressions, even though, or rather *because*, they do not expect these expressions to be effective. For example, blocked expressions are often used in jokes, to lead the hearer first toward one interpretation, which is typically mischievous, and then enforce a revision of that interpretation, as if “nobody ever did anything wrong”. Here is our attempt at an illustration:

(15) a. Jessica’s 80 year old mother came by to show me her new car.
b. She is absolutely gorgeous!
c. I mean the Jaguar, of course.

Intentional flouting of the hearer’s rationality assumption leads to (not necessarily good) humor rather than incoherence when it is detectable by the hearer. This indicates a shared meta-level of understanding between the conversational participants that a coordinated rationality assumption is being employed by all.

Blocking may also play a role in determining the optimal expression for a given referent and therefore, indirectly, in a hearer’s recognition of a speaker’s additional purposes. In (4-c) for example, “Betsy” is optimal because the more efficient pronoun “she” is blocked. This leads the hearer to conclude that the speaker did not have any additional purposes in using the expression.

3. Applications and Comparison with Prior Theories

3.1 Centering

The main generalization provided by CENTERING theory is referred to in the literature as *Rule 1*. It says that whenever speakers use a pronoun to refer to an entity that was also referred to in the previous utterance, they always use a pronoun for anaphoric reference to the most salient entity which is referred to both in the current and in the previous utterance. RCA theory provides an explanation of the general robustness of this generalization, but also identifies some important limitations: in many situations it is rational for speakers to obey Rule 1, but in some situations it is not rational to do so.

RCA theory holds that a rational speaker uses a pronoun for anaphoric reference if (i) his only purpose is to specify a certain referent and (ii) the pronoun is expected to be effective in the given context.

Let us show that this explains why rational speakers often obey Rule 1. First, in most contexts a pronoun will be effective for anaphoric reference to the most salient entity that is mentioned both in the previous and in the current utterance. So condition (ii) is usually satisfied. The few cases in which it is not satisfied give rise to situations in which it is indeed not rational for a speaker to obey Rule 1, as we will demonstrate below.

Second, condition (i), which says that the speaker's only purpose is to specify a certain referent, holds for most cases of anaphoric reference. Moreover, Rule 1 only applies to utterances which contain at least one anaphoric pronoun. This has the practical effect of preventing Rule 1 from applying in the substantial category of additional purposes in which the speaker's additional purpose is to indicate the beginning of a new discourse segment because pronouns are typically not used for anaphoric reference across discourse segment boundaries.

This leaves only two cases in which it is not rational for a speaker to obey Rule 1. The first is when condition (i) is violated because the speaker has an additional purpose other than indicating the beginning of a new discourse segment. Fragment (1) illustrates this:

(1) a. My dog has been quite obstreperous lately.
b. I took him to the groomer yesterday.
c. The mangy old beast hates him.

Rule 1 is violated in (1-c), because a description is used for anaphoric reference to the most salient entity which is referred to both in (1-b) and in (1-c), i.e. the dog, while a pronoun is used for anaphoric reference to a less salient entity which is referred to both in (1-b) and in (1-c), i.e. the groomer. RCA theory provides a straightforward explanation: the description "the mangy old beast" is used for anaphoric reference to the speaker's dog in (1-c), instead of the pronoun "he", to present new information about the dog. RCA theory predicts that the use of a pronoun to refer to the groomer does not depend on a pronoun being used to refer to the dog, but rather on whether that pronoun is blocked for the groomer. In this case "him" is not blocked because Condition B removes the more salient dog from the set of potential referents.

The second case in which it is not rational for a speaker to obey Rule 1 is when condition (ii) is violated: a pronoun cannot be expected to be effective for anaphoric reference to the most salient entity which is referred to both in the previous and in the current utterance. This is the case, for example, if an even more salient entity is referred to in the previous, but *not* in the current utterance. Fragment (4) illustrates this:

(4) a. Susan has been very generous lately.

b. She gave Betsy a bottle of wine yesterday.

c. Betsy said it was delicious.

Rule 1 is violated in (4-c), because a name is used for anaphoric reference to the most salient entity which is referred to both in (4-b) and in (4-c), i.e. Betsy, while a pronoun is used for anaphoric reference to a less salient entity that is referred to both in (4-b) and in (4-c), i.e. the bottle of wine. Again, RCA provides a straightforward explanation: the name “Betsy” is used, because the more efficient pronoun “she” is blocked. This has no effect on the availability of “it” to refer to the bottle of wine, so “it” remains the optimal anaphoric expression.

In conclusion, RCA explains the general robustness as well as some important empirical limitations of the main generalization provided by CENTERING theory.

3.2 Anomalous Expressions

Anaphoric expressions are sometimes perceived as anomalous, as is illustrated by the following two fragments from (Grosz *et al.*, 1995):

(2) a. John has been acting quite odd lately.

b. He called up Mike yesterday.

c. He wanted to meet him urgently.

(16) a. John has been acting quite odd lately.

b. He called up Mike yesterday.

c. *John wanted to meet him urgently.

“John” in (16-c) is perceived as anomalous, in contrast with “he” in (2-c). The explanation that CENTERING theory offers for this kind of anomaly is that Rule 1 is obeyed in

(2-c) and violated in (16-c). But this hypothesis seems too strong: violations of Rule 1 do not always result in anomaly, as shown by the illustrations given below.

RCA theory provides a more subtle explanation of anomaly: an anaphoric expression is perceived as anomalous if and only if the expression is marked (which tells the hearer that the speaker must have had some additional purpose) and the hearer cannot identify any additional purpose that the speaker may plausibly have had in using the expression. Put simply, an expression is anomalous if the rationale behind the speaker's use of a marked expression cannot be determined.

The *if*-part of this explanation is illustrated by fragment (16). In (16-c), the optimal expression for anaphoric reference to John is the pronoun "he", but the name "John" is used instead. This tells the hearer that the speaker must have had some additional purpose in using the expression. But it is not clear which purpose that might be. The expression does not provide any new information about John, and (16-c) does not mark the beginning of a new discourse segment either. Although the hearer may go to substantial lengths to find an additional purpose before assuming lack of rationality on the part of the speaker, in this case no rational explanation can be found. This is what explains that "John" in (16-c) is perceived as anomalous. The use of proper names where pronouns are optimal are perhaps the clearest case of this sort of anomaly because both the pronoun and name generally contribute no information other than the identity of the referent. This makes it particularly difficult for the hearer to maintain her rationality assumption about the speaker by generating some additional purpose as an explanation for the use of the marked expression.

The *only if*-part of the explanation is illustrated by fragments (4) and (1).

(4) a. Susan has been very generous lately.

b. She gave Betsy a bottle of wine yesterday.

c. Betsy said it was delicious.

(1) a. My dog has been quite obstreperous lately.

b. I took him to the groomer yesterday.

c. The mangy old beast hates him.

In (4-c), “Betsy” is the optimal expression for anaphoric reference to Betsy, because the more efficient pronoun “she” is blocked. Therefore, the hearer is not lead to believe that the speaker must have had additional purposes in using the expression. This explains why “Betsy” in (4-c) is not anomalous. Note that CENTERING theory predicts that (4-c) should be anomalous because Betsy is the most salient entity in that utterance that was also referred to in the previous utterance and the less salient bottle of wine is referred to with a pronoun.

The case of (1-c) is different. Here, the optimal expression for anaphoric reference to the dog is the pronoun “he”. The speaker used the marked description “the mangy old beast” instead. So the hearer concludes that the speaker must have had an additional purpose in using this expression. But, unlike in the case of (16-c), the speaker’s additional purpose is easily identified here: the expression conveys new information about the dog. This, then, explains why “the mangy old beast” in (1-c) is not perceived as anomalous. Again, CENTERING theory will make the erroneous prediction that (1-c) is anomalous because the less salient groomer is referred to using a pronoun while the most salient entity in the utterance, the dog, is not.

3.3 Reading Times

Much psycholinguistic research has been devoted to measuring reading times for alternative anaphoric expressions. RCA theory can be used to explain the results of many of these experiments. In this paper, we focus on some particularly striking results obtained by Gordon *et al.* (1993) considering discourse fragments like the ones in (17), (18), and (19).

(17) a. Bruno was the bully of the neighborhood.
b. He chased Tommy all the way home from school one day.
c. He watched him hide behind a big tree and start to cry.
d. He yelled at him so loudly that all the neighbors came out.

(18) a. Bruno was the bully of the neighborhood.
b. He chased Tommy all the way home from school one day.
c. He watched Tommy hide behind a big tree and start to cry.
d. He yelled at Tommy so loudly that all the neighbors came out.

(19) a. Bruno was the bully of the neighborhood.
b. Bruno chased Tommy all the way home from school one day.
c. Bruno watched Tommy hide behind a big tree and start to cry.
d. Bruno yelled at Tommy so loudly that all the neighbors came out.

Gordon *et al.* (1993) observed the following reading times when subjects were presented with these and similar discourse fragments:

Paradigm	Example	Reading Time (in ms)
Pro-Pro	(17)	7.518
Pro-Name	(18)	7.624
Name-Name	(19)	8.460

Apparently, “Bruno” in (19) requires significantly more reading time than “he” in (18). This observation, which is often referred to as the *repeated-name penalty*, challenges theoretical models such as that of Gernsbacher (1989), which predict that explicit names or descriptions are always easier to process than pronouns.

The data is striking in another respect as well. Whereas “Bruno” in (19) requires significantly more reading time than “he” in (18), “Tommy” in (18) does *not* require significantly more reading time than “him” in (17). So proper names do not *always* require significantly more reading time than pronouns.

Gordon *et al.* (1993) hypothesize that extra reading time is required just in case the most salient entity to which anaphoric reference is made in an utterance is not referred to using a pronoun, but a repeated name. That is, “Bruno” in (19) requires more reading time than “he” in (18), because Bruno is the most salient entity in both fragments. “Tommy” in (18) does not require more reading time than “he” in (17), because Tommy is not the most salient entity in these fragments. Gordon *et al.* hold that this hypothesis is in accordance with CENTERING theory: extra reading time is required just in case Rule 1 is violated.

These considerations are unsatisfactory, because Gordon *et al.* assume an oversimplified formulation of Rule 1. In its original formulation by Grosz *et al.* (1995), Rule 1 says that *whenever a speaker uses a pronoun to refer to an entity that was also referred to in the previous utterance*, he must use a pronoun for anaphoric reference to the most salient entity which is referred to both in the current and in the previous utterance. Gordon *et al.* ignore the qualifying clause (given in italics), which, according to a corpus study by Poesio *et al.* (2004), destroys the rule’s descriptive adequacy: under standard assumptions as to the meaning of *utterance* and *salience*, almost 97% of the utterances considered by Poesio *et al.* obey Rule 1 as formulated by Grosz *et al.* (1995), while less than 45% of those utterances obey the simplified version of Rule 1 that is assumed by Gordon *et al.* (1993).

RCA theory provides a more satisfactory explanation: first, “Bruno” in (19) requires more reading time than “he” in (18) because “Bruno” is a marked expression in the given context, so the hearer will spend some time attempting to reconstruct the additional purpose that the speaker must have had in using the expression (without success in this case).

“Tommy” in (18) also requires more reading time than “him” in (17) for this reason. On the other hand, “him” occurs in a Condition B environment in (17), and as is well known from the acquisition literature, restricting the set of potential referents of pronouns in Condition B environment requires a significant processing load (cf. Reinhart, 2006). This complexity is avoided in (18). Apparently, the optimality of “him” which makes it relatively easy to process is balanced by the additional complexity involved in determining its potential referents, so that, cumulatively, (18) does not require significantly more reading time than (17).

3.4 Salience Alignment

We have shown thus far that RCA theory explains data that had previously been observed but unexplained. In this section we present an interesting prediction which, to the best of our knowledge, has not been discussed in the literature before. The prediction is merely stated here. Future empirical research is necessary to corroborate it.

Whenever an utterance involves anaphoric reference to several discourse referents, RCA theory predicts that the most salient discourse referent is most likely to be referred to early in the utterance where earliness is determined by the processing order in which hearers perform resolution.[7] The reason for this is twofold. First, at the beginning of an

[7] Two possible hypotheses that have some support in the literature are that processing of anaphoric expressions occurs in left-to-right order in a sentence and that it occurs, at least in English, in subject-first order. Either of these theories may be used here to define earliness and generate a prediction of expected speaker behavior.

utterance syntactic constraints and world knowledge cannot apply to resolve ambiguity because there is no syntactic or pragmatic context at that point. Salience rankings over the discourse referents are available right from the beginning of an utterance. Thus comparative salience is the only basis on which to resolve ambiguity. By placing the most salient discourse referent early in the utterance, the speaker is guaranteed to be able to effectively use the simplest anaphoric expression for that referent. Second, using the most salient discourse referent first is likely to induce syntactic and contextual constraints on the resolution of anaphoric expressions which occur later in the utterance, making it more likely that the speaker will be able to use highly efficient expressions for those referents as well even though they are not the most salient discourse referents.

On the other hand, if the speaker chooses to refer to a discourse referent that is not the most salient one early in the utterance, it is likely that she will not be able to effectively use the most efficient anaphoric expression for that referent. The introduction of that referent early in the utterance will still impose some syntactic and contextual constraints on other positions in the utterance but until the most salient discourse referent is used any ambiguity between it and other referents is likely to remain. A speaker may still choose to make early reference to a discourse referent that is not the most salient for other reasons. For instance, language-specific lexical constraints imposed by the best or only choice of verb for a particular utterance might force a particular ordering over the introduction of discourse referents.

4. Conclusion

It is often and quite uncontroversially assumed that conversational participants minimize, and assume one another to minimize their communicative effort. RCA theory spells out what this means exactly in the specific case of generating and interpreting anaphoric referring expressions. We have shown that the theory explains a wide range of empirical observations, and also leads to interesting new predictions.

There are several promising directions for future work. First, RCA theory makes predictions about many more phenomena than the ones discussed above. Some of these phenomena have been studied empirically, some have not. The predictions made by RCA theory must be compared with the actual empirical findings, either to further corroborate the theory or to refine it.

Second, RCA theory may be applied to improve systems for automated generation and interpretation of anaphoric expressions, and it may also be used for automated discourse segmentation.

Finally, the theory could be extended in various ways. First, we could consider speech instead of written discourse. Preliminary investigations indicate that various observations regarding the interpretation of stressed pronouns (cf. Kameyama, 1999; Beaver, 2004; de Hoop, 2004) could be explained in the general framework of RCA theory. Second, we could consider languages other than English, in which the inventory of anaphoric referring expressions may be more extensive. In particular, it would be interesting to consider languages with so-called zero pronouns. We are confident that zero pronouns, intonation, and other factors which have been kept out of consideration in the present paper can be incorporated into the general framework in a natural way, and that the resulting theory will be applicable to spoken and written discourse cross-linguistically.

References

Asher, N., Denis, P., and Reese, B. (2006). Names and pops and discourse structure. In C. Sidner, J. Harpur, A. Benz, and P. Kuhnlein, editors, *Proceedings of the workshop on constraints in discourse*.

Beaver, D. (2004). The optimization of discourse anaphora. *Linguistics and Philosophy*, **27**(1), 3–56.

Büring, D. (2005). *Binding Theory*. Cambridge University Press.

Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. Mouton de Gruyter.

de Hoop, H. (2004). On the interpretation of stressed pronouns. In R. Blutner and H. Zeevat, editors, *Optimality Theory and Pragmatics*. Palgrave/Macmillan.

Donnellan, K. (1966). Reference and definite descriptions. *Philosophical Review*, **75**, 281–304.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, **32**, 99–156.

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

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics*, volume 3, pages 41–58. Academic Press, New York.

Grosz, B., Joshi, A., and Weinstein, S. (1995). Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, **21**(2), 203–225.

Horn, L. (1984). Towards a new taxonomy of pragmatic inference: Q-based and R-based implicatures. In D. Schiffrin, editor, *Meaning, Form, and Use in Context*, pages 11–42. Georgetown University Press.

Kameyama, M. (1999). Stressed and unstressed pronouns: complimentary preferences. In P. Bosch and R. van der Sandt, editors, *Focus: linguistic, cognitive, and computational perspectives*. Cambridge University Press.

Poesio, M., Stevenson, R., di Eugenio, B., and Hitzeman, J. (2004). Centering: a parametric theory and its instantiations. *Computational Linguistics*, **30**(3).

Reinhart, T. (1983). *Anaphora and Semantic Interpretation*. Croom Helm, London.

Reinhart, T. (2006). *Interface Strategies*. MIT Press.

Roelofsen, F. (2007). *Constraints on the Interpretation of Pronouns*. Manuscript, University of Amsterdam.

Zipf, G. (1949). *Human behavior and the principle of least effort*. Addison-Wesley, Cambridge.