

Game Theory and Discourse Anaphora^{*}

Robin Clark[†] (rclark@babel.ling.upenn.edu)

Department of Linguistics, University of Pennsylvania

Prashant Parikh

IRCS, University of Pennsylvania

Abstract. We develop an analysis of discourse anaphora—the relationship between a pronoun and an antecedent earlier in the discourse—using games of partial information. The analysis is extended to include information from a variety of different sources of information, including lexical semantics, contrastive stress, grammatical relations and decision theoretic aspects of the context.

In this paper, we will develop a game theoretic treatment of some simple cases of discourse anaphora.¹ Our attention will be mainly focused on short texts of the type shown in (1):

- (1) a. A cop saw a hoodlum. He yawned.
- b. A cop saw a hoodlum. He chased him.

The simple texts in (1) have the interesting property that, all else being equal, the pronouns in the second sentence are taken by most speakers to be unambiguous. For example, the pronoun *he* in (1)a preferentially refers to the cop and not to the hoodlum. Equally in (1)b, *he* tends to refer to the cop and *him* refers to the hoodlum, and not the other way around.

It is useful to compare anaphors across sentences with those that occur inside a sentence, something we will not treat in this paper. Consider in this light (2):

- (2) The priest told John his pants were on fire.

It is our judgment that the pronoun *his* in (2) is fully ambiguous; the pronoun can pick out either the priest or John with little or no bias for one or the other. This intuition contrasts sharply with the intuitions about the pronouns in (1). Indeed, a speaker who wishes to refer to the hoodlum, and not the cop, in the subject of the next clause, would do well to use a definite description:

^{*} Both authors wish to thank Richard Breheny and Bill Labov for valuable comments made on an earlier draft of this paper.

[†] I acknowledge the generous support of a grant from the NIH, grant NS44266.

¹ On a game theoretic approach to language, see Parikh (2001), Parikh (2006) or Parikh & Clark (2005). Myerson (1991) gives a good overview of analytic game theory.

- (3) a. A cop saw a hoodlum. The hoodlum yawned.
- b. A cop saw a hoodlum. The hoodlum ignored him.

Given the simple paradigms in (1) and (3) we will develop a game theoretic treatment in section 1. In section 2, we will extend the analysis to cover some apparent counterexamples. Finally, in section 3 we summarize the theory we have developed and point out some of its more distinctive aspects.

The basic idea we will develop is a simple one. Speakers use discourse anaphors (pronouns, definite descriptions and, in some cases, names that have already been introduced) when they have reason to suppose that their audience can appropriately assign the correct discourse entity to the expression. Speakers, in other words, use expressions strategically based on their assessment of how hearers will choose to interpret their utterances. Hearers interpret utterances strategically based on their assessment of how speakers would most likely encode the information. While this kind of strategic interaction might seem hopelessly circular, the analytic tools provided by game theory provide a straightforward non-circular analysis of exactly this kind of strategic interaction. The non-circularity of the treatment follows from the fact that the speaker and the audience are both aware of the structure of the game. That is, they know what entities are available in the discourse model for the speaker to choose from, they are aware of the various ways in which the speaker could encode the referent (pronoun, definite description, name and so forth) and they are aware of the structure of the preceding utterance. This shared knowledge allows the participants to compute a Pareto-Nash Equilibrium profile for the game. This is a strategy profile for both the speaker and the audience that defines a set of optimal choices for the speaker and the audience. They are optimal in the sense that neither the speaker nor the audience can do better by deviating from the strategy profile. The profile indicates, in particular, the best way for the speaker to encode her meaning. Then, given an encoding, the profile indicates the best way for the hearer to assign that expression a referent.

We will use games of partial information to model the strategic inferences involved in discourse anaphora. In a game of partial information, there is a forest of game trees. In the present case, the game involves referring to and picking out a particular discourse entity or sequence of discourse entities. Some of the nodes in the trees correspond to information states that are ambiguous: in the case at hand, the audience does not know with certainty, given that the speaker has produced a pronoun, which part of the game tree the speaker is in. Thus, the game involves only partial information. As we will demonstrate, we can solve

such a game, producing an optimal strategy profile for the participants. We turn to this task in the next section.

1. Discourse anaphora and strategic interaction

We turn, now, to the analysis of the basic cases of discourse anaphora. For simplicity, our analysis will be focused on languages like English which lack null pronouns. Although it is a straightforward project to extend the analysis to languages with null pronouns, doing so would make our presentation unhelpfully complex; we will leave these cases for future analysis. We will follow the uncontroversial assumption that indefinite noun phrases are used to introduce entities into the discourse.² For example, a sentence like:

- (4) A cop saw a hoodlum.

introduces two entities into the discourse model, one for the cop and one for the hoodlum. We will suppose, further, that expressions like pronouns and definite descriptions draw on the discourse model to establish their reference; in particular, we will assume that both singular pronouns and, contrary to Russell, singular definite descriptions refer to entities. Thus, given the sentence:

- (5) He yawned.

the hearer is faced with a decision problem: Which element of the discourse model should she take as the referent of *he*? If we extend these considerations slightly, we should observe that the speaker is also faced with a decision; namely, given a context and a discourse model, \mathcal{D} , should he use a pronoun to refer to a particular element of \mathcal{D} or should he use a definite description? Clearly there are costs for both choices. If the speaker chooses a pronoun, then he risks that the hearer will select an incorrect element of \mathcal{D} . If the speaker chooses a definite description, he reduces the risk of misunderstanding, but increases the amount of work that must be expended in producing and processing the utterance; definite descriptions are longer and syntactically more complex than pronouns.

We propose that the problem can best be solved as a game of partial information. The speaker and hearer share some common knowledge

² Exactly how indefinites do this is beyond the scope of this paper. Stalnaker (1998) is a basic contribution. For a general discussion, see Roberts (2004). See Pietarinen (2004), and Clark (in press) for some discussion grounded in games. This area has been subject to a great deal of research from a number of perspectives including Dynamic Semantics (Groenendijk & Stokhof, 1991; Dekker, 2004) and Discourse Representation Theory (Kamp & Reyle, 1993; van Eijck & Kamp, 1997).

and have some interests in common. We can represent their common knowledge, their choices and their interests as a set of game trees. By finding the *Pareto-dominant Nash equilibrium* of the game, the players can most efficiently solve their problem and communicate.³ Suppose, then, that the speaker has uttered the sentence in (4), introducing the following discourse entities:

- (6) d_1 = the cop
 d_2 = the hoodlum

Now suppose that the speaker wishes to encode the meaning that the cop yawned. Both the speaker and the hearer know that the speaker could refer either to the cop or the hoodlum using either a pronoun or a definite description. Having encoded the intended meaning, the hearer must decide whether to associate the expression with d_1 , the cop, or d_2 , the hoodlum.

The game tree in figure 1 shows the various moves available to the speaker and hearer as well as the payoffs that they can expect once the choices have been made. Figure 1 shows two trees, one rooted in information state s and the other rooted in information state s' . Consider, first, the tree that is rooted at information state s . Information state s is associated with probability ρ and shows the case where the speaker intends to refer to d_1 , the cop in the discourse model \mathcal{D} , while the tree rooted at information state s' (associated with probability ρ') shows the case where the speaker intends to refer to d_2 , the hoodlum. The branches from the root show the possible moves that can be made by the speaker, while the branches emanating from these show the hearer's possible moves. The leaves show a set of ordered pairs of payoffs, where the first element is the payoff to the speaker while the second is the payoff to the hearer. Finally, the circled nodes are states which the hearer cannot distinguish.

Thus, if the speaker intends to refer to d_1 she can either use a definite description, *the cop*, or a pronoun, *he*. If she uses the definite description, she succeeds in referring to d_1 unambiguously but at a cost of some work for both her and the hearer; she must go to the effort of actually producing the definite description—which is work—and the hearer must go to the effort of processing it—which is more work. Furthermore, we will suppose that referring to a prominent element

³ A Nash equilibrium is a strategy that offers each participant the best payoff given the strategies of the other players. That is, in a Nash equilibrium a player has no reason to change his or her strategy since any other move results in a lower payoff. A game may have several Nash equilibria. A Pareto-dominant Nash equilibrium is a Nash equilibrium whose payoffs are at least as high as the payoffs in any other Nash equilibrium; no other Nash equilibrium offers a better payoff.

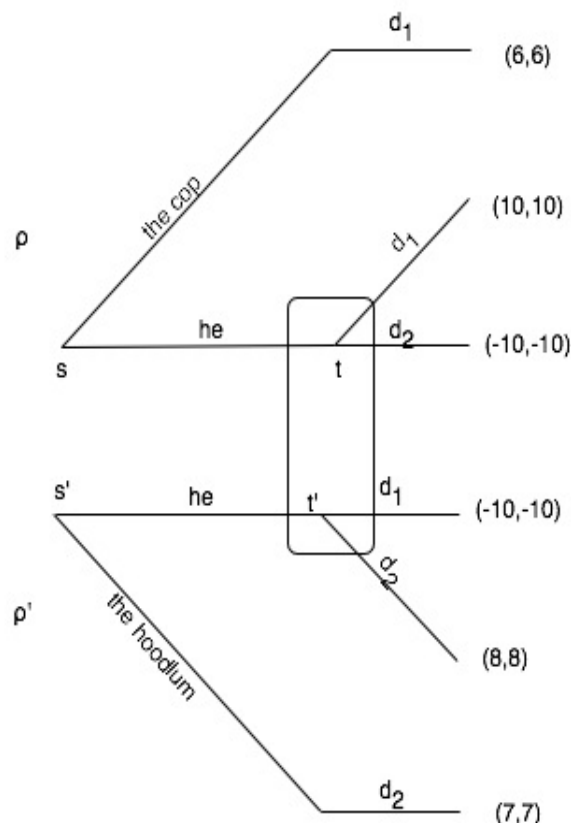


Figure 1. A Game Tree for Simple Discourse Anaphora

(the subject of the preceding sentence) with a full description rather than a pronoun entails some cost. Thus, I have shown a payoff of $(6, 6)$; the speaker and the hearer have communicated successfully, but at a cost. Suppose she uses a pronoun. Now, the hearer can either pick d_1 , the intended referent, or d_2 the boy. In the former case, communication has succeeded at the cost of very little work. Both the speaker and the hearer are happy and get a payoff of $(10, 10)$. If, however, the hearer selects d_2 , then communication has failed, an eventuality that both the speaker and hearer find unpleasant and wish to avoid. We therefore assign this outcome a payoff of $(-10, -10)$. The tree rooted at s' is nearly symmetrical, with d_1 and d_2 substituted for each other throughout the discussion. The one difference in payoffs—choosing d_2 for *he* has a payoff of $(8, 8)$ and not $(10, 10)$ —reflects our assumption that it

is slightly less efficient to pronominalize a less prominent element (in this case, an object).

Clearly, a complete theory would include a detailed account of the costs and benefits of various expressions. One possibility is that elements with high conventional semantic content like names and definite descriptions incur cost than elements like pronouns (or zero elements like null pronouns). This account would build on the notion of *efficiency* discussed in Parikh (2001); natural languages are context dependent and prefer means that can draw on context.

We suspect that this notion of efficiency is built on the minimization of effort, although this notion clearly requires clarification. Bill Labov (p.c.) has suggested that the costs could be related to the size of the set from which the speaker draws the expression. The set of pronouns in any given language is quite small, particularly as compared to the set of names and possible descriptions. On this account, it is easier for the speaker to make a choice from a small, fixed set than from a large (in fact, infinite) open set.⁴ Thus, the cost of a pronoun will be less than the cost of descriptions. Names present something of a problem here, since the set of names that an object has can be quite small. Often, however, we might not have access to the name of an object—many objects simply lack names. A full investigation of this problem is well outside the scope of the present paper. For the moment, we will summarize our method of apportioning payoffs as based on the interaction of three principles:

- (7) a. It is generally more costly to use longer expressions.
- b. It is generally more costly to use expressions with “high” conventional content (thus, names and descriptions are costlier than pronouns).⁵
- c. It is cheaper to refer to a more prominent element with a pronoun; it is correspondingly marked to refer to a more prominent element with a description or name when a pronoun could be used. Prominence is, here, calculated on the basis of

⁴ For example, the policeman introduced in our examples could be referred to as *the cop*, *the policeman*, *the officer of the law*, *the blue coated minion of the state* and so on. We enter, here, into questions of stylistics that lie far outside the scope of this paper.

⁵ By conventional content we mean the content of an element that is fixed by the language independent of context. Thus, pronouns have little conventional content but names and descriptions have high conventional content.

the grammatical function the element plays in the preceding sentence.⁶

We will discuss the prominence ranking below. The three principles in (7) rely on linguistic structure to establish the basic game trees. We will argue, below, that contextual information can condition the probabilities associated with information states. Since Nash equilibria are computed on the basis of both the payoffs and the probabilities, we will see that context can change the preferred interpretation of discourse anaphors.

Now, since we have no further information, we suppose that $\rho = \rho'$. This means that information state s is as likely as information state s' . The Pareto-dominant Nash equilibrium of this game is the strategy:

$$\{(s, \text{he}), (s', \text{the hoodlum}), (\{t, t'\}, d_1)\}$$

That is, if the speaker is in information state s where he wishes to refer to d_1 , she will use *he*. The hearer, who is now indeterminate between information state t and information state t' , will select d_1 . If, on the other hand, the speaker is in state s' (where she wishes to refer to d_2 , she will use *the hoodlum*); since the hearer's choice is determined unambiguously in the example, we have not included it in the strategy profile. Notice that the same strategy profile explains the data in (1)a and (3)a; namely, the preference for using a pronoun to refer to the subject of the preceding sentence and a description to refer to the object.

Now let us turn to the slightly more complex case, noted above in (1)b where two pronouns are used:

(8) A cop saw a hoodlum. He chased him.

Again assuming that *a cop* invokes a discourse entity, d_1 , and *a hoodlum* invokes entity d_2 , there is a strong preference to take *he* as referring to d_1 and *him* as referring to d_2 . The situation can again be represented as a game tree as shown in figure 2.

We have represented the game as involving two moves, as above. In this tree, we have shown only the sequence of the two possible referring expressions and the choice of two discourse entities. Thus, the speaker must decide whether to use two definite descriptions, a pronoun and a definite description or two pronouns.⁷ Of course, if the speaker uses

⁶ Of course, grammatical function is confounded with linear precedence in English, so the relevant principle could be grounded in linear precedence. Thus, it could be that this principle follows from general properties of short term memory.

⁷ We could, of course, represent each choice as a separate edge in the game tree. Thus, the speaker would first choose how to encode the subject of the sentence, the

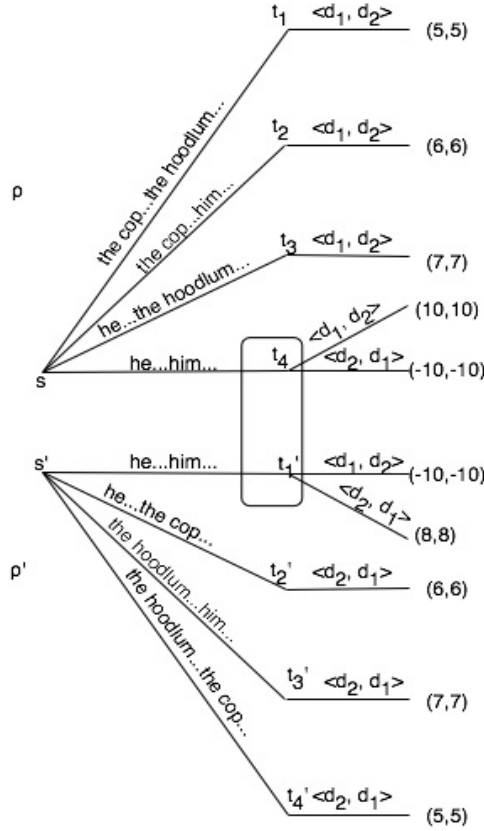


Figure 2. A Game Tree for Two Discourse Anaphors

two pronouns, then the hearer cannot know with certainty which game tree he is in, a fact represented by circling the ambiguous information state nodes in the diagram. Both the speaker and the hearer can exploit properties of the grammar to narrow down the choices. For example, since the grammar rules out the case where the two pronouns refer to the same entity, we need not include branches where the hearer chooses the same discourse entity twice.

The payoffs associated with each sequence of choices reflect the work of production and perception as well as success of communication. For example, the choice *the cop...the hoodlum...* guarantees success-

hearer would pick an entity from the choice set; then, the speaker would choose how to encode the second entity. The resulting game tree is more complex than the one in figure 2 and rather harder to read. Since it doesn't really add information that is not already contained in the simpler figure, we have not shown it.

ful communication but incurs work for both the speaker (in terms of production) and the hearer (in terms of perception). The choice *the cop...him...* is ranked slightly higher for both the speaker and the hearer since it involves successful communication with less work due to the replacement of one definite description by a pronoun. Notice that the best option for both the speaker and the hearer in terms of payoffs (and, in particular, effort, which is reflected in the payoffs) is *he...him...* where the hearer correctly chooses the discourse entities. The problem, of course, is that this encoding also runs the risk of miscommunication. A final factor we will take into account in determining payoffs is the relative prominence of an element; specifically, the subject of the preceding sentence is more likely to be the target of a discourse anaphor in the next sentence.

How should the speaker and the hearer play the game? If $\rho = \rho'$, as above, then the Pareto-Nash equilibrium is the strategy:

$$\{(s, \text{he...him...}), (s', \text{the hoodlum...him...}), (\{t_4, t'_1\}, \langle d_1, d_2 \rangle)\}$$

That is, if the speaker is in information state s , where he wishes to refer to the sequence of discourse entities $\langle d_1, d_2 \rangle$, then he should use the pronouns *he* followed by *him*. The hearer should respond by picking the pair $\langle d_1, d_2 \rangle$; that is, the choice where $he = d_1$ and $him = d_2$. If the speaker is in state s' then he should use *the hoodlum* and *him*, with hearer's choice being determined as shown in figure 2. Notice that the strategy profile accounts for the data in (1)b and (3)b.

The following two short texts:

- (9) a. A man saw a boy. He kicked the man in the shins.
- b. A man saw a boy. The boy kicked him in the shins.

Although (9)a is interpretable, it is decidedly odd. The text in (9)b is entirely acceptable. The analysis of the game in Figure 2 correctly distinguishes between (9)a and (9)b on the assumption that the calculation of payoffs takes grammatical prominence into account.

We have assumed that the game trees are associated with probability mass functions, ρ and ρ' . In fact, these probabilities are crucial in working out the Pareto-Nash equilibria of the games. In general, when there are n discourse entities to choose from, we will have n game trees whose roots are associated with probability mass functions ρ_1, \dots, ρ_n , each ρ_i would correspond to the case where discourse entity i is taken as the intended referent of the speaker. The game tree associated with each ρ_i would have 2^k branches, where k is the number of discourse anaphors in the expression since for each discourse anaphor, the speaker can select either a pronoun or a definite description.

We will suppose that the payoffs are structured to reflect the hierarchy in (10), where the grammatical functions in the ranking refer to the grammatical function played by the phrase that refers to the discourse entity in the sentence preceding the current sentence. The probability mass function ρ_1 is associated with an information state where the speaker intends the subject of the preceding sentence is selected as the target of a discourse anaphor or definite description, ρ_2 is associated with an information state where the indirect object is most prominent and so forth. Consider the ranking in (10):⁸

(10) Subject > Indirect Object > Direct Object > Others

Notice, though, that the relative prominence of elements is reflected in the payoffs of the game. The idea is that violating the ranking in (10) would carry a cost that is directly reflected in the payoffs given to the players. We will maintain this approach, although the ranking in (10) may be subject to linguistic variation (see Prasad, 2003, and the references cited there). This approach suggests that cases where the strategy profile provided by the Pareto-Nash equilibrium has apparently been violated are due to the fact that the probabilities associated with the information states have changed due to conditioning from other information sources.

2. Extending the analysis

In section 1 we laid out the basic analysis of discourse anaphora. In this section we turn to some extensions of the basic analysis that show how games of partial information can be used to account for a variety of interesting phenomena. Needless to say that, given the available space, we cannot account for every conceivable example. But we can, nevertheless, consider some interesting examples of discourse anaphora and consider their implications for the model.

We considered, earlier, small texts like the following:

(11) A man spotted a boy. The boy kicked him.

where the second sentence has a definite description, referring to the object of the preceding sentence, in subject position and a pronoun, referring to the subject of the preceding sentence, in object position. As

⁸ The ranking in (10) is the same as is assumed in much of Centering Theory. It is also in accord with our assumption that grammatical function correlates with ease of pronominalization. Centering theory is discussed in Joshi & Weinstein (1981), Grosz, Joshi & Weinstein (1983), Grosz, Joshi & Weinstein (1986) and Walker & Prince (1996) among many other sources.

we saw, this option was part of the strategy profile of the Pareto-Nash equilibrium.

The speaker could have initiated the discourse by using a passive in the first sentence:

(12) A boy was spotted by a man.

Although the passive form encodes the same meaning as the first sentence in (11), it should have a smaller payoff than the active since it is longer. Suppose, however, that the following sentence is:

(13) He kicked him.

This sentence has the same interpretation as the second sentence in the sequence in (11)—the boy is taken as the kicker and the man is the recipient of the kick. As we have seen, the sentence in (13) yields a higher payoff than the second sentence in (11) since it uses two pronouns in place of a pair consisting of a definite description and a pronoun.

This contrast between the text in (11), on the one hand, and the text consisting of (12) followed by (13) suggests that speakers could strategically plan texts in such a way that they will accept a lower short term payoff in order to achieve a higher payoff overall.

We can think of a text as consisting of an iterated sequence of games where the same players play against each other on a sequence of different games; in other words, each sentence in the text would correspond to a game, the games differing since the sentences themselves differ. In addition, we have seen that the expected payoffs of the simple discourse anaphora games will be sensitive to the grammatical prominence of the elements in the discourse model. Thus, the games are linearly dependent on each other. The probabilities and payoffs of one game are conditioned by the preceding game and, in turn, condition the following game.

We suggest, then, that speakers think strategically about the sequence of games, choosing a form that will maximize their payoffs in the long run. Thus, the sequence of (12) and (13) establishes the boy as the most prominent element in the sequence but it does so at a cost in the initial sentence. The text in (11) begins by establishing the man as the most prominent element in the first sentence and then switches to the boy in the second sentence, incurring the cost of using a definite description. Compare that text with the following:

(14) A man saw a boy. He was/got kicked by him.

In (14), the second sentence is passivized. This structure is presumably a more costly form, but it maintains the man in the most prominent position so that, if the text is continued by:

(15) He shouted obscenities.

the strategy profile predicts that the man should be the antecedent of the pronoun.

We do not, as yet, have a mathematical theory of iterated linearly dependent games, though we believe that such a theory will be of interest both to linguists and game theorists. We should be careful to distinguish these iterated games from repeated games. In the latter case, the same players face each other in the same game; for example, they might be playing a Prisoner's Dilemma game. Repeated games have generated an extensive literature in game theory.

It is tempting to try to analyze the problem by representing the entire text as a super-game. That is, the game is to choose between sequences of games, the Nash equilibria consisting of those sequences of games that yield the best payoffs for the speaker and hearer. We believe, however, that this line of attack is far too complex to yield much insight into the process. We leave this, for the moment, as an open research problem.

Let us turn, now, to some other factors that can influence the probabilities and payoffs that influence the strategies. First consider the following well-known contrast:

- (16) a. John called Bill a Republican. Then he insulted him.
 b. John called Bill a Republican. Then *he* insulted *him*.

We intend, by the above, that the second sentence in (16)a be spoken with relatively unmarked, flat intonation while the second sentence in (16)b be spoken with heavy, contrastive stress on the pronouns.

The partial information game for (16)a is straightforward and can be reconstructed from the partial game shown in figure 2. Assuming that *John* denotes d_1 in the discourse model and *Bill* denotes d_2 , I have shown the game in figure 3 for convenience. It is easy to demonstrate that we predict that *he* in (16)a should refer to John and *him* should refer to Bill, given $\rho = \rho'$.

The partial game for (16)b should be identical to that for (16)a, the sole difference is that the contrastive stress changes the probabilities associated with the information states s and s' , so that $\rho' \gg \rho$ in figure 3.⁹ In this case, the preferred interpretation would be that *he* refers to Bill and *him* refers to John, with the implicature that calling someone a Republican is an insult. Contrastive stress has the result of

⁹ Bill Labov has observed to us that the production of stress would require effort and, thus, might reduce the payoffs slightly. We will ignore this possibility, here, since it does not alter the outcome of the game if the payoffs are reduced across the board for all uses of the stressed pronouns.

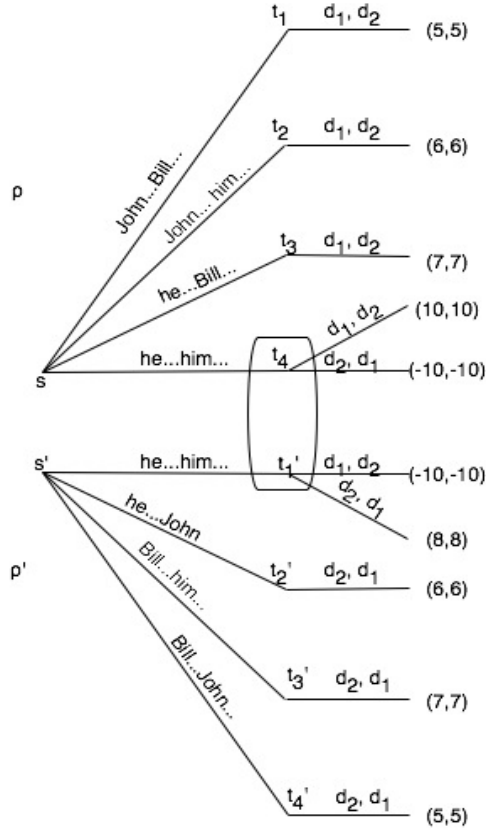


Figure 3. The Effect of Stress on Games

reordering the likelihoods of the trees rooted at information states s and s' ; for example, contrastive stress might reverse the ordering. We can think of these probabilities as being conditioned by various factors. Example (16)b shows that contrastive stress on the pronouns is one conditioning factor:

$$\begin{aligned}\rho &= P\{s \mid \text{he bears contrastive stress}\} \\ \rho' &= P\{s' \mid \text{he bears contrastive stress}\} \\ \rho' &\gg \rho\end{aligned}$$

Notice that ρ' must be strictly greater than ρ ; the difference must be sufficient to offset the slight preference for pronominalizing the subject of the preceding clause that we have built into the payoffs of the game. The fundamental idea, though, is that contrastive stress is an information source that alters the subjective probabilities of the information states. Both the speaker and the hearer know that contrastive stress

does this, so that the speaker can use contrastive stress to signal this change in the likelihoods of the information states.

The example of contrastive stress suggests that the probabilities associated with various information states can be conditioned by a variety of sources of information. Intuitively, contrastive stress signals that s' is more likely than s and this has a corresponding effect on the Pareto-Nash equilibrium. Consider, in this light, the following examples:¹⁰

- (17) a. John can open Bill's safe. He knows the combination.
b. John can open Bill's safe. He should change the combination.

The first sentence in both examples is the same. All else being equal, we should construct a game tree equivalent to the one we have already seen in figure 1. The first sentence establishes three possible antecedents: John, Bill and Bill's safe. Since only John and Bill are animate, they are the only possible antecedents for the pronoun, *he*, so we need not consider the safe. The question, now, is what values to give to ρ and ρ' , the probabilities associated with the root information states. In particular, is $\rho > \rho'$ or vice versa?

In (17)a, our unadorned model gives the correct result. Example (17)b, on the other hand, shows that lexical information and world knowledge can be used to tune these probabilities. Presumably, Bill has an interest in keeping John out of his safe and so Bill would be well advised to change the combination to the safe. These considerations result in the probabilities associated with the root information states being adjusted, causing a corresponding change in the Pareto-Nash equilibrium; that is, the discourse entity corresponding to *Bill* is preferentially chosen to interpret the pronoun. Note, however, that the following discourse is slightly stilted and unnatural:

- (18) John can open Bill's safe. John should change the combination, just to teach Bill a lesson.

This result may seem odd at first, but notice that there is a shorter and, presumably, less costly way to encode the meaning in (18):

- (19) John can open Bill's safe and should change the combination, just to teach Bill a lesson.

The availability of coordination as in (19) presumably makes (18) a less than optimal choice for encoding the meaning.

Similar considerations apply to the example in (20):

- (20) Mary insulted Sue. She slugged her.

¹⁰ We draw these examples, and many of those following, from Breheny (2002).

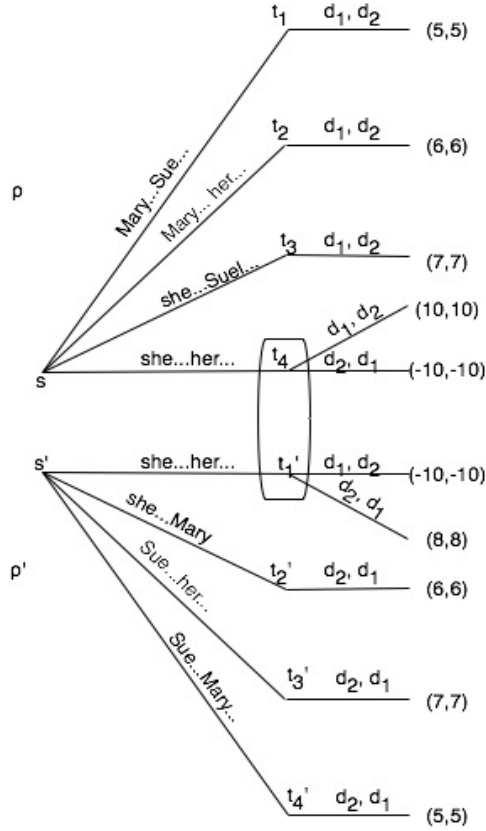


Figure 4. The contribution of lexical semantics

In the case of example (20) two discourse entities, Mary and Sue, are introduced by the first sentence. All else being equal, we would expect *Mary* to be a likelier antecedent for *she* and *Sue* to be a likelier antecedent for *her* due to their relative prominence in the preceding sentence; that is, we would expect $\rho \geq \rho'$ in figure 4. The lexical semantics of both *insult* and *slug* intervene, however, in the calculation of likelihoods for a given information state. In general, it is plausible to suppose that someone might slug a person who had insulted her, although someone might insult and then slug someone else. In other words, the object of *insult* is a likely subject of *slug* (particularly if the subject of *insult* is the object of *slug*). We can think of this in terms of conditional probabilities. Let V_k be the main verb of the k th utterance and V_{k-1} be the main verb of the preceding utterance; we then have:

$$\begin{aligned}\rho &= P\{s|V_k = \text{slug} \wedge V_{k-1} = \text{insult}\} \\ \rho' &= P\{s'|V_k = \text{slug} \wedge V_{k-1} = \text{insult}\}\end{aligned}$$

Given the conditioning by *insult* and *slug*, we find that $\rho' \gg \rho$ so that the Pareto-Nash equilibrium becomes:

$$(21) \quad \{s, \text{Mary} \dots \text{her} \dots\}, (s', \text{she} \dots \text{her} \dots), (\{t_4, t'_1\}, \langle d_2, d_1 \rangle)\}$$

That is, the best choice for the hearer, given conditioning by the lexical semantics of the verbs, would be to interpret *she* as Sue and *her* as Mary.

Of course, other words can condition the probabilities of the information states. Consider, for example, discourse connectives (Miltsakaki, 2003) like *then*:

$$(22) \quad \text{Mary insulted Sue. Then, she slugged her.}$$

In the case of *then*, part of its work is to increase the probability of information state *s*, even given the presence of *insult* and *slug*. Given this, the pronouns in the second sentence of (22) are best interpreted as *she* = Mary and *her* = Sue.

The situation is quite different with the connective *so*; *so* conditions the probabilities of *s* and *s'* (referring again to figure 4) so that $\rho' \gg \rho$, giving the strategy profile in (21). Thus, in (23), Sue is a better antecedent for *she* than Mary is:

$$(23) \quad \text{Mary insulted Sue. So she slugged her.}$$

In other words, *so* conditions the information state that the game players judge to be likeliest. Notice that the conditioning is sufficiently great that the following seems rather bizarre:

$$(24) \quad \text{Mary insulted Sue. So Mary slugged her.}$$

That is, my insulting someone is not generally taken as sufficient reason for me then to slug them.

We will not pursue the issue of discourse connectives further. The general approach to them in the game framework seems clear: discourse connectives are important in conditioning the likelihood of information states which, in turn, affects the Pareto-Nash equilibrium of the discourse anaphora games.

So far, our estimates of the likelihood of various information states have rested on interactions between syntactic structures and lexical semantics. There are cases where neither the syntactic structure nor the lexical semantics influence the assessment of likelihoods in a decisive manner. Consider the following example:

- (25) John's spaghetti spilled on Bill's jacket. He didn't notice.

Our judgment is that either John or Bill can be taken as the antecedent for the pronoun in the second clause. Notice that both *John* and *Bill* are genitives in the first clause and, thus, stand outside the ranking in *ranking*, above. Neither John nor Bill is more prominent than the other, so the payoffs are entirely symmetrical. Since neither possible antecedent for the pronoun can be ranked and the lexical semantics provides no clues, we assume that the probability of information state s and the probability of information state s' are essentially indistinguishable. This in turn suggests that the best that the players can do is to adopt a mixed strategy so that the anaphor in (25) is truly indeterminate and the sentence is truly ambiguous (see Parikh, 2006, for a discussion of indeterminacy within the game theoretic framework). Notice that the indeterminacy could be resolved by a further sentence:

- (26) John's spaghetti spilled on Bill's jacket. He didn't notice. His jacket, however, was ruined.

Although the second is indeterminate, the third sentence entirely resolves the ambiguity. Notice that this happens without any sense of a garden path, indicating that both possibilities were available (via a mixed strategy).

Consider, next, a case where the conditioning information is redundant. An example would be a discourse connective like *so* combined with contrastive stress:

- (27) Mary insulted Sue. So *shé* slugged *hér*.

Example (27) strikes us as peculiar. In this case, the discourse connective *so* and the contrastive stress both condition the probabilities in the same direction. This redundant marking seems to be self-defeating; the hearer can only speculate as to why the speaker would choose to heavily mark the same contrast.

Another interesting case is where the lexical information and contrastive stress are at cross purposes:

- (28) John can open Bill's safe. *Hé* should change the combination.

It seems to us that this example is more acceptable than the peculiar (27), with the interpretation that John should change the combination of Bill's safe. It would seem that contrastive stress can trump the rather weak conditioning by the lexical semantics in this case, although both conditioning factors are present:

- (29) $\rho = P\{s | \text{hé} \wedge V_k = \text{change} \wedge V_{k-1} = \text{open} \dots\}$

We leave it to the reader to construct further examples and judge their effects.

Finally, let's consider a case where world knowledge over-rides the rankings in (10).¹¹ Suppose that Bill has been invited swimming by his friend John. Bill asks his father who, in turn consults with Bill's mother. It seems that Bill's father can use either of the following sentences:

- (30) a. John wants to take Bill swimming.
b. Bill wants to go swimming with John.

Bill's mother can respond to either (30)a or (30)b with:

- (31) He's sick.

Meaning, in either event, that Bill is sick. How can this be? The main verb in both (30)a and (30)b is the same, the choice of subject being contingent on the structure of the embedded predicate. We can think of no reason why either *go* or *take* should condition the best interpretation of the pronoun in (31). Since lexical semantics does not seem to help, we appear to predict that grammatical structure should be decisive. This would predict, incorrectly, that *he* should be interpreted as John in response to (30)a and as Bill in response to (30)b.

Notice, however, that Bill's father is faced with a decision problem: Should he allow Bill to go swimming with John? Bill's mother supplies a piece of information that, when interpreted correctly, allows his father to solve the decision problem. Both of Bill's parents recognize this; his mother, in particular, knows that providing the information that Bill is sick will solve the problem and she knows that Bill's father knows this. Providing the information that John is sick is not so decisive. Thus, faced with the decision problem at hand, the utility of both parents is served by interpreting *he* as Bill.

We have seen, in this section and section 1, that the game trees for discourse anaphora are established by the following conditions:

1. The grammatical roles carried by various expressions in the preceding sentence, which affects the payoffs,
2. Lexical semantics, which affects the probabilities associated with the various information states,
3. Contrastive stress, which likewise affects the probabilities of the information states,
4. Decision problems, which may affect both probabilities and payoffs.

¹¹ This example was pointed out to Parikh by Richard Breheny (p.c.)

When more than one factor is present, it may be unclear how they interact, thus making the resolution of the anaphor genuinely indeterminate, something that might be represented by a mixed strategy. It is also possible that different speakers and addressees would have different assessments of these conditioning factors. In this case, we might expect miscommunication or partial communication to occur. Empirically, then, one would expect speakers to avoid multiple conditioning factors from influencing their choices. Rather than further refining the framework, we turn to a summary and discussion of its properties.

3. Discussion

We have presented the basic elements of a theory of discourse anaphora. The elements of the theory are presented in section 1 where we gave elementary game trees which formalize the strategic interaction between the speaker and the hearer. We have given basic trees for one and two discourse anaphors, given two possible discourse antecedents. We have also suggested how these game trees can be extended for an arbitrary number of potential discourse antecedents. The game trees can also be extended to cover more discourse anaphors within an utterance.

In developing the game trees, we have developed payoffs for the players. These payoffs represent a system of inequalities that correspond to preferences on the parts of the players. Thus, pronouns are preferred to longer expressions with greater processing loads and successful identification of the referent of the expression is preferred to miscommunication. We have associated these payoffs with integer values but we should note that any assignment of values will do so long as the inequalities are respected; that is, the preference ordering remains invariant up to affine transformations. In brief, the reader should refrain from assigning too much importance to the actual numerical values given here; the theory lies in the inequalities.

Finally, each root information state is associated with a probability. As we saw in section 2, these probabilities could be conditioned by a number of information sources: Grammatical function of an element in the preceding sentence, contrastive stress, lexical semantics and decision problems. Although the model admits some degree of freedom, it is, nevertheless, constrained by the form of the game trees and the payoffs associated with the various choices. Once the model is spelled out, many a priori possible outcomes are ruled out.

Given the game trees, payoffs and probabilities, it is straightforward to compute the Pareto-Nash Equilibrium of the game. The result-

ing strategy profile corresponds to the possible patterns of discourse anaphora. Thus, the model we have described here relates discourse anaphora to rational choice. It seems to us that analytic game theory provides a method for exploring and formalizing *linguistic choice*, an idea that was important in Saussure (1916/1972).¹² The study of choice as part of the study of language has been largely neglected because the problem of choice was thought to lie outside the grammar. Game theory, however, provides us with a formal language for seamlessly linking the statistical aspect of language with its algebraic aspect. It seems to us that this makes the game theoretic approach outlined here genuinely different from other approaches like Centering Theory or Relevance Theory (see Sperber & Wilson, 1995).

There are other differences between analytic game theory and other theories. Most importantly, the system developed in this paper is grounded in a theory of common knowledge. The participants in a discourse are able to communicate efficiently—in this case, track the referents of discourse anaphors—because they share the same information, which we have encoded in the game tree, the probabilities and the payoffs. Thus, the theory is normative in the sense that it predicts how participants in a discourse ought to behave in order to maximize their utility, where utility is measured in terms of accuracy and efficiency in transmitting information. The theory does not, however, characterize the mental computations that speakers and hearers might go through in approximating these strategies. Thus, the theory does not provide rules which are intended to correspond to steps in a mental calculation, nor does it make claims about mental representations.

Crucially, the game tree is part of public knowledge, available to anyone with knowledge of the language who has been following the discourse and not a mental object. The problem of how speakers and hearers approximate the information contained in the game tree would be the domain of psycholinguistics. The theory presented here is a social theory of information exchange grounded in rational choice. We should note, however, that we are not committed to the view that the participants in the game are “hyper-rational;” our view is consistent with models of bounded rationality (see Rubinstein (1998)) and other models grounded in behavioral game theory (see Camerer (2003)).

A number of open questions remain. As noted briefly above, we have not considered typological variation in this account. Thus, we have put aside the question of whether the ranking strategy in (10) can vary across languages, although it surely does. Our inclination

¹² See, for example, Chapter IV in Part 2 of the *Cours* where he discusses the idea of exchanging one element for another in a linguistic system.

would be to treat this variation at the level of the payoffs of the game, as we did in section 1. We have also not considered languages with phonologically empty pronominals, although it is clear how to extend the game trees in section 1 to do so. Finally, we have only considered inter-sentential anaphora. We have not considered coreference relations within a sentence. Again, we will leave this problem for future research. Even without these extensions, it is clear that analytic game theory provides a promising framework for the analysis of anaphora in particular and meaning in general.

References

- Breheny, R. Pragmatic analyses of anaphoric pronouns: Do things look better in 2-d?. Manuscript, RCEAL, University of Cambridge, 2002.
- Camerer, C. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton, NJ, 2003.
- CLARK, R. Games, quantification and discourse structure. In *Logic, Games and Philosophy: Foundational Perspectives*, A.-V. Pietarinen, Ed. Kluwer Academic Publishers, in press.
- DE SAUSSURE, F. *Cours de linguistique générale*. Éditions Payot, Paris, 1916/1972.
- DEKKER, P. Grounding dynamic semantics. In *Descriptions and Beyond*, M. Reimer and A. Bezuidenhout, Eds. Oxford University Press, 2004, pp. 484–502.
- GROENENDIJK, J., AND STOKHOF, M. Dynamic predicate logic. *Linguistics and Philosophy* 14 (1991), 39–100.
- GROSZ, B., JOSHI, A., AND WEINSTEIN, S. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the ACL* (1983), pp. 44–50.
- GROSZ, B., JOSHI, A., AND WEINSTEIN, S. Towards a computational theory of discourse interpretation. Unpublished manuscript, 1986.
- JOSHI, A., AND WEINSTEIN, S. Control of inference: Role of some aspects of discourse structure-centering. In *Proceedings of the International Joint Conference on Artificial Intelligence* (1981), pp. 385–387.
- KAMP, H., AND REYLE, U. *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1993.
- MILTSAKAKI, E. The syntax-discourse interface: Effects of the main-subordinate distinction on attention structure. Doctoral Dissertation, University of Pennsylvania, 2003.
- MYERSON, R. B. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, MA, 1991.
- PARIKH, P. *The Use of Language*. CSLI Publications, Stanford, CA, 2001.
- PARIKH, P. Radical semantics: A new theory of meaning. *Journal of Philosophical Logic* (Forthcoming).
- PARIKH, P., AND CLARK, R. The meaning of *THE*: A new account of definite descriptions. manuscript, University of Pennsylvania, 2005.
- PIETARINEN, A.-V. Semantic games and generalised quantifiers. manuscript, University of Helsinki, 2004.
- PRASAD, R. Constraints on the generation of referring expressions, with special reference to hindi. Doctoral Dissertation, University of Pennsylvania, 2003.

- ROBERTS, C. Pronouns as definites. In *Descriptions and Beyond*, M. Reimer and A. Bezuidenhout, Eds. Oxford University Press, 2004, pp. 503–543.
- RUBINSTEIN, A. *Modelling Bounded Rationality*. The MIT Press, Cambridge, MA, 1998.
- SPERBER, D., AND WILSON, D. *Relevance: Communication and Cognition*, 2nd ed. Basil Blackwell, London, 1995.
- STALNAKER, R. On the representation of context. *Journal of Logic, Language and Information* 7 (1998), 3–19.
- VAN EIJCK, J., AND KAMP, H. Representing discourse in context. In *Handbook of Logic and Language*, J. van Benthem and A. ter Meulen, Eds. The MIT Press, Cambridge, MA, 1997, pp. 179–237.
- WALKER, M., AND PRINCE, E. A bilateral approach to givenness: A hearer-status algorithm and a centering algorithm. In *Reference and referent accessibility*, T. Fretheim and J. Gundel, Eds. John Benjamins, Amsterdam/Philadelphia, 1996, pp. 291–306.

Address for Offprints: Department of Linguistics
University of Pennsylvania
Philadelphia, PA 19104
USA