

# Issues in Communication Game

HASIDA Kôiti

Electrotechnical Laboratory

1-1-4 Umezono, Tukuba, Ibaraki 305, Japan.

hasida@etl.go.jp

## Abstract

As interaction between autonomous agents, communication can be analyzed in game-theoretic terms. *Meaning game* is proposed to formalize the core of intended communication in which the sender sends a message and the receiver attempts to infer its meaning intended by the sender. Basic issues involved in the game of natural language communication are discussed, such as salience, grammaticality, common sense, and common belief, together with some demonstration of the feasibility of game-theoretic account of language.

## 1 Introduction

Communication is a game (interaction among autonomous agents) by definition. So it can be analyzed in game-theoretic terms.<sup>1</sup> In this paper we study a fundamental aspect of linguistic communication from the point of view of game theory, and enumerate some basic issues involved in the communication games of natural language.

Let  $I$  be the proposition that the sender  $S$  intends to communicate a semantic content  $c$  to the receiver  $R$ . Then  $I$  entails that  $S$  intends that  $R$  should both recognize  $c$  and believe  $I$ . This is the core of *nonnatural meaning* (Grice, 1957; Grice, 1969). Grice's original notion of nonnatural meaning further entails ( $S$ 's intention of)  $R$ 's believing (when  $c$  is a proposition or a reference) or obeying (when it is an order or a request)  $c$ , but we disregard this aspect and concentrate on this core.

This restricted sense of nonnatural meaning implies that communication is inherently collaborative, because both  $S$  and  $R$  want that  $R$  should recognize  $c$  and  $I$ .  $S$  of course wants it, and so does  $R$  because it is beneficial in general to know what  $S$  intends to make  $R$  believe or obey.  $S$  might be lying or trying to mislead  $R$ , but even in

<sup>1</sup>See Osborne and Rubinstein (1994), among others, for general reference on game theory.

such a case  $S$  is still intending to communicate a content  $c$  by way of making  $R$  recognize this intention. Even if  $R$  doubts  $S$ 's honesty,  $R$  will try to know what  $c$  is, because knowing what  $c$  is would help  $R$  infer what the hidden intent of  $S$  may be, among others. For instance, when  $S$  tells  $R$  that it is raining,  $R$  will learn that  $S$  wants to make  $R$  believe that it is raining.  $R$  would do so even if  $R$  knew that it is not raining. Even if  $R$  were unsincere and misunderstood  $S$ 's message on purpose,<sup>2</sup> the nonnatural meaning is still properly conveyed, because otherwise the intended misunderstanding would be impossible.

The present study concerns this aspect of communication, the nonnatural meaning in the restricted sense, which is a core of intended communication. Lies, ironies, indirect speech acts, and so forth (Perrault, 1990; Perrault and Allen, 1980) all share this core. Our understanding about it will hence help us understand basic workings of natural communication systems. As an example, centering theory (Grosz et al., 1995) could be attributed to game-theoretic accounts, as demonstrated later in this paper.

## 2 Communication Games

Communication has been discussed in the game-theory literature. A *signaling game* consists of sender  $S$ 's sending a *message* (or a *signal*) to receiver  $R$  and  $R$ 's doing some *action* in response to that message. Here  $S$  knows something that  $R$  did not know before receiving the message. This is formulated by assuming that  $S$  belongs to some *type*, which  $S$  knows but  $R$  does not know at first. Let  $T$  be the set of the types,  $P$  be the probability distribution over  $T$ . Let  $M$  be the set of the messages and  $A$  be the set of  $R$ 's possible actions. Finally, let  $U_X$  be the *utility function* for player  $X$ .  $U_S(t, m, a)$  and  $U_R(t, m, a)$  are real numbers for  $t \in T$ ,  $m \in M$  and  $a \in A$ . A signaling game with  $T = \{t_1, t_2\}$ ,  $M = \{m_1, m_2\}$  and  $A = \{a_1, a_2\}$  is illustrated by a *game tree* as shown

<sup>2</sup>If  $R$  is sincere and unintentionally misunderstands, that is just a failure of sharing the same context with  $S$ .

in Figure 1. Here the game proceeds downwards.

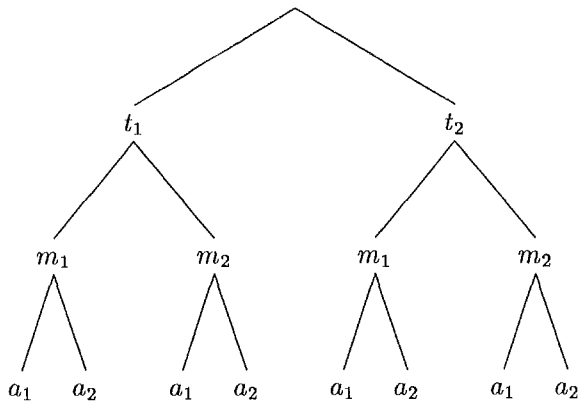


Figure 1: A signaling game.

The top branch is the nature's initial choice of  $S$ 's type according to  $P$ , the middle layer is  $S$ 's decision on which message to send, and finally the bottom layer is  $R$ 's choice of her action. When  $R$  has just received  $m_i$  ( $i = 1, 2$ ), she does not know whether the game has been played through  $t_1$  or  $t_2$ .

Let  $\sigma_S$  and  $\sigma_R$  be  $S$ 's and  $R$ 's strategies,<sup>3</sup> respectively. That is,  $\sigma_S(m|t)$  is the conditional probability of  $S$ 's sending message  $m$  provided that she is of type  $t$ , and  $\sigma_R(a|m)$  the conditional probability of  $R$ 's doing action  $a$  provided that she has received  $m$ . The combination  $\langle \sigma_S, \sigma_R \rangle$  of strategies is an *equilibrium*<sup>4</sup> of a signaling game when  $\sigma_S$  and  $\sigma_R$  are the optimal responses to each other; that is, when  $\sigma_X$  maximizes  $X$ 's expected utility

$$\sum_{t,m,a} P(t) \sigma_S(m|t) \sigma_R(a|m) U_X(t, m, a)$$

given  $\sigma_Y$ , for both  $X = S \wedge Y = R$  and  $X = R \wedge Y = S$ .

In typical applications of signaling game,  $T$ ,  $M$  and  $A$  are not discrete sets as in the above example but connected subsets of real numbers, and  $S$ 's preference for  $R$ 's action is the same irrespective of her type. In this setting,  $S$  should send a costly message to get a large payoff. For instance, in job market signaling (Spence, 1973), a worker  $S$  signals her competence (type) to a potential employer  $R$  with the level of her education as the message, and  $R$  decides the amount of salary to offer to  $S$ . A competent worker will have high education and the employer will offer her a high salary. In mate selection (Zahavi, 1975), a deer  $S$  indicates its strength by the size of its antlers

<sup>3</sup>Or *mixed strategies*, which are probability distributions over the *simple strategies* (actions).

<sup>4</sup>Or *complete Bayesian equilibrium*, in a more precise, technical term.

to potential mates  $R$ . A strong deer will grow extra large antlers to demonstrate its extra survival competence with this handicap.

*Cheap-talk game* is another sort of communication game. It is a special case of signaling game where  $U_S$  and  $U_R$  do not depend on the message; that is, composing/sending and receiving/interpreting message are free of cost. In a cheap-talk game,  $S$ 's preference for  $R$ 's action must depend on her type for non-trivial communication to obtain, because otherwise  $S$ 's message would give no information to  $R$  about her type.

### 3 Meaning Game

Now we want to formulate the notion of *meaning game* to capture nonnatural meaning in the restricted sense discussed in Section 1. Let  $C$  be the set of semantic contents and  $P$  the probability distribution over the linguistic reference to the semantic contents. That is,  $P(c)$  is the probability that  $S$  intends to communicate semantic content  $c$  to  $R$ . As before,  $M$  is the set of the messages. A meaning game addresses a *turn* of communication  $\langle c_S, m, c_R \rangle$ , which stands for a course of events where  $S$ , intending to communicate a semantic content  $c_S$ , sends a message  $m$  to  $R$  and  $R$  interprets  $m$  as meaning  $c_R$ .  $c_S = c_R$  is a necessary condition for this turn of communication to be successful. It seems reasonable to assume that the success of communication is the only source of positive utility for any player.

So a meaning game might be a sort of signaling game in which  $S$ 's type stands for her intending to communicate some semantic content, and  $R$ 's action is to infer some semantic content. That is, both  $T$  and  $A$  could be simply regarded as  $C$ . Strategies  $\sigma_S$  and  $\sigma_R$  are defined accordingly.

In a simple formulation, the utility function  $U_X$  of player  $X$  would thus be a real-valued function from  $C \times M \times C$  (the set of turns). It would be sensible to assume that  $U_X(c_S, m, c_R) > 0$  holds only if  $c_S = c_R$ .  $U_X$  reflects the grammar of the language (which might be private to  $S$  or  $R$  to various degrees). The grammar evaluates the (computational, among others) cost of using content-message pairs. The more costly are  $\langle c_S, m \rangle$  and  $\langle m, c_R \rangle$ , the smaller is  $U_X(c_S, m, c_R)$ . The notion of equilibria in a meaning game is naturally derived from that in a signaling game.

If the players want something like *common belief*,<sup>5</sup> however, meaning games are not signaling games. This is because  $c_S = c_R$  is not a sufficient condition for the success of communication in that case.  $U_X$  should then depend on not just  $c_S$ ,  $m$ , and  $c_R$ , but also the players' nested beliefs

<sup>5</sup>People have common belief of proposition  $p$  when they all believe  $p$ , they all believe that they all believe  $p$ , they all believe that they all believe that they all believe  $p$ , and so on, ad infinitum.

about each other. We will come back to this issue in Section 4.

Note also that the typical instances of meaning game in natural language communication is not like the typical applications of signaling game such as mentioned before, even if meaning games are special sort of signaling games. That is, meaning games in natural language would normally involve discrete sets of semantic contents and messages.

Natural-language meaning games are not cheap-talk games, either, because we must take into consideration the costs of content-message pairs. It is not just the success of communication but also various other factors that account for the players' utility.  $S$  and  $R$  hence do not just want to maximize the probability of successful communication.

To illustrate a meaning game and to demonstrate that meaning games are not cheap-talk games, let us consider the following discourse.

- (1)  $u_1$ : Fred scolded Max.  
 $u_2$ : He was angry with the man.

The preferred interpretation of 'he' and 'the man' in  $u_2$  are Fred and Max, respectively, rather than the contrary. This preference is accounted for by the meaning game as shown in Figure 2. In this

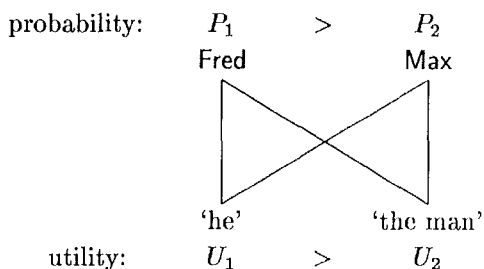


Figure 2: A meaning game about references of NPs.

game, Fred and Max are semantic contents, and 'he' and 'the man' are messages.<sup>6</sup> We have omitted the nature's selection among the semantic contents. Also, the nodes with the same label are collapsed to one.  $S$ 's choice goes downward and  $R$ 's choice upward, without their initially knowing the other's choice. The complete bipartite connection between the contents and the messages means that either message can mean either content grammatically (without too much cost).

$P_1$  and  $P_2$  are the prior probabilities of references to Fred and Max in  $u_2$ , respectively. Since Fred was referred to by the subject and Max by the object in  $u_1$ , Fred is considered more salient than Max in  $u_2$ . This is captured by assuming  $P_1 > P_2$ .  $U_1$  and  $U_2$  are the utility (negative

<sup>6</sup>Perhaps there are other semantic contents and messages.

cost) of using 'he' and 'the man,' respectively.<sup>7</sup> Utilities are basically assigned to content-message pairs, but sometimes it is possible to consider costs of messages irrespective of their contents. We assume  $U_1 > U_2$  to the effect that 'he' is less complex than 'the man' both phonologically and semantically; 'he' is not only shorter than 'the man' but also, more importantly, less meaningful in the sense that it lacks the connotation of being adult which 'the man' has.

There are exactly two equilibria entailing 100% success of communication, as depicted in Figure 3 with their expected utilities  $\mathcal{E}_1$  and  $\mathcal{E}_2$  apart from the utility of success of communication.<sup>8</sup>  $P_1 > P_2$

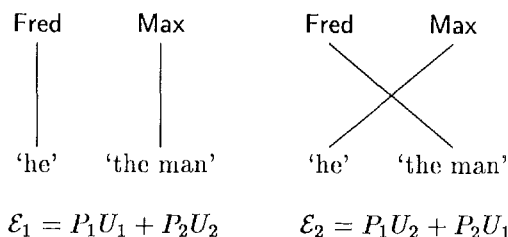


Figure 3: Two equilibria of the meaning game in Figure 2.

and  $U_1 > U_2$  imply  $\mathcal{E}_1 - \mathcal{E}_2 = (P_1 - P_2)(U_1 - U_2) > 0$ . So the equilibrium in the left-hand side is preferable for both  $S$  and  $R$ , or *Pareto superior*. This explains the preference in (1). It is straightforward to generalize this result for cases with more than two contents and messages: A more salient content should be referred to by a lighter message when the combinations between the contents and the messages are complete. A general conjecture we might draw from this discussion is the following.

- (2) Natural-language meaning games are played at their Pareto-optimal equilibria.

An equilibrium is *Pareto optimal* iff no other equilibrium is Pareto superior to it.

Note that we have derived an essence of centering theory (Joshi and Weinstein, 1981; Kameyama, 1986; Walker et al., 1994; Grosz et al., 1995). Centering theory is to explain anaphora in natural language. It considers list  $Cf(u_i)$  of *forward-looking centers*, which are the semantic entities *realized*<sup>9</sup> in  $u_i$ , where  $u_i$  is the  $i$ -th utterance. The forward-looking centers of utterance  $u$

<sup>7</sup>For the sake of simplicity, here we assume that  $U_S$  and  $U_R$  are equal. See Section 4 for discussion.

<sup>8</sup>Common belief about the communicated content is always obtained in both cases. So the current discussion does not depend on whether the success of communication is defined by  $c_S = c_R$  or common belief.

<sup>9</sup>A linguistic expression *realizes* a semantic content when the former directly refers to the latter or the situation described by the former involves the latter.

are ranked in  $Cf(u)$  according to their saliences. In English, this ranking is determined by grammatical functions of the expressions in the utterance, as below.

subject > direct object > indirect object  
> other complements > adjuncts

The highest-ranked element of  $Cf(u)$  is called the *preferred center* of  $U$  and written  $Cp(u)$ . *Backward-looking center*  $Cb(u_i)$  of utterance  $u_i$  is the highest-ranked element of  $Cf(u_{i-1})$  that is realized in  $u_i$ .  $Cb(u)$  is the entity which the discourse is most centrally concerned with at  $u$ .

Centering theory stipulates the following rule.

- (3) If an element of  $Cf(u_{i-1})$  is realized by a pronoun in  $u_i$ , then so is  $Cb(u_i)$ .

In (1),  $Cb(u_2) = \text{Fred}$  because  $Cf(u_1) = [\text{Fred}, \text{Max}]$ , if either 'he' or 'the man' refers to Fred. Then rule (3) predicts that Fred cannot be realized by 'the man' if Max is realized by 'he' — the same prediction that we derived above. Moreover, (3) itself is a special instance of our above observation that a more salient content should be referred to by a lighter message, provided that the backward-looking center is particularly salient.

(3) is common in all the version of centering theory, but of course there are further details of the theory, which vary from one version to another. To derive all of them (which are right) in a unified manner requires further extensive study.

## 4 Playing the Same Game

We have so far assumed implicitly that  $S$  and  $R$  have common knowledge about (the rule of) the game (that is,  $P$ ,  $U_S$  and  $U_R$ ). This assumption will be justified as a practical approximation in typical applications of signaling games (and cheap-talk games). For instance, there may well be a body of roughly correct, stable common-sense knowledge about the correlation between the competence of workers and the degree of effort they make to have higher education, about how much an employer will offer to an employee with a certain competence, and so on.

However, common knowledge on the game might be harder to obtain in natural-language meaning games, because the game lacks such stability of the typical signaling games as mentioned above. A natural-language meaning game is almost equivalent to the context of discourse, which changes dynamically as the discourse unfolds.

In general, to figure out her own best strategy,  $S$  ( $R$ ) attempts to infer  $R$ 's ( $S$ 's) strategy by simulating  $R$ 's ( $S$ 's) inference. If  $S$  and  $R$  do not have common knowledge about the game, this inference will constitute an infinite tree.<sup>10</sup> For instance,

For instance, after an utterance of 'a house,' 'the door' realizes the house referred to by 'a house.'

<sup>10</sup>This is not a game tree but a tree of belief embedding.

Figure 4 depicts  $S$ 's inference when she wants to communicate  $c_1$ , where the players have common knowledge of  $C = \{c_1, c_2\}$  and  $M = \{m_1, m_2\}$  but not of their utility functions. The nodes labeled by  $c_i$  represent  $S$  when she wants to communicate  $c_i$ , and those labeled by  $m_i$  represent  $R$  when she wants to interpret  $m_i$ , for  $i = 1, 2$ . The inference by  $R$  when interpreting message  $m_i$  is a similar tree rooted by  $m_i$ .

Although it is impossible to actually have common knowledge in general (Halpern and Moses, 1990), there are several motivations for the players to pretend to have common knowledge about the game. First, they can avoid the computational complexity in dealing with infinite trees such as above. Second, common belief on the game is a simple means to obtain common belief on the communicated content. Third, the best payoff is obtained when the players have common knowledge about the game, if their utility functions are equal. In fact, the utility functions are probably equal, because language use as a whole is a *repeated game*. That is, provided that communicating agents play the role of  $S$  and  $R$  half of the time each, they can maximize their expected utility by setting their utility functions to the average of their selfish utilities. Fortunately, this equalization is very stable, as long as the success of communication is the only source of positive utility for both the players.

In communication games, common knowledge on which message  $S$  has sent should help the players converge on common belief on the game. That is, when the players have common knowledge that message  $m$  was sent, they may be able to detect errors in their embedded beliefs. In fact, an embedded belief turns out wrong if it implies  $\sigma_S(m|c) = 0$  for every  $c$  in the embedded context. This common knowledge about  $m$  may be even incorporated in the meaning game. That is, it may affect the cost of retrieving or composing various content-message pairs, thus biasing the scope of the game towards those content-message pairs closely associated with  $m$ . Contents and messages very difficult to envisage given  $m$  will be virtually excluded from the game. Once the game is defined, however, both players must take into consideration the entire maximal connected subgraph containing the content she wants to convey or the message she wants to interpret.

## 5 Composite Game

Natural-language communication is a composite game in two senses. First, as mentioned in the previous section, it is considered a *repeated game*, which is a sequence of smaller games. Second, each such smaller game is a compound game consisting of temporally overlapping meaning games. These facts introduce several complications into the communication game.

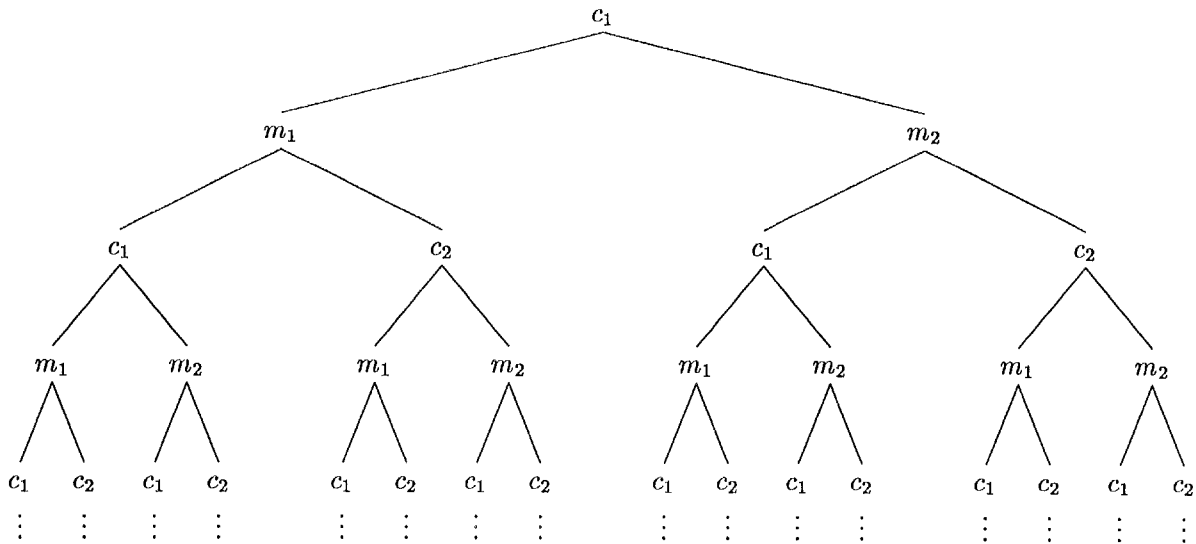


Figure 4: Inference by  $S$  to communicate semantic content  $c_1$ .

In a repeated game, one stage may affect the subsequent stage. In natural-language communication, a meaning game can influence the next meaning game. For instance, if a semantic content  $c$  is referred to by a message with a low cost, then the probability of reference to  $c$  may increase as a sort of accommodation,<sup>11</sup> because a reference by a lightweight message presupposes high prior probability of reference, as discussed in Section 3. For instance, a reference to Fred by ‘he’ will raise the salience of Fred.

Another type of contextual effect shows up the following discourse.

- (4)  $u_1$ : Fred scolded Max.  
 $u_2$ : The man was angry with him.

Here ‘the man’ and ‘he’ in  $u_2$  are more readily interpreted as Fred and Max, respectively, which violates (3) and hence our game-theoretic account.

This preference is accounted for by the preference for parallelism concerning the combination of semantic content and grammatical function: In both  $u_1$  and  $u_2$  Fred is realized by the subject NP and Max is realized by the object NP. This is the same sort of preference that is addressed by property-sharing constraint (Kameyama, 1986). This effect is attributed to the utility assignment as shown in Figure 5. That is, the utility  $U_1$  of associating the proposition  $\text{angry}(\text{Fred}, \text{Max})$  (that Fred is angry with Max) with the sentence ‘The man was angry with him’ is greater than the utility  $U_2$  of associating  $\text{angry}(\text{Max}, \text{Fred})$  (the proposition that Max is angry with Fred) with the same

<sup>11</sup>Lewis (1979) discusses several types of accommodation for conversational score, of which the most relevant here is accommodation for comparative salience:  $x$  becomes more salient than  $y$  when something is said which presupposes  $x$  to be more salient than  $y$ .

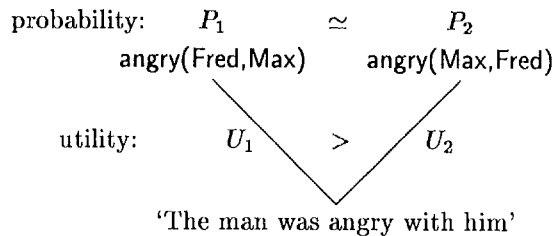


Figure 5: A meaning game about propositions and sentences.

sentence. This game might involve other possible associations such as that between  $\text{angry}(\text{Max}, \text{Fred})$  and ‘The man made him angry,’ but as mentioned at the end of Section 4 contents and messages other than included in Figure 5 probably accompany great costs and hence may be neglected.

In general, several meaning games are played possibly in parallel during linguistic communication using a compound expression. A turn of communication with an utterance of ‘the man was angry with him’ consists of the sentence-level game mentioned above, the two noun phrase-level games — one concerning the subject NP (shown in Figure 2) and the other the object NP of ‘with’ — and so on. A strategy of each player in such a compound game associated with a compound expression is a combination of her strategies for all such constituent games. Each player attempts to maximize the expected utility over the entire compound game, rather than for each constituent game.

Different constituent games often interact. For instance, if the speaker chooses to say ‘the man’ for the subject NP, then the whole sentence can-

not be ‘he was angry with the man.’ So a global solution, which maximizes the utility from the entire game, may maximize the utility from some constituent games but not from others. In the above example, the global solution, which involves saying ‘the man was angry with him’ and interpreting it as `angry(Fred,Max)`, maximizes the utility from the sentence-level game but not from the NP-level games. Incidentally, the players will gain greater utility if they use the combination of `angry(Fred,Max)` and ‘he was angry with the man,’ which is consistent with the optimal equilibrium of the NP-games. When ‘the man was angry with him’ is used despite the smaller default utility associated with it, Max will probably be assigned a greater salience than otherwise, which is again a sort of accommodation.

Extralinguistic context enters sentence-level games and plays an important role in language use. For example, if it is known that Max never gets angry and that Fred is short-tempered, then both in (1) and (4) the second utterance will preferably be interpreted as meaning `angry(Fred,Max)`.

## 6 Conclusion

*Meaning game* captures nonnatural meaning in the restricted sense which obtains in basically all the cases of natural-language communication. The factors which define a meaning game include grammatical function, reference by lightweight message, extralinguistic information (these affect *P*), grammar, cost of recalling (these affect the utility), and so on. To have a more complete game-theoretic account of natural language, we need a quantitative characterization of how those factors contribute to the game.

We have almost restricted ourselves to references of nouns phrases, but the seemingly promising targets of game-theoretic account of natural language apparently include binding theory, conversational implicature (Parikh, 1992), and so on. Since our account is very general in nature, however, it should apply to language as a whole. For instance, the notion of grammaticality may well be attributed to the computational difficulty in convergence to a common game-theoretic equilibrium. Also, abductive inference involved in language understanding (Hobbs et al., 1993) (hence in language production, too, from the game-theoretic, reciprocal viewpoint) is closely related with our theory. That is, the particular usefulness of abduction in natural language communication could be ascribed to the fact that language use in essence is a collaborative interaction such as discussed so far.

## References

- H. Paul Grice. 1957. Meaning. *Philosophical Review*, 66:377–388.
- H. Paul Grice. 1969. Utterer’s meaning and intentions. *Philosophical Review*, 68(2):147–177.
- Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.
- Joseph Y. Halpern and Yoram Moses. 1990. Knowledge and common-knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587.
- Jerry R. Hobbs, Mark E. Stickel, Douglass E. Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142.
- Aravind K. Joshi and Scott Weinstein. 1981. Control of inference: Role of some aspects of discourse structure — centering. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 385–387.
- Megumi Kameyama. 1986. A property-sharing constraint in centering. In *Proceedings of the 24th Annual Meeting of ACL*, pages 200–206.
- David Lewis. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8:339–359.
- Martin J. Osborne and Ariel Rubinstein. 1994. *A Course in Game Theory*. The MIT Press, Cambridge, Massachusetts.
- Prashant Parikh. 1992. A game-theoretic account of implicature. In *Proc. of the Fourth Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 85–94, Monterey, CA.
- C. Raymond Perrault and James F. Allen. 1980. A plan-based analysis of indirect speech act. *American Journal of Computational Linguistics*, 6(3-4):167–182.
- C. Raymond Perrault. 1990. An application of default logic to speech act theory. In Philip R. Cohen, J. Morgan, and Martha E. Pollack, editors, *Intentions in COMMUNICATION*, pages 161–185. MIT Press.
- A. Michael Spence. 1973. Job market signaling. *Quarterly Journal of Economics*, 87:355–74.
- Marilyn Walker, Masayo Iida, and Sharon Cote. 1994. Japanese discourse and the process of centering. *Computational Linguistics*, 20(2):193–232.
- A. Zahavi. 1975. Mate selection — a selection for a handicap. *Journal of Theoretical Biology*, 53:205–214.