A Unified Theory of Irony and Its Computational Formalization

Akira Utsumi
Department of Systems Science, Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku, Yokohama 226, Japan
utsumi@sys.titech.ac.jp

Abstract

This paper presents a unified theory of verbal irony for developing a computational model of irony. The theory claims that an ironic utterance implicitly communicates the fact that its utterance situation is surrounded by ironic environment which has three properties, but hearers can assume an utterance to be ironic even when they recognize that it implicitly communicates only two of the three properties. Implicit communication of three properties is accomplished in such a way that an utterance alludes to the speaker's expectation, violates pragmatic principles, and implies the speaker's emotional attitude. This paper also describes a method for computationally formalizing ironic environment and its implicit communication using situation theory with action theory.

1 Introduction

Although non-literal language such as metaphor has become a popular topic in computational linguistics (Fass et al., 1991), no attention has been given to ironic uses of language. One reason for this imbalance is that traditional accounts of irony — and even default logic formalization (Perrault, 1990) — assume that irony communicates the opposite of the literal meaning. This assumption leads to the misconception that irony is governed only by a simple inversion mechanism, and thus it has no theoretical interest. Another reason is that studies of irony have been regarded as of no practical use for NLP systems. However, recent accounts denying the meaning-inversion assumption have revealed that irony is a more complicated pragmatic phenomenon governed by several mental processes (Kumon-Nakamura et al., 1995), and that irony offers an effective way of accomplishing various communication goals that are difficult to convey literally (Roberts and Kreuz, 1994).

The aim of this paper is to propose a unified theory of irony that answers to three crucial questions in an unequivocal manner: (Q1) what are properties that distinguish irony from non-ironic utterances, (Q2) how do hearers recognize utterances to be ironic, and (Q3) what do ironic utterances convey to hearers? Our theory provides a computationally feasible framework of irony as the first step toward a full-fledged computational model of irony, and it can account for several empirical findings from psycholinguistics. The essential idea underlying our theory is that an ironic utterance *implicitly displays ironic environment*, a special situation which has three properties for being ironic, but the hearer does not have to see all the three properties implicitly communicated in order to recognize the utterance to be ironic. Note that this paper focuses only on verbal irony, and thus situational irony[1] (i.e., situations are ironic) is beyond the scope of our theory.

This paper is organized as follows: Section 2 discusses the problems of previous irony theories. Section 3 presents our unified theory of irony that can cope with the problems, and its computational formalization. Finally, Section 4 suggests that our theory agrees well with several empirical findings.

2 Previous theories of irony

Several irony theories have been proposed in the last few decades, but all the theories, as we will explain, make the same mistake in that they confuse the two different questions (Q1) and (Q2).

The traditional pragmatic theory (Grice, 1975; Haverkate, 1990) assumes that an utterance is recognized to be ironic when the hearer becomes aware of an apparent violation of some pragmatic principles (e.g., the maxim of quality or the sincerity conditions for speech acts), and as a result it conveys the opposite of the literal meaning. This theory, however, completely fails to explain many ironic utterances. First, irony can be communicated by various expressions that do not include such violation: true assertions such as (2a) in Figure 1, understatements such as (2c), and echoic utterances such as (5a). Moreover, Candy's husband of Example 1 can perceive Candy's utterances (1a)~(1e) as ironic even under the situation where he does not know or is careless of Candy's expectation of satisfy her hunger, in other words, where he is not aware of the violation. This im-

[1] Situational irony can be indicated by metareferential expressions such as "It is ironic that...", but verbal irony is incompatible with such expressions.

Example 1: Candy baked a pizza to satisfy her hunger. When she was dishing it up, her husband entered the kitchen and gobbled up the whole pizza. Candy said to her husband:

(1) a. I'm not hungry at all.
b. Have you seen my pizza on the table?
c. I'll get to sleep.
d. I'm really satisfied to eat the pizza.
e. How about another small slice of pizza?

Example 2: A mother asked her son to clean up his messy room, but he did a sloppy, half-hearted job. After a while, she discovered that his room is still messy, and said to her son:

(2) a. I love children who keep their rooms clean, indeed.
b. Would you mind if I asked you to clean up your room, please?
c. This room may be slightly messy.

Example 3: Peter broke his wife's favorite teacup when he washed the dishes awkwardly. Looking at the broken cup, his wife said:

(3) a. Thank you for washing my cup carefully.
b. Thank you for crashing my treasure.

Example 4: Nancy and Jane were planning a trip to the beach, but that day was a cold and stormy one. As she looked out the window, Nancy said:

(4) a. Oh, the weather is really nice.
b. Maybe the beach is crowded with people.

Example 5: Just after his colleague Jesse said to him "I'd be promoted before you", Peter replied:

(5) a. You'd be promoted before me, huh?
b. You'd better work harder.

Figure 1: Five examples of ironic utterances

plies that violation of pragmatic principles is not an answer to (Q2). Secondly, it is not an answer to (Q1) because of its incompetence to discriminate irony from other non-literal utterances (e.g., a lie) in which the maxim of quality is flouted. Finally, the notion of "the opposite of the literal meaning" is problematic because it is applicable only to declarative assertions but many ironic utterances can take non-declarative forms: questions such as (1b); requests such as (2b); offerings such as (1e); and expressives such as (3a).

Other recent theories – e.g., mention theory (Wilson and Sperber, 1992) and echoic reminder theory (Kreuz and Glucksberg, 1989) – share a common view that by mentioning or alluding to someone's thought, utterance, expectation or cultural norm, an ironic utterance communicates a speaker's attitude toward a discrepancy between what actually is and what has been expected. This view may be essential to irony, but these theories are still incomplete as a comprehensive framework for irony for at least three reasons. First, their concepts of mention/allusion – Sperber and Wilson's echoic interpretation and Kreuz and Glucksberg's echoic reminder – are too narrow to capture the allusive nature of irony (e.g., (1b), (1e), (4b)), and they are not clear enough to be formalized in a computable fashion. For example, Nancy's utterance (4a) in Figure 1 is an echoic interpretation of Nancy's expectation of the fine weather, but (4b) does not interpretively echo any states of affairs: (4b) is an implication derived from the failed expectation. Second, they implicitly assume that the properties that characterize irony can be applied to recognition of ironic utterances as they stand or they do not focus on how hearers recognize utterances to be ironic. Thus they cannot also explain a certain kind of ironic utterances in which hearers are not aware of any pragmatic violation. Finally, these theories provide no plausible explanation of how irony is discriminated from non-ironic echoic utterances.

Allusional pretense theory (Kumon-Nakamura et al., 1995) is the most powerful one in that it can explain ironic utterances of five speech act classes using the two crucial notions of allusion (including echoic interpretation and reminder) and pragmatic insincerity. They claimed that all ironic utterances allude to a failed expectation and violate one of the felicity conditions for well-formed speech acts. However, allusional pretense theory still suffers from the same disadvantage as other theories: their notion of allusion is not clear enough, and it does not focus on how hearers recognize utterances to be ironic.

3 A unified theory of irony

3.1 Ironic Environment and Its Implicit Display

Our unified theory of irony claims as an answer to (Q1) that irony is a figure of speech that implicitly displays the fact that its utterance situation is surrounded by ironic environment. To make this claim realizable, we must explain two important notions: ironic environment and implicit display.

In order for an utterance to be ironic, a speaker must utter in a situation surrounded by ironic environment. Given two temporal locations t_0 and t_1 such that t_0 temporally precedes t_1, the utterance situation where an utterance is given is surrounded by ironic environment if and only if it satisfies the following three conditions:

1. The speaker has an expectation E at t_0.
2. The speaker's expectation E fails at t_1.
3. As a result, the speaker has a negative emotional attitude toward the incongruity between what is expected and what actually is.

Note that our notion of speaker's expectaions subsumes cultually expected norms and rules. Furthermore previous theories assume echoic irony like (5a) to allude to other person's thoughts or

Example 1

Instantiated Causal Relations:

$S_1 \models \langle\!\langle\, accessible, x, l_t \,\rangle\!\rangle \wedge \underbrace{\langle\!\langle\, loc, a, l_t \,\rangle\!\rangle}_{X} \wedge \langle\!\langle\, eatable, a \,\rangle\!\rangle : [\underbrace{eat(x,a)}_{A}] \Rightarrow \underbrace{S_2 \models \langle\!\langle\, hungry, x; 0 \,\rangle\!\rangle}_{Q} \wedge \langle\!\langle\, loc, a, l_x \,\rangle\!\rangle$

$S_1 \models \langle\!\langle\, accessible, y, l_t \,\rangle\!\rangle \wedge \langle\!\langle\, loc, a, l_t \,\rangle\!\rangle \wedge \langle\!\langle\, eatable, a \,\rangle\!\rangle : [\underbrace{eat(y,a)}_{B}] \Rightarrow S_2 \models \langle\!\langle\, hungry, y; 0 \,\rangle\!\rangle \wedge \langle\!\langle\, loc, a, l_y \,\rangle\!\rangle$

$S_1 \models \langle\!\langle\, hungry, x; 0 \,\rangle\!\rangle \Rightarrow \underbrace{S_2^{\langle\!\langle\, precedes, S_1, S_2 \,\rangle\!\rangle} \models \langle\!\langle\, get\text{-}to\text{-}sleep, x \,\rangle\!\rangle}_{Y}$

Ironic Environment:

$\langle\!\langle\, Candy, x \,\rangle\!\rangle \wedge \langle\!\langle\, husband, x, y \,\rangle\!\rangle \wedge \langle\!\langle\, pizza, a \,\rangle\!\rangle \wedge \langle\!\langle\, eatable, a \,\rangle\!\rangle \wedge \langle\!\langle\, on, l_t, b \,\rangle\!\rangle \wedge \langle\!\langle\, table, b \,\rangle\!\rangle \wedge$
$\langle\!\langle\, accessible, x, l_t \,\rangle\!\rangle \wedge \langle\!\langle\, in, l_x, c \,\rangle\!\rangle \wedge \langle\!\langle\, stomach, x, c \,\rangle\!\rangle \wedge \langle\!\langle\, in, l_y, d \,\rangle\!\rangle \wedge \langle\!\langle\, stomach, y, d \,\rangle\!\rangle$

$t_0 \models \langle\!\langle\, loc, a, l_t \,\rangle\!\rangle \wedge \langle\!\langle\, hungry, x \,\rangle\!\rangle \wedge \langle\!\langle\, hope, x, T^{\langle\!\langle\, precedes, t_0, T \,\rangle\!\rangle} \models \langle\!\langle\, hungry, x; 0 \,\rangle\!\rangle \,\rangle\!\rangle$
$\Downarrow \quad eat(y,a)$

$t_1 \models \langle\!\langle\, loc, a, l_y \,\rangle\!\rangle \wedge \langle\!\langle\, hungry, x \,\rangle\!\rangle \wedge \langle\!\langle\, hope, x, T^{\langle\!\langle\, precedes, t_1, T \,\rangle\!\rangle} \models \langle\!\langle\, hungry, x; 0 \,\rangle\!\rangle \,\rangle\!\rangle \wedge \langle\!\langle\, did, eat(y,a) \,\rangle\!\rangle \wedge$
$\langle\!\langle\, hungry, y; 0 \,\rangle\!\rangle \wedge \langle\!\langle\, did, eat(x,a); 0 \,\rangle\!\rangle \wedge \langle\!\langle\, angry\_at, x, y, eat(y,a) \,\rangle\!\rangle$

Example 2

Instantiated Causal Relations: $S_1 \models \langle\!\langle\, messy, a \,\rangle\!\rangle : [clean\text{-}up(y,a)] \Rightarrow S_2 \models \langle\!\langle\, clean, a \,\rangle\!\rangle$

Ironic Environment:

$\langle\!\langle\, mother, x, y \,\rangle\!\rangle \wedge \langle\!\langle\, son, y, x \,\rangle\!\rangle \wedge \langle\!\langle\, room, a \,\rangle\!\rangle \wedge \langle\!\langle\, owns, y, a \,\rangle\!\rangle$

$t_0 \models \langle\!\langle\, messy, a \,\rangle\!\rangle \wedge \langle\!\langle\, ask, x, y, clean\text{-}up(y,a) \,\rangle\!\rangle \wedge \langle\!\langle\, hope, x, T^{\langle\!\langle\, precedes, t_0, T \,\rangle\!\rangle} \models \langle\!\langle\, clean, a \,\rangle\!\rangle \,\rangle\!\rangle$
$\Downarrow \quad \neg clean\text{-}up(y,a)$

$t_1 \models \langle\!\langle\, messy, a \,\rangle\!\rangle \wedge \langle\!\langle\, did, clean\text{-}up(y,a); 0 \,\rangle\!\rangle \wedge \langle\!\langle\, hope, x, T^{\langle\!\langle\, precedes, t_1, T \,\rangle\!\rangle} \models \langle\!\langle\, clean, a \,\rangle\!\rangle \,\rangle\!\rangle \wedge$
$\langle\!\langle\, angry\_at, x, y, \neg clean\text{-}up(y,a) \,\rangle\!\rangle$

Figure 2: Representation of ironic environments for Examples 1 and 2

utterances, but our theory contends that such irony alludes to a speaker's expectation that "the speaker wants the hearer to know the hearer's utterances or thoughts are false". For example, the speaker's expectation of (5a) is that Jesse knows he cannot be promoted before Peter.

Ironic environment can be classified into the following four types.

- a speaker's expectation E can be caused by an action A performed by intentional agents
  - E failed because A failed or cannot be performed by another action B (type-1)
  - E failed because A was not performed (type-2)
- a speaker's expectation E is not normally caused by any intentional actions
  - E failed by an action B (type-3)
  - E accidentally failed (type-4)

For example, ironic environment of Example 1 falls in type-1: Candy's expectation of staying her hunger can be realized by an action of eating a pizza, but her husband's action of eating the whole pizza hindered her expected action. In the same way, ironic environments of Examples 2-4 fall in type-2∼type-4, respectively, and that of Example 5 falls in type-3.

An utternace implicitly displays all the three conditions for ironic environment when it

1. alludes to the speaker's expectation E,
2. includes pragmatic insincerity by intentionally violating one of pragmatic principles, and
3. implies the speaker's emotional attitude toward the failure of E.

For example, utterances (2d) and (2e) for Example 2 are not ironic even when they are given in the situation surrounded by ironic environment: (2d) and (2e) directly express the speaker's expectation and the speaker's emotional attitude, respectively, and both do not include pragmatic insincerity.

(2) d. I've expected a clean room.
e. I'm disappointed with the messy room.

On the other hand, all the utterances of Figure 1 are ironic because they implicitly express the three components of ironic environment, as we will show in Sections 3.3-3.5.

3.2 Representing Ironic Environment

In order to formalize ironic utterances and ironic environment in a computational fashion, we use situation theory (Barwise, 1989) and situation calculus. Our representational scheme includes discrete items of information called *infons*, *situations* capable of making infons true (i.e., supporting infons), and *actions*. For example, information that Candy eats the pizza is represented as the infon $\langle\!\langle\, eat, x, a \,\rangle\!\rangle$ in which x and a denote "Candy" and "the pizza", and its negation as $\langle\!\langle\, eat, x, a; 0 \,\rangle\!\rangle$. A fact/event that Candy eats the pizza is represented as $t \models \langle\!\langle\, eat, x, a \,\rangle\!\rangle$ where the situation t expresses the spatiotemporal location of that event. An action of eating the pizza performed by Candy is expressed by the predicate $eat(x,a)$ and its negation (i.e., an action of not performing $eat(x,a)$) by $\neg eat(x,a)$. The state of affairs that an action A is performed is expressed by $\langle\!\langle\, did, A \,\rangle\!\rangle$. Also, a proposition p expressing the claim that Candy eats the pizza is written as $(t \models \langle\!\langle\, eat, x, a \,\rangle\!\rangle)$. The proposition $p = (s \models \sigma)$

Speech Act:	$Inform(S, H, P)$
Preconditions:	$\langle\!\langle\, speaker, S \,\rangle\!\rangle$, $\langle\!\langle\, hearer, H \,\rangle\!\rangle$, $\langle\!\langle\, proposition, P \,\rangle\!\rangle$, $u_S \models P$
Effects:	$u_H \models P$, $u_H \models u_S \models P$
Speech Act:	$RequestIf(S, H, P)$
Preconditions:	$\langle\!\langle\, speaker, S \,\rangle\!\rangle$, $\langle\!\langle\, hearer, H \,\rangle\!\rangle$, $\langle\!\langle\, proposition, P \,\rangle\!\rangle$, $\neg KnowIf(S, P)$
Effects:	$u_H \models \langle\!\langle\, intend, S, InformIf(H, S, P) \,\rangle\!\rangle$

Notes: u_S and u_H denote the speaker's and hearer's mental situations, $KnowIf(S, P) = u_S \models P \vee \neg P$, and $\neg KnowIf(S, P) = u_S \not\models P \wedge \neg P$.

Figure 3: Speech act definitions

is true if s supports σ, and otherwise false. Situations are partially ordered by the *part-of* relation denoted by $\trianglelefteq$. A situation s_1 is a part of a situation s_2 (i.e., $s_1 \trianglelefteq s_2$) if and only if every infon supported by s_1 is also supported by s_2. In this paper we also represent an agent X's mental situation as u_X and his/her beliefs as support relations between u_X and infons. For example, the fact that Jim believes/knows the above event is represented as $u_{Jim} \models t \models \langle\!\langle\, eat, x, a \,\rangle\!\rangle$. Infons and actions can include parameters denoted by capital letters. Parameters can be restricted by infons: for example, $T^{\langle\!\langle\, precedes, t_0, T \,\rangle\!\rangle}$ is a parameter for temporal situations which temporally succeed t_0. A *causal relation* between two events $s_1 \models \sigma_1$ and $s_2 \models \sigma_2$ is expressed by $s_1 \models \sigma_1 : [A] \Rightarrow s_2 \models \sigma_2$. This relation means that if an action A is executed in a situation s_1 supporting the infon σ_1, then it causes the infon σ_2 to be true in the resulting situation s_2. Thus it follows that $s_2 \models \langle\!\langle\, did, A \,\rangle\!\rangle$. When we omit an action A from a causal relation, that relation becomes a *constraint* in situation theory, denoted by $s_1 \models \sigma_1 \Rightarrow s_2 \models \sigma_2$. Figure 2 illustrates the representation of ironic environments of Examples 1 and 2. Although Figure 2 does not include any mental situations (i.e., ironic environment is represented from god's eye view), when a speaker intends the utterance to be ironic the speaker's mental situation must support all states of affairs, events and causal relations in this figure.

An utterance U is characterized by its propositional content P and the illocutionary act that the speaker performs in saying U, some of which are shown in Figure 3 (Litman and Allen, 1987). For example, the propositional content of (1a) is $(t_1 \models \langle\!\langle\, hungry, x; 0 \,\rangle\!\rangle)$ and its illocutionary act is *Inform*. Also (1b) is characterized by $P = (t_1 \models \langle\!\langle\, see, y, T^{\langle\!\langle\, precedes, T, t_1 \,\rangle\!\rangle} \models \langle\!\langle\, loc, a, l_t \,\rangle\!\rangle \,\rangle\!\rangle)$ and the illocutionay act *RequestIf*.

3.3 Allusion

We give a formal definition of allusion in our theory. Given P expressing the propositional content of U, and Q expressing the speaker's expected event/state of affairs, an utterance U *alludes* to the expectation E if it satisfies one of the conditions shown in Table 1. The relation $\rightsquigarrow$ in Table 1 is defined as follows: assuming that

- $\langle\!\langle\, hope, P, (S \models X) \,\rangle\!\rangle \Leftarrow \langle\!\langle\, want, P, (S \models X) \,\rangle\!\rangle \wedge \langle\!\langle\, anticipate, P, (S \models X) \,\rangle\!\rangle$
- $S_1 \models \langle\!\langle\, disappointed, P, (S_1 \models X) \,\rangle\!\rangle \Leftarrow S_0 \models \langle\!\langle\, hope, P, (S \models \overline{X}) \,\rangle\!\rangle \wedge S_1 \models X \wedge S_1 \trianglelefteq S \wedge \langle\!\langle\, precedes, S_0, S_1 \,\rangle\!\rangle \wedge \langle\!\langle\, precedes, S_0, S \,\rangle\!\rangle$
- $S_1 \models \langle\!\langle\, angry\_at, P_1, P_2, A \,\rangle\!\rangle \Leftarrow S_0 \models \langle\!\langle\, want, P_1, (S \models \overline{X}) \,\rangle\!\rangle \wedge S_1 \models X \wedge S_1 \trianglelefteq S \wedge \langle\!\langle\, precedes, S_0, S_1 \,\rangle\!\rangle \wedge \langle\!\langle\, precedes, S_0, S \,\rangle\!\rangle \wedge \langle\!\langle\, agent, A, P_2 \,\rangle\!\rangle \wedge S_0 \models * : [A] \Rightarrow S_1 \models X \wedge S_1 \models \langle\!\langle\, did, A \,\rangle\!\rangle \wedge \langle\!\langle\, blameworthy, A \,\rangle\!\rangle$

Figure 4: Emotion-eliciting rules

$P_1 = P_2$ means that both are conceptually identical or unifiable, $P_1 \rightsquigarrow P_2$ holds if

$$\begin{cases} P_1 = (P_2) \text{ or } P_1\text{'s constituent} = \{P_2 \text{ or } (P_2)\} \\ \qquad\qquad \text{(when } P_2 \text{ is an event)} \\ P_1 = P_2 \text{ or } P_1\text{'s constituent} = P_2 \\ \qquad\qquad \text{(when } P_2 \text{ is an action)} \end{cases}$$

This definition allows all utterances in Figure 1 to allude speaker's expectations, but it does not allow (2d) to allude to it. Table 1 shows which condition each of these utterances satisfies. For example, the utterance (1b) that mention theory cannot explain alludes to Candy's expectation by referring to one of the conditions $X = S_1 \models \langle\!\langle\, loc, a, l_t \,\rangle\!\rangle$ in Figure 2 for an action $A = eat(x, a)$ since the part of its propositional content P and X are unifiable. Other utterances for Example 1, (1a) and (1c)$\sim$(1e), also refer to Q, Y, A, B shown in Figure 2, respectively. In the same way, (2b) satisfies Condition 4 since its content $P = clean\text{-}up(y, a)$ is identical to A.

3.4 Pragmatic Insincerity

Table 2 lists the pragmatic principles violated by the ironic utterances in Figure 1. In many cases an ironic utterance is pragmatically insincere in the sense that it intentionally violates one of the preconditions in Figure 3 (i.e., sincerity, preparatory and propositional conditions) that need to hold before its illocutionary act is accomplished, but pragmatic insincerity also occurs when an utterance violates other pragmatic principles. Requests often become insincere when they are over-polite like (2b) since they violate the politeness principle (although (2b) also becomes insincere when the mother no longer intends her son to clean up his room). Understatements like (2c) are also insincere since they do not provide as much information as required. The true assertion (2a) violates the principle of relevance in that it does not yield any contextual implication. As mentioned earlier, the last three cases have been problematic for all the previous theories of irony because none of these theories recognized a wide variety of principles violated by ironic utterances. Although this paper does not describe how these pragmatic principles should be formalized, they should be taken into account for the next steps of our study.

Table 1: Allusion of ironic utterances in Figure 1

Conditions for allusion	Utterances satisfying the condition
1. $P \rightsquigarrow Q \wedge P \not\rightsquigarrow T \models \langle\langle R, S, Q \rangle\rangle$	(1a) (2a) (4a) (5a)
2. $P \rightsquigarrow X$ where $X : [A] \Rightarrow Q$ or $X \Rightarrow Q$	(1b) (2c) (5b)
3. $P \rightsquigarrow Y$ where $Q \Rightarrow Y$	(1c) (4b)
4. $P \rightsquigarrow A$ where $X : [A] \Rightarrow Q$ (type-1 or type-2)	(1d) (2b)
5. $P \rightsquigarrow B$ or W or Z where $W : [B] \Rightarrow Z$ (type-1 or type-3)	(1e) (3a) (3b)

Notes: In Condition 1, T, R and S denote parameters for situations, relations about expecting, and speakers, respectively. In Condition 5, B denotes actions which disable an action A of Condition 4.

Table 2: Pragmatic insincerity of ironic utterances in Figure 1

Violated pragmatic principles		Utterances violating the principle
Sincerity condition	for Inform (S believes P)	(1a) (1c) (1d) (4a) (4b) (5a)
	for Question (S does not know P)	(1b)
	for Advise (S believes P will benefit H)	(5b)
	for Offer (S wants to do an action P for H)	(1e)
	for Thank (S feels grateful for an action P)	(3b)
Propositional content condition for Thank (P is a past action done by H)		(3a)
Preparatory condition for Offer (S is able to do an action P)		(1e)
Maxim of relevance (P is relevant in Sperber and Wilson's (1986) sense)		(2a)
Politeness principle (U should be made at an appropriate level of politeness)		(2b)
Maxim of quantity (P is as informative as required)		(2c)

Notes: S, H and P denote the speaker, the hearer and the propositional content, respectively.

3.5 Emotional Attitude

Speakers can use a variety of signals/cues — intonation contour, exaggerated stress, tone of voice, hyperbole, facial expression, etc. — for implicitly communicating their emotional attitude. The use of the interjection "Oh" with a special tone of voice in (4a) offers one typical example of this. Implicit communication can also be accomplished by utterances explicitly referring to the pleased emotion that speakers would experience if their failed expectation became true. For example, the utterance (3a) explicitly expresses speaker's counterfactual emotion.

At the same time, many ironic utterances make emotion-eliciting rules for the speaker's attitude (some of which are shown in Figure 4) accessible by the hearers by alluding to one of premises of the rule. In the case of (3a), it alludes to Peter's action of washing the dishes so that the rule for "angry_at" emotion becomes more accessible.

3.6 Recognizing and Interpreting Irony

In many cases, all the three components for implicit communication of ironic environment are easily recognized by the hearer. As we mentioned in Section 2, however, there are also many cases such as Example 1 that an utterance can be ironically interpreted even though all the three components cannot be recognized by the hearer because the hearer's mental situation differs from the speaker's one. Furthermore, in the case of (5a), after recognizing the utterance to be ironic Jesse turns out to know that the speaker Peter thinks Jesse cannot be promoted before Peter.

Hence we propose the following condition for recognizing irony as an answer to (Q2):

> Hearers can assume an utterance to be ironic (with high possibility) if they can recognize that the utterance implicitly displays at least two of the three components for ironic environment, and if the utterance situation does not rule out the possibility of including the unrecognized components, if any.[2]

This "2-of-3" criterion makes it possible that hearers can recognize utterances as ironic even though speakers do not intend their utterances to be understood as irony. It provides empirical evidence of our theory since such unintentional irony has been found in a number of psychological experiments (Gibbs and O'Brien, 1991).

By recognizing an utterance to be ironic, the hearer becomes aware of an illocutionary act of irony, that of conveying the fact that the utterance situation is surrounded by ironic environment (i.e., all the three components for ironic environment hold in a current situation). That is an answer to (Q3), and then the hearer interprets/understands the ironic utterance by adding that information to his/her mental situation. In

[2]Practically speaking, whether an utterance is ironic is a matter of degree. Thus the degree of ironicity might be a better criterion for recognizing irony. If we can quantitatively evaluate, though do not in this paper, to what degree an utterance alludes to the speaker's expectation, to what degree it includes pragmatic insincerity, and to what degree it implies the speaker's emotional attitude, we think the proposed condition for recognizing irony can also be quantitatively defined.

many cases, since the hearer already knows the fact that the three components hold in the situation, interpretation of irony results in confirmation of the most uncertain information, that is, the speaker's emotional attitude. However, when the hearer does not recognize all components, he/she also obtains new information that the unrecognized component holds in a current situation. Therefore, our theory includes many previous theories claiming that irony communicates an ironist's emotional attitude. For example, in the case of (5a), after recognizing Peter's utterance (5a) to be ironic, Jesse turns out to know that Peter thinks Jesse's preceding utterance is absurd, and tries to confirm Peter's emotional attitude by interpreting (5a) ironically. Furthermore, as we mentioned in Section 1, an ironic utterance achieves various communication goals held by the speaker -- e.g., to be humorous, to emphasize a point, to clarify -- as perlocutionary acts.

4 Implications of the Theory

Distinction between ironic and non-ironic utterances: Our theory can distinguish ironic utterances from non-ironic ones. For example, lies and other non-ironic utterances violating the pragmatic principle do not allude to any antecedent expectation and/or do not offer cues for reasoning about the speaker's emotional attitude. Non-ironic echoic utterances do not include pragmatic insincerity and/or do not implicitly communicate the speaker's attitude.

Ironic cues: Some theories assume that irony can be identified by special cues for irony, but the empirical finding in psychology shows that people can interpret ironic statements without any special intonational cues (Gibbs and O'Brien, 1991). Our theory agrees with this finding: such kind of cues is only a part of Component 3 as we described in Section 3.5, and thus ironic utterances without these cues can be recognized as ironic.

Victims of irony: Several irony studies, e.g., (Clark and Gerrig, 1984), have pointed out that irony generally has victims. Our theory suggests that ironic utterances have potential victims when their ironic environments fall in one of types-1,2,3: in the case of type-1 or type-3 an agent of B becomes a victim, and in the case of type-2 an agent of A becomes a victim.

Sarcasm and irony: We argue that explicit victims and display of the speaker's counterfactual pleased emotion described in Section 3.5 are distinctive properties of sarcasm. Thus the utterances (3a) and (3b) are sarcastic because they have an explicit victim, Peter, and they refer to the wife's counterfactual pleased emotion. In particular, an utterance "Thanks a lot!" for Example 3 is non-ironic sarcasm since it does not allude to any expectation.

5 Conclusion

In this paper we have proposed a unified theory of irony that overcomes several difficulties of previous irony theories. Our theory allows us to give plausible answers to what irony is, how irony is recognized and what irony communicates. The properties of irony -- allusion, pragmatic insincerity, and emotional attitude -- are formalized unequivocally enough to build a computational model of irony. From this point of view, we believe that this paper provides a basis for dealing with irony in NLP systems, and we are developing computational methods for interpreting and generating irony (Utsumi, 1995).

References

J. Barwise. 1989. *The Situation in Logic*. Stanford: CSLI Publications.

H.H. Clark and R.J. Gerrig. 1984. On the pretense theory of irony. *Journal of Experimental Psychology: General*, 113(1):121–126.

D. Fass, E. Hinkelman, and J. Martin, editors. 1991. *Proceedings of the IJCAI Workshop on Computational Approaches to Non-Literal Language: Metaphor, Metonymy, Idioms, Speech Acts, Implicature*.

R.W. Gibbs and J. O'Brien. 1991. Psychological aspects of irony understanding. *Journal of Pragmatics*, 16:523–530.

H.P. Grice. 1975. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and semantics, Vol.3: Speech acts*, pages 41–58. Academic Press.

H. Haverkate. 1990. A speech act analysis of irony. *Journal of Pragmatics*, 14:77–109.

R.J. Kreuz and S. Glucksberg. 1989. How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General*, 118(4):374–386.

S. Kumon-Nakamura, S. Glucksberg, and M. Brown. 1995. How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, 124(1):3–21.

D.J. Litman and J.F. Allen. 1987. A plan recognition model for subdialogues in conversations. *Cognitive Science*, 11:163–200.

C.R. Perrault. 1990. An application of default logic to speech act theory. In P.R. Cohen, J. Morgan, and M.E. Pollack, editors, *Intentions in Communication*, pages 161–185. The MIT Press.

R.M. Roberts and R.J. Kreuz. 1994. Why do people use figurative language? *Psychological Science*, 5(3):159–163.

D. Sperber and D. Wilson. 1986. *Relevance: Communication and Cognition*. Oxford, Basil Blackwell.

A. Utsumi. 1995. How to interpret irony by computer: A comprehensive framework for irony. In *Proceedings of RANLP*, pages 315–321, Bulgaria.

D. Wilson and D. Sperber. 1992. On verbal irony. *Lingua*, 87:53–76.