Implicit Display Theory of Verbal Irony:
Towards A Computational Model of Irony

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

Abstract
This paper proposes the implicit display theory of verbal irony that overcomes several difficulties of previous irony theories, and then describes a computational model of irony interpretation and generation based on the theory. The theory claims that irony 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 do not see all the three properties implicitly displayed by the utterance. Implicit communication of three properties is accomplished in such a way that an utterance alludes to the speaker’s expectation, violates pragmatic principles, and is accompanied by several cues for implying the speaker’s emotional attitude.

1 INTRODUCTION

Irony is one of the most ubiquitous ingredients of nonliteral (or humorous) language. Imagine the following situation and an utterance (1a):

[Situation 1] 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:

(1) a. This room is very clean!

This utterance is a typical example of irony, and the speaker, a mother, intends to convey something other than what the words explicitly mean.

Then what do ironic utterances convey? One notorious answer provided by traditional accounts of irony is that irony communicates the opposite of the literal meaning. This account is problematic for several reasons, and what is worse, it leads to the misconception that irony is governed only by a simple inversion mechanism. Recent studies denying the meaning-inversion assumption have revealed that irony is a more complicated pragmatic phenomenon governed by several mental processes (e.g., Wilson and Sperber 1992; Kamon-Nakamura et al. 1995). However, they still do not give plausible answers to the following three essential questions: 
(Q1) what properties distinguish irony from non-ironic utterances?; (Q2) how do hearers recognize utterances to be ironic?; and (Q3) what do ironic utterances convey to hearers?

In the domain of computational linguistics, surprisingly little attention has been given to ironic uses of language, although other nonliteral language such as metaphor and implicature has been a popular topic (Fass et al. 1991). However, it is important to construct a computational model of irony for at least two reasons: irony, which is heavily dependent upon context, can throw new light on computational studies of many pragmatic phenomena, and it offers an effective way of accomplishing various communication goals that are difficult to convey literally (Roberts and Kreuz 1994).

The author has been exploring the mechanism of recognizing and interpreting irony and trying to construct a computational model of irony. This paper presents the implicit display theory, a unified theory of irony that answers the three questions (Q1)∼(Q3), and illustrates how irony is interpreted and generated by computer. The 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 the theory is that an ironic utterance implicitly displays ironic environment, a special situation which has three properties for being ironic — expectation, incongruity, emotional attitude —, but the hearer does not have to see all the three properties implicitly communicated in order to recognize the utterance to be ironic.

This paper focuses only on verbal irony which can be distinguished from situational irony (or irony of fate). Situations are ironic when an expectation is violated in specific ways and such
nonverbal irony appears even without verbal communication. Note that situational irony can be indicated by metareferential expressions like “It is ironic that...”, but those expressions themselves are not ironic.

This paper is organized as follows: Section 2 discusses the problems of previous irony theories and then Section 3 explains, in detail, implicit display theory of irony that overcomes the difficulties. Section 4 gives a computational formalization of the theory and a basic framework for irony interpretation and generation.

2 Previous Approaches to Irony

A number of studies have so far tackled the three questions (Q1)∼(Q3). They are classified into two essentially different approaches: the pragmatic approach and the cognitive approach.

The traditional pragmatic theory (Grice 1975; Searle 1979) 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 sincerity conditions for speech acts), and as a result it conveys the opposite of the literal meaning. This view applies well to typical ironies such as (1a), but it completely fails to explain what irony is, how irony is recognized and what irony communicates. First, irony can be communicated by various expressions that do not include such violation. For example, the following utterances — a true assertion (1b), a understatement (1c), and an over-polite request (1d) — are interpreted as ironic under Situation 1 though they violate neither the maxim of quality nor any felicity conditions.

(1) a. You'd be promoted before me, huh?
   b. I love children who keep their rooms clean.
   c. This room may be slightly messy.
   d. Would you mind if I asked you to clean up your room, please?

Also, Peter’s echoic reply (2a) of the following exchange is also one of ironies without such violation.

[Situation 2] Just after his colleague Jesse said to him “I’d be promoted before you”, Peter replied:

(2) a. You’d be promoted before me, huh?

Moreover, in some cases, ironic intent can be communicated under the situation where hearers are not aware of any violation. For example, Mary can perceive Peter’s utterance (3) in Situation 3 as ironic even when she does not at all think that she is not a good cook. All these facts imply that violation of such principles is not a necessary condition of irony and the traditional approach cannot give an answer to (Q2).

[Situation 3] Eating an unsavory fruit cake made by Mary, who is proud of her skill in cooking, Peter says:

(3) You are a good cook indeed.

Although Haverkate (1990) extended a target of intentional violation to felicity conditions for five speech act classes, it is still incomplete for the same reasons. Secondly, the traditional approach cannot discriminate irony from other non-literal utterances (e.g., a lie) in which the maxim of quality is flouted, and therefore does not provide a plausible answer to (Q1). 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 (4c); requests such as (1d); offerings such as (4e); and expressives such as (5a).

[Situation 4] 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:

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

[Situation 5] Peter broke his wife’s favorite teacup when he washed the dishes awkwardly. Looking at the broken cup, his wife said:

(5) a. Thank you for washing my cup carefully.
   b. Thank you for crushing my treasure.

Mention theory proposed by Sperber and Wilson (Sperber and Wilson 1981, 1986; Wilson and Sperber 1992) is the first theory to focus on the allusive nature of irony. They have argued that verbal irony is a variety of echoic mention of an attributed thought or utterance. By mentioning or alluding to someone’s thought, utterance, expectation or cultural norm, irony communicates a speaker’s attitude toward a discrepancy between what actually is and what has been expected. This view of irony is shared by other cognitive approaches such as Kreuz and Glucksberg’s (1989) echoic reminder theory, and it has been supported by psychological experiments (Jorgensen et al. 1984). These theories, however, 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. For example, Nancy’s utterance (6a) is an echoic interpretation of Nancy’s expectation of the fine weather, but (6b) does not interpretively echo any states of affairs: (6b) is an
implication derived from the failed expectation.

[Situation 6] 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:

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

Likewise their theories cannot explain what (4c) and (4e) allude to. Thus their notions are not a necessary condition of irony. Second, these theories provide no plausible explanation of how irony is discriminated from non-ironic echoic utterances, and therefore do not provide a plausible answer to (Q1). Finally, they explicitly assume that 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.

Allusional pretense theory (Kumon-Nakamura et al. 1995) integrated these two approaches and claimed that all ironic utterances allude to a failed expectation and violate one of the felicity conditions for well-formed speech acts. This theory has a powerful ability to explain more ironic utterances than previous studies, resolving some difficulties of previous studies. However, allusional pretense theory still suffers from the same disadvantage as other theories: 1) their notion of allusion is not clear enough; 2) some ironies — e.g., (1b), (1c), (2a) — do not include any violation of the felicity conditions as we described above; 3) they do not explain how hearers recognize utterances to be ironic; and 4) the role of ironic cues (e.g., ironic tone of voice) is not addressed.

Clark and Gerrig’s (1984) pretense theory takes a rather different approach to irony: “in being ironic, the theory goes, a speaker is pretending to be an injudicious person speaking to an uninhibited audience; the speaker intends the addressees of the irony to discover the pretense and thereby see his or her attitude toward the speaker, the audience and the utterance” (ibid., p.121). According to pretense theory, each and every time the speaker says irony he/she is pretending to be an injudicious person, which corresponds to the victim of irony. For example, this theory explains ironies in Situation 4 in a way that Candy, the speaker of (4a)~(4e), is pretending to her husband who ate Candy’s pizza, exaggerating how ridiculous his behavior is. However, that there are ironies that do not have victims such as (6a) makes pretense theory less convincing. Furthermore, as Kreuz and Glucksberg (1989) pointed out, the notion of pretense is too powerful for a comprehensive theory of irony in that it can be applied to all indirect speech acts. Though pretense theory claims to be superior to the mention theory, there is a negative suggestion, which we agree with, that both theories may not differ significantly from one another (Williams 1984).

From the above discussion, we have made the following observations on a comprehensive theory of irony which will be presented in Section 3.

- The notions of allusion and pragmatic violation are essential to irony. However, previous theories are too general in that each of their proposed properties for being ironic also covers a part of non-ironic utterances, and at the same time, too specific in that each of them does not cover all ironic utterances. To make an adequate theory that explains both what irony is and what irony is not, we must integrate allusion, pragmatic violation, and other properties of irony into the theory.
- All the previous theories make the same mistake in that they confuse the two different questions (Q1) and (Q2), since there are several cases like Situation 3 in which after recognizing a given utterance to be ironic the hearer becomes aware that one of properties is satisfied. Thus the theory of irony must distinguish conditions for perceiving an utterance as ironic from those for an utterance being ironic.

3 IMPLICIT DISPLAY THEORY OF IRONY

3.1 WHAT IS IRONY?

The implicit display theory claims as an answer to (Q1) that irony is an utterance which implicitly displays to the hearer the situation surrounded by ironic environment. It can be divided into two parts: ironic environment as a situational property and implicit display as a linguistic property.

3.1.1 IRONIC ENVIRONMENT

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 equals or precedes $t_1$, a situation in which 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 (i.e., $E$ is incongruous with the reality) 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 the notion of expectations here subsumes culturally expected norms and conven-
tions, but does not include other person’s thoughts or utterances. For example, Situation 4 is surrounded by ironic environment, since the ironist Candy has an expectation of staying her hunger, but her expectation has not been fulfilled, and she is disappointed or angry with the result.

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 Situation 4 falls in type-1: Candy’s expectation 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 environment of Situation 1 falls in type-2, those of Situations 2, 3 and 5 fall in type-3, and that of Situation 6 falls in type-4.

### 3.1.2 Implicit display

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

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

For example, utterances (1e) and (1f) for Situation 1 are not ironic even when they are given in the situation surrounded by ironic environment:

(1e) I’ve expected a clean room.
(1f) I’m disappointed with the messy room.

(1e) and (1f) directly express the speaker’s expectation and the speaker’s emotional attitude, respectively, and both do not include pragmatic insincerity. On the other hand, all ironic utterances implicitly express the three components of ironic environment, as we will see below.

### 3.1.3 Allusion

We give a formal definition of allusion to E. Given P expressing the propositional content of U, Q expressing the speaker’s expected event/state of affairs, and PE expressing the state of affairs that the speaker expects E, an utterance U alludes to the expectation E if and only if it satisfies one of the conditions shown in Table 1. The relation \( U \sim X \) in Table 1 holds if \( U’ \)’s propositional content \( P \) and \( X \) are conceptually/semantically identical or unifiable (i.e., \( P = X \)) or if \( U \) includes a referring expression \( U_r \) (i.e., phrases or words) whose referent is \( X \).

For example, the following ironic utterances for Situation 1 show a variety of the relation \( U \sim Q \):

(1) a. This room is very clean! \((P = Q)\)
   b. I love children who keep their rooms clean.
      (the phrase “who keep their rooms clean” refers to \( Q \))
   c. How do you feel in such a comfortable place?
      (the phrase “such a comfortable place” anaphorically refers to \( Q \))

The notion of allusion here is wider than Sperber and Wilson’s echoic mention/interpretation and clearer than that of allusional pretense theory. The definition allows ironic utterances to allude speaker’s expectations, but it does not allow (1e) to allude to it because of the condition \( U \not\sim PE \). Table 1 also shows which condition each of ironies presented in this paper satisfies. For Situation 4, the irony (4c) that mention theory cannot explain alludes to Candy’s expectation by the phrase “my pizza on the table” referring to one of the conditions \( X = “Candy’s pizza was on the table” \) for an action \( A = “eat Candy’s pizza” \). Other four utterances for Situation 4, (4a), (4b), (4d) and (4e), also refer to \( Q, A, Y, B \), respectively.

### 3.1.4 Pragmatic insincerity

Table 2 lists the pragmatic principles violated by the ironic utterances in this paper. In many cases an ironic utterance is pragmatically insincere in the sense that it intentionally violates one of the preconditions (i.e., sincerity, preparatory and propositional conditions) that need to hold before its illocutionary act is accomplished, but pragmatic insincerity also occurs when an ut-

<table>
<thead>
<tr>
<th>Conditions for allusion</th>
<th>Utterances satisfying the condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ( U \sim Q ) and ( U \not\sim PE )</td>
<td>(1a) (1b) (2b) (3) (4a) (6a)</td>
</tr>
<tr>
<td>2. ( U \sim A ) where ( A ) is an action that brings about ( Q ) (type-1, type-2)</td>
<td>(1d) (4b)</td>
</tr>
<tr>
<td>3. ( U \sim X ) where ( X ) is a premise of ( Q ) or a precondition of ( A )</td>
<td>(1c) (4c)</td>
</tr>
<tr>
<td>4. ( U \sim Y ) where ( Y ) is an implication derived from ( Q )</td>
<td>(4d) (6b)</td>
</tr>
<tr>
<td>5. ( U \sim B, W ) or ( Z ) where ( B ) is an action that precludes ( A, W ) is a premise of ( B ) and ( Z ) is an effect of ( B ) (type-1, type-3)</td>
<td>(2a) (2c) (4e) (5a) (5b)</td>
</tr>
</tbody>
</table>
### Table 2: Pragmatic principles violated by ironic utterances

<table>
<thead>
<tr>
<th>Violated pragmatic principles</th>
<th>Utterances violating the principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sincerity condition for Inform (S believes (P)) (1a) (2a) (2b) (3) (4a) (4b) (4d) (6a) (6b)</td>
<td></td>
</tr>
<tr>
<td>for Question (S does not want to know (P))</td>
<td>(4c)</td>
</tr>
<tr>
<td>for Offer (S wants to do an action (P) for H)</td>
<td>(4e)</td>
</tr>
<tr>
<td>for Thank (S feels grateful for an action (P))</td>
<td>(2c) (5b)</td>
</tr>
<tr>
<td>Propositional content condition for Thank ((P) is a past action done by H) (4c)</td>
<td></td>
</tr>
<tr>
<td>Preparatory condition for Question (S does not know (P))</td>
<td>(4c)</td>
</tr>
<tr>
<td>for Offer (S is able to do an action (P))</td>
<td>(4e)</td>
</tr>
<tr>
<td>Maxim of relevance ((P) is relevant in Sperber and Wilson’s (1986) sense) (1b)</td>
<td></td>
</tr>
<tr>
<td>Politeness principle (U should be made at an appropriate level of politeness) (1d)</td>
<td></td>
</tr>
</tbody>
</table>

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

### Table 3: Examples of cues for implicitly displaying emotional attitudes

<table>
<thead>
<tr>
<th>Verbal cues</th>
<th>Nonverbal cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. hyperbole, exaggeration — adjectives (e.g., amazing, splendid), adverbs (e.g., certainly, really, absolutely), metaphors</td>
<td>1. facial expression — quizzical, sneering, deadpan</td>
</tr>
<tr>
<td>2. interjection — “Oh!”, “ah!”, “O!”, “Dear me!”, “Oh dear!”, “huh”</td>
<td>2. gesture</td>
</tr>
<tr>
<td>3. prosody(paralinguistic cues) — accent, intonation contour, exaggerated stress, slow speaking rate, tone of voice, nasalization</td>
<td></td>
</tr>
<tr>
<td>4. speaker’s counterfactual pleased emotions — thank, compliment</td>
<td></td>
</tr>
</tbody>
</table>

...continued...
tional irony has been found in a number of psychological experiments (Gibbs and O’Brien 1991).

### 3.3 What Does Irony Convey?

By recognizing an utterance to be ironic, the hearer understands an intended illocutionary act that the speaker performs in saying ironic: the act of informing 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). By understanding the illocutionary act, the hearer turns out to know that the informed fact is shared by both the hearer and the speaker.

In many cases, however, the hearer already knows 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. That is why previous irony theories regard the speaker’s negative attitude as what irony communicates. Therefore, when the hearer does not recognize all components, he/she also obtains new information that the unrecognized component holds in a current situation. For example, in the case of (2a), after recognizing Peter’s utterance (2a) 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.

Furthermore, together with that illocutionary act, irony achieves various communication goals intended by the speaker as perlocutionary acts, some of which are as follows (Roberts and Kreuz 1994):

- to give pain
- to emphasize a point
- to get attention
- to be humorous
- to clarify
- to dissemble

### 3.4 Implications of The Theory

#### 3.4.1 Echoic irony and expectation

It appears that some particular kind of echoic irony like (2a), which Cutler (1974) called “provoked irony”, does not allude to speaker’s expectation, since its propositional contents must be assumed to be desirable less from the point of view of the speaker than from that of the hearer. Therefore, in order to explain provoked irony, previous theories (e.g., mention theory and echoic reminder theory) assume that what such irony alludes to is not a speaker’s expectation but other person’s thought or utterance. However, what might be the echoed materials of the following utterances that can be interpreted as ironic under the same situation as (2a)?

(2) b. You are very smart, so you already think I’d be promoted before you.

c. Thank you for informing your valuable opinion.

To overcome this difficulty, implicit display theory argues that provoked irony also alludes to the speaker’s expectation: expectation that “the hearer knows his/her own utterance or thought is false”. Such expectation is often brought up by the hearer’s preceding utterance, and at the same time (i.e., \( t_0 = t_1 \)), the ironist becomes aware that the expectation is incongruous with the reality. Hence we may safely say that the act of saying the hearer’s thought or opinion corresponds to, B, an action that makes E incongruous.

In the case of Situation 2, by hearing Jesse’s utterance of “I’d be promoted before you”, Peter has an expectation of convincing Jesse that Jesse would not be promoted before Peter, knows that Jesse does not think his own utterance is false, and consequently, the situation is surrounded by ironic environment. Provoked irony (2a) alludes to that expectation by referring to one of the effects Z = “Peter believes that Jesse would be promoted before Peter”, which would normally be brought about by an action B of “informing that Jesse would be promoted before Peter”, and at the same time, it violates the sincerity condition of Inform in that Z is false. Also, (2b) and (2c), which do not allude to one's utterance/thought, can be explained within implicit display theory: both allude to that expectation, because the former refers to Peter’s expectation Q and the latter refers to Jesse’s action B.

You may doubt the appropriateness of the above explanation, but Kaufer (1981) mentioned the similar view in the discussion of what assumptions about the context ironize clearly false utterances like “Columbus discovered America in 1900”. He argued that, in order to be perceived as irony, such utterance must be given in the contextual setting in which “the ironist knows the utterance is false (and thus rejects it), knows that the addressee does not know this, and (most importantly) also believes that the latter should know it’ (ibid., p.503). The last assumption clearly corresponds to the speaker’s expectation of provoked irony explained above.

#### 3.4.2 Ironic cues

Several theories assume that irony can be identified by a number of cues specific to irony, in particular, ironic tone of voice and intonational cues. Although there is little doubt that such intonational cues may accompany ironic utterances, the empirical finding in psychology (Gibbs and
O’Brien (1991) has shown that people can interpret ironic statements without any special intonational cues. Furthermore, both ironic and non-ironic interpretations are derived from the use of the same intonational contour in different contexts (Ward and Hirschberg 1985). These observations imply that such cues themselves are neither sufficient nor necessary conditions for irony. Implicit display theory is consistent with this implication since these cues are only a part of Component 3 for implicit display.

Kreuz and Roberts (1995) have addressed another cue for irony, hyperbole, and they claimed that ironic tone of voice is nothing more than the use of hyperbole. However, the same discussion can be done about the hyperbolic features of irony: there are ironies that do not include hyperbole such as (1c). Implicit display theory also covers hyperbolic features of irony and explains non-hyperbolic irony.

3.4.3 Victims of irony

Several studies (e.g., Clark and Gerrig 1984) have pointed out that most (but not all) ironies have victims and thus that the notion of victims may be an important property for irony. Implicit display theory implies 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.

3.4.4 Sarcasm and irony

Sarcasm is a figure of speech designed to cut or give pain and often expressed by verbal irony. Thus explicit victims of irony often become the target of sarcasm, and by displaying the speaker’s counterfactual pleased emotion, which is one of cues for emotional attitudes, sarcasm has the effect of giving the target pain. We argue that both are distinct properties of sarcasm. For example, (5a) and (5b) are sarcastic ironies because they have an explicit victim, Peter, and they refer to the wife’s counterfactual pleased emotion.

4 TOWARDS A COMPUTATIONAL MODEL OF IRONY

4.1 Representation

In order to represent ironic utterances and ironic environment, this paper uses situation theory (Barwise 1989) and situation calculus. As an example, Figure 1 illustrates the representation of ironic environment of Situation 4. In the figure, all events/states of affairs are represented as pairs s \( \vdash \sigma \) of an infon \( \sigma \) and a situation \( s \). For example, given that the situation \( t_0 \) expresses a spatiotemporal location and \( x \) denotes “Candy”, the state of affairs that Candy is hungry is represented as \( t_0 \vdash \langle \text{hungry}, x \rangle \) and its negation as \( t_0 \vdash \lnot \langle \text{hungry}, x \rangle \). Actions are expressed by predicates: for example, an action of eating the pizza preformed by Candy’s husband is expressed by the predicate \( \text{eat}(y, a) \). The state of affairs that an action \( A \) is performed is expressed by \( \langle \text{did}, \langle \text{eat}(y, a) \rangle \rangle \). Furthermore, a proposition \( p \) expressing the claim that \( s \vdash \sigma \) is written as \( (s \vdash \sigma) \). The proposition \( p \) is true if \( s \vdash \sigma \) and otherwise false. Infons and actions can include parameters denoted by capital letters. Parameters can be restricted by infons: for example, \( T^1 \langle \text{precedes}, t_0, T \rangle \) is a parameter for temporal situations which temporally succeed \( t_0 \). A causal relation between two events \( s_1 \vdash \sigma_1 \) and \( s_2 \vdash \sigma_2 \) is expressed by \( s_1 \vdash \sigma_1 : \langle A \rangle \Rightarrow s_2 \vdash \sigma_2 \). This relation means that if an action \( A \) is executed in a situation \( s_1 \) supporting the infon \( \sigma_1 \), then it causes the infon \( \sigma_2 \) to be true in the resulting situation \( s_2 \). When we omit an action \( A \) from a causal relation, that relation becomes a constraint in situation theory, denoted by \( s_1 \vdash \sigma_1 
Rightarrow s_2 \vdash \sigma_2 \).

A conversational participant such as a hearer and a speaker has a finite number of beliefs in his/her mental situation that are manifest, then perceptible and made available to him/her. An agent \( X \)’s mental situation is represented as a situation \( u_X \) and his/her beliefs as support relations between \( u_X \) and infons. For example, the fact
Speech Act: Inform$(S,H,P)$
Preconditions: $u_H \models \langle proposition,P \rangle$, $u_S \models P$
Effects: $u_H \models \langle intend,S,Convince(S,H,P) \rangle$

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

Speech Act: Thank$(S,H,P)$
Preconditions: $\langle action,P \rangle$, $\langle agent,H,P \rangle$, $T \models \langle did,P \rangle$, $\langle precedes,T,T_u \rangle$, $u_S \models T \models \langle grateful,S,H,P \rangle$
Effects: $u_H \models T \models \langle grateful,S,H,P \rangle$

Notes: $S$ and $H$ denote the speaker and the hearer, and thus $u_S$ and $u_H$ denote the speaker’s and hearer’s mental situations. Also, $T_u$ denotes the utterance situation. $KnowIf(S,P) = u_S \models P \lor \neg P$, and $\neg KnowIf(S,P) = u_S \not\models P \land \neg P$.

Figure 2: Action schemes for speech acts

that Jim believes/knows Candy is hungry is represented as $u_{Jim} \models \langle hungry, x \rangle$. Although Figure 1 does not include any mental situations (i.e., ironic environment is represented from god’s eye view), when Candy intends the utterance to be ironic her mental situation must support all states of affairs, events and causal relations in this figure. These manifest beliefs constitute a context, which corresponds to what Sperber and Wilson (1986) call a cognitive environment.

An utterance $U$ is characterized by its propositional content $P$ and the illocutionary force $F$ that the speaker performs in saying $U$. For example, $(a)$ consists of the propositional content $P = \langle t_1 \models \langle see,y,T \langle precedes,T,T_u \rangle \models \langle loc,a,t_1 \rangle \rangle \rangle$ and the illocutionary act $F = \text{RequestIf}$. These illocutionary acts are defined as action schemes as in (Litman and Allen 1987), some of which are shown in Figure 2. However, in ordinary conversations, the intended illocutionary act of utterances often differs from such literal illocutionary act. Thus in this paper literal illocutionary acts are denoted by $F_i$ and the intended illocutionary acts of utterances by $F_f$. In the case of irony, as we mentioned in Section 3.3, $F_i$ is to inform the fact that the utterance situation is surrounded by ironic environment.

Perlocutionary communication goals that ironists usually intend to achieve by irony may affect the decision of how implicit display is accomplished (i.e., which proposition should be referred to, which pragmatic principle should be violated, and what cues should be used to convey the attitude). Conversely the hearer can infer what communication goals are intended by the speaker by examining the way of implicit display. For example, to achieve the goal give_pain by irony, the following strategies can be applied:

- $U$ refers to the victim’s action, its premises, or its effects.
- $U$ has a surface speech act Thank.
- $U$ expresses the counterfactual pleased emotion toward the victim’s action.

4.2 Algorithms for Interpreting and Generating Irony

A rough sketch of an interpretation algorithm is given in Figure 3. The algorithm takes as input an utterance $U$ under interpretation and a hearer’s context $W$, and it produces a set of goals $G$ the speaker has intended to achieve by $U$. It must be noted that we do not argue here that irony interpretation is an optional process done before other (literal) interpretation process (Utsumi 1995). The interpretation algorithm checks whether each of the three conditions for implicit display is satisfied by $U$, and then calculates the degree of ironicalness $d(U)$ by the method described in Section 4.3. At line 4, the algorithm judges that $U$ is ironic when $d(U)$ exceeds a given constant $C_{irony}$. The degree of ironicalness takes a real value between 0 and 3, and thus this criterion can be seen as a comparative version of the recognition condition described in Section 3.2. If $U$ is judged ironic, the algorithm then rea-
\( \langle \text{hope}, X, (S \models I) \rangle \leftarrow \langle \text{want}, X, (S \models I) \rangle \wedge \langle \text{anticipate}, X, (S \models I) \rangle \)

\( S_1 \models \langle \text{disappointed}, X, (S_1 \models I) \rangle \leftarrow \)

\( S_0 \models \langle \text{hope}, X, (S \models I) \rangle \wedge \langle \text{antecedent}, S_0, S_1 \rangle \wedge \langle \text{precede}, S_0, S \rangle \)

\( S_1 \models \langle \text{angry-at}, X_1, X_2, A \rangle \leftarrow \)

\( S_0 \models \langle \text{want}, X_1, (S \models I) \rangle \wedge \langle \text{antecedent}, A, X_2 \rangle \wedge \langle \text{precede}, S_0, S \rangle \wedge \langle \text{agent}, A, X_2 \rangle \wedge \)

\( S_1 \leftarrow \ast : [A] \Rightarrow S_1 \models I \wedge \langle \text{did}, A \rangle \wedge \langle \text{blameworthy}, A \rangle \)

Figure 4: Emotion-eliciting rules

GenerateIrony(\( G, W \))

0. \( G_F \leftarrow \phi \)

1. Select possible subgoals (plans) for achieving \( G \), and add them to \( G_F \).

2. Construct pairs \( (P, F_i) \) that implicitly display the ironic environment in \( W \) so that they satisfy \( G_F \).

3. Select the pair \( (P^*, F^*_i) \) with the highest degree of ironicness among them.

4. Construct \( U \) from \( (P^*, F^*_i) \).

Figure 5: Irony generation algorithm

sons about the speaker’s emotional attitude using emotion-eliciting rules, some of which are shown in Figure 4. Moreover, the algorithm tries to recognize communication goals using relations between the way of implicit display and the goals.

On the other hand, the generation algorithm shown in Figure 5 takes as input the set of the speaker’s communication goals \( G \) (including the ironic illocutionary goal \( F_1 \)) and a context \( W \), and then it generates an ironic utterance \( U \) appropriate to the communication goals and to the ironic environment. More concretely, the algorithm first selects more specific goals that serve for many decisions of what should be said ironically and of how it should be said. To take a simple example, a communication goal “to give pain” is decomposed into several subgoals such as “to refer to actions or events ascribed to the victim” and “to directly express the counterfactual pleasured emotion”. The next step is that the algorithm decides what to say and how to say ironically. The propositional content is mainly concerned with the decision of what should be referred to by irony, while the literal illocutionary act is mainly concerned with the decision of which pragmatic principle should be violated. Decision of what cues should be used to convey the attitude mainly affects stylistic and prosodic features. These decisions are done so that the selected subgoals are satisfied and that their degrees of ironicness is much more than a threshold.

4.3 A METHOD FOR ASSESSING THE DEGREE OF IRONICALNESS

To evaluate the degree of ironicness of \( U \), we use the following measures whose values are real numbers ranging from 0 to 1:

1. For the allusiveness of \( U \)
   (a) context-independent desirability of the referring expression \( U_R \): \( d_1 \)
   (b) degree of similarity between \( Q \) and \( R \): \( d_2 \)
   (c) expectedness of \( E \): \( d_3 \)
   (d) indirectness of expressing the fact that the speaker expects \( E \): \( d_4 \)

2. For pragmatic insincerity of \( U \)
   (e) degree of pragmatic insincerity of \( U \): \( d_5 \)

3. For emotional attitudes in \( U \)
   (g) degree to which \( U \) implies the speaker’s attitude: \( d_6 \)
   (h) indirectness of expressing the attitude: \( d_7 \)

The desirability \( d_1 \) reflects the asymmetry of irony that many studies have pointed out: more positive utterances are more likely to be ironic. On the other hand, some studies (Kreuz and Glucksberg 1989; Kumon-Nakamura et al. 1995) suggested that a negative utterance can also convey irony if the speaker’s negative expectation is incongruous with a positive event and if that is obvious to both the speaker and the hearer. The measure \( d_3 \) reflects this fact: whether they are positive or negative, utterances which allude to stronger or more obvious expectations are more likely to be ironic. Hence, although \( d_1 \) of the following irony (7) is much less than that of (1a), \( d_3 \) of (7) is nearly equal to that of (1a).

[Situation 7] A wife asked her husband to scold their son for leaving his room messy. They entered into his room and found that it was sparkling clean. As he looked around the room, her husband whispered to her:

(7) This room is certainly dirty!

Furthermore, the expectedness \( d_1 \) reflects an intuitively plausible hypothesis that personal expectations should be stronger than culturally/socially expected norms and conventions. The similarity \( d_2 \) serves to evaluate to what degree an utterance alludes to an expectation. It implies that an utterance referring to \( Q \) must have the greatest similarity \( d_2 = 1 \), and that the similarity of an utterance referring to the action \( A \) or \( B \) might be greater than that of an utterance referring to other propositions. When the interpretation algorithm does not decide a speaker’s expectation \( E \), it can be assumed that \( d_3 = d_4 = 0 \). The last measure for allusion, the indirectness \( d_4 \), is necessary for ruling out
non-ironic utterances that directly express the speaker’s expectation. Thus the value of $d_4$ for (1e) is 0.

The measure $d_5$ is equal to 0 when utterances do not violate any pragmatic principles, and otherwise it is greater than 0. For example, $d_5$ of provoked irony (2a) is 0 on the hearer Jesse’s side, while it is much greater on the speaker Peter’s side. In particular, $d_5$ should be 1 when utterances violate the sincerity condition like (1a), since it is the most frequently violated principle by irony. Likewise, the measure $d_6$ is 0 when utterances do not include any cues listed in Table 3, and otherwise it is greater than 0. The last measure $d_7$ is also necessary for ruling out non-ironic utterances like (1f) that directly express the speaker’s attitude.

Using these seven measures, we define the degree of ironicalness $d(U)$ of an utterance $U$ as follows:

$$d(U) = d_4 \cdot d_7 \cdot \left( \frac{d_1 + d_2 + d_4}{3} + d_5 + d_6 \right)$$

This equation means that direct expressions of expectations and of emotional attitudes cannot be ironic even if they implicitly display other components. Three measures $d_1 \sim d_3$ for allusion are averaged so that the three components for implicit display can equally contribute to the degree of ironicalness.

5 Concluding Remarks

This paper has presented implicit display theory of verbal irony and illustrated a rough sketch of a computational model of irony. Through the discussion about many features of irony, we have confirmed the superiority of the theory over existing irony theories and also shown that irony is a more complicated pragmatic phenomenon than was supposed. To deal with irony by computer, several computational techniques in NLP, especially methods for treating pragmatic phenomena — e.g., referring expressions, speech acts, plan recognition and inference, conversational implicatures, politeness — are required to be cooperatively implemented in a computer program.

References


