

アイロニー理解の計算モデル

内海 彰

東京工業大学 大学院総合理工学研究科知能システム科学専攻

utsumi@utm.dis.titech.ac.jp

1 はじめに

アイロニーは一見現実に反することを言いながら真意をほのめかす非字義的な言語表現である。例えば、以下の発話はいずれもアイロニーとして機能する。

状況1 妻がおなかが減って眠れないので食べようとしたピザをまるごと平らげてしまった夫に向かって

- (1) a. すごくおなかがいっぱいだわ。
- b. ああ、ピザが食べれて満足!
- c. ピザはここにあったのを見なかった?
- d. これでぐっすり眠れるわ。
- e. ピザをのこらず食べてくれてありがとう。

このようなアイロニカルな発話は、夫の行為への非難やいやみを伝えながらも直接的な表現に比べて相手に与えるダメージが弱まるなど対人関係の維持に効果的であり、それゆえにアイロニーは知的で機知に富んだ会話や文章には欠かせない言語現象である。しかし現状ではアイロニーの計算機処理を指向した研究はいくつかの例外 [7] を除いて皆無である。

そこで本稿では筆者が提案している暗黙的提示の概念 [8, 9] を用いて、上記のようなアイロニー表現の解釈過程の計算モデルを提案する。本稿のモデルはアイロニーに関するいくつかの重要な知見に基づいている。

2 アイロニー環境の暗黙的提示

暗黙的提示理論では、「アイロニーは現在の発話状況がアイロニー環境 (ironic environment) によって囲まれた状況であることを聞き手に暗黙的に提示する (implicitly display) 言語表現である」と考える。この定義で重要なのは、従来のアイロニー論でしばしば混同されてきたアイロニーであるために必要な状況設定とアイロニーの表現上の特徴を明確に区別している点である。なお従来のアイロニー論に比べて優れている点などの暗黙的提示理論の詳細については [8, 9] を参照されたい。

発話状況が以下の3つの条件を満たしているとき、その状況はアイロニー環境によって囲まれていると言う。

1. 時間 T_0 において話し手がある期待 E を持っている。
2. T_0 と同じかまたは T_0 より後の時間 T_1 において話し手の期待 E が満たされていない。
3. 上記の期待と現実との間の不一致に対して、話し手が否定的な心的態度や感情 (失望, 怒り, 非難, ねたみなど) を持っている。

例えば状況1は話し手である妻の期待「空腹をみたく」が夫の行為 (ピザを食べる) によって満たされており、妻がそれに対して不満や怒りを抱いているので、アイロニー環境によって囲まれた状況であると言える。

そして言語表現や発話 U が以下の3条件を満たすとき、 U はアイロニー環境を暗黙的提示するという。

1. U は話し手の期待 E に婉曲的に言及 (allude) する。
2. U は何らかの語用論の原則に表面上違反することによって語用論的不誠実性を含む。
3. U は話し手の心的態度・感情を暗示する。

条件1の婉曲的言及 (allusion) は言及理論 [6, 10] などのアイロニー論の言及 (mention) の概念を大幅に拡張したものであり、「原因・結果」や「可能化」などの接続関係 (coherence relation) によって定義される。具体的には、 P を言語表現 U の命題内容を表す命題または行為、 Q を期待 E の内容を表す命題とすると、 P または P の構成要素となる命題や行為 (これらをまとめて P_i と表す) と Q を関係づける接続関係の連鎖が存在するとき、言語表現 U は期待 E に婉曲的に言及すると定義する。例えば、状況1のそれぞれの発話は図1に示すように妻の期待に婉曲的に言及している。

語用論的不誠実性は、従来のアイロニーへの語用論的アプローチ [3] で用いられている選択制限違反や Grice の質の公準違反を拡張した概念であり、少なくとも何らかの語用論的な原則に表面上違反しているとき、その発話は語用論的に不誠実になる。違反される原則には質の公準や言語行為の適切性条件 (誠実性, 準備, 命題内容条件) だけでなく、関連性の原理, 丁寧さの原理や量の公準なども含まれる。例えば以下の発話では、

状況2 散らかっている部屋をきれいにしなさいと注意した子が何もしない息子に対して母親が

- (2) a. お母さんは部屋をきれいにしている子が好きよ。
- b. もうしわけありませんが、どうか部屋を片付けて頂けませんか?
- c. ちょっと散らかっているみたいね。

一般化表現である (2a) は、自分の部屋をきれいにしている子供が存在する文脈において (2a) が「母はその人が好きである」を含意する点で関連性の原理 [6] を満たすことができるが、状況2の場合にはそのような文脈ではないので関連性の原理に違反している。同様に、過度に丁寧な表現 (2b) は適切な丁寧さで発話を行なうとい

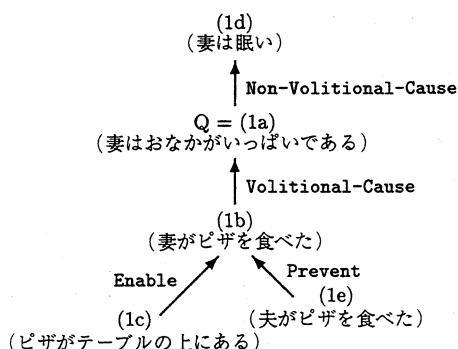


図 1: 状況 1 における婉曲的言及

う丁寧さの原理に違反しているし、控え目な表現 (2c) は量の公準に違反している。

条件 3 の話し手の心的態度・感情の暗示はさまざまな標識の使用によって達成される。これらの標識には修飾語句による誇張表現 (例: (1a) の「すごく」) やアイロニカルな意味を含む語彙 (例: 「おめでたい」), 感嘆詞 (例: (1b) の「ああ」), 韻律上の特徴 (イントネーション・音調やアクセントなど) などの言語的情報の他に表情や身振りなどの非言語的な標識も含まれる。さらに発話 (1e) のように感謝や賛辞などの実際とは反対の感情を表出しても本当の心的態度が間接的に伝達される。

3 アイロニー理解の計算モデル

3.1 アイロニーの認識と解釈

アイロニーは比喩などの他の非字義的な表現と異なり、アイロニーでない表現との境界が曖昧である。例えば、相手の発話に対して「それって皮肉？」とは言いが、「それって比喩？」とは言わない。よってアイロニーの解釈機構をモデル化するためには、明らかにアイロニーでない表現を除外したあとに何らかの評価基準を用いてアイロニーである可能性が高い表現を検出するという方が望ましい。

前章の定義から明らかにアイロニーでない場合として (a) 発話状況がアイロニー環境ではない, (b) アイロニー環境を直接的に表現している, の 2 つが考えられる。例えば以下の発話 (1f) の下線部は (1a) と同じ表現であるが、明らかにアイロニカルな意図を含んでいない。

状況 1' 妻はおなかが減って眠れないのでピザを食べようとしたが、夫がまるごと平らげてしまった。それに気付いて「ごめんね」と謝った夫に対して

- (1f) 気にしないでよ。 すごくおなかがいっぱい だわ。

これは妻の「気にしないで」という発話によりアイロニー環境の条件 3 の否定的な態度が成立しないためである。同様に以下の発話 (1g) は話し手の期待を直接的に表現しているの、たとえアイロニー環境である状況 1 で発せられたとしてもアイロニーではない。

- (1g) おなかが減ったからピザを食べたかったのに。

次にアイロニーである可能性が高い表現とは何かを求めるために、アイロニーに関するいくつかの経験的な知見を考えてみよう。

- (A1) 聞き手がアイロニー環境であることに気付いていなくても、アイロニーを理解することができる。
- (A2) 話し手がアイロニーと意図していない表現を聞き手がアイロニーと解釈する場合がある [1, 2]。
- (A3) 特別なアイロニー標識がなくてもアイロニーと解釈されるし [2], 逆にこれらの標識はアイロニー以外の発話にも使われる [1]。
- (A4) 字づらの意味が肯定的な (positive) 内容のものほどアイロニーとして解釈されやすいが、言及の対象となる期待が話し手と聞き手双方にとって明らかな状況下では、以下の (3) のような否定的な (negative) 表現でもアイロニーと容易に解釈される [4]。

状況 3 いくら注意しても息子が部屋を散らかしたままなので、妻が夫に息子を叱ってほしいと頼んだ。そこで 2 人は彼の部屋に行ってみると、部屋はきれいに片付いていた。夫が妻に向かって

- (3) おお、なんと汚い部屋だ。

(A1) は解釈前にアイロニー環境の成立が既知である必要はないことを示しており、さらに (A2) によって聞き手がアイロニーと解釈することによってアイロニー環境の 3 条件を推論していることがわかる [2]。つまりアイロニーである可能性が高いと判断するためには、アイロニー環境が満たされる可能性があればよい。さらに (A2) や (A3) から暗黙的提示の 3 条件がすべて同定される必要はないことがわかる。また (A4) は、話し手の期待が既知の場合には、その期待に言及していれば (内容が肯定的か否定的かにかかわらず) アイロニーと理解されるが、話し手の期待が未知のときには表現内容が肯定的であるほどアイロニーである可能性が高くなると解釈できる。これらの知見からアイロニーの可能性が高いと判断する基準として、言及の婉曲度 d_a 、不誠実性 d_i 、感情の暗示度 d_e 、話し手の期待の明白度 d_o 、 U の内容の評価値 d_d の 5 つの尺度を用いた言語表現 U のアイロニー度 $d(U)$ を以下の式で定義する。

$$d(U) = d_o \cdot d_a + (1 - d_o) \cdot d_d + f(d_a, d_o) \cdot (d_i + d_e)$$

ただし、 $f(d_a, d_o) = \begin{cases} 1 - d_o & (d_a = 0 \text{ のとき}) \\ 1 & (\text{その他}) \end{cases}$

この式の最初の 2 項が婉曲的言及、残りの項が不誠実性と心的態度の暗示をどれだけ満たしているかを表現しており、さらに話し手の期待が未知 ($d_o = 0$) の場合には (A4) を満たすように婉曲的言及の度合いを表現内容の評価値で置き換えている。5 つの尺度はすべて 0 から 1 までの実数値なのでアイロニー度は 0 から 3 までの実数値をとり、上記の議論よりアイロニー度の値が 1.5 より大きい場合には U がアイロニーであると判断する。

以上をまとめると、アイロニー解釈の過程として図 2 を考えることができる。アイロニーの可能性が高いと判

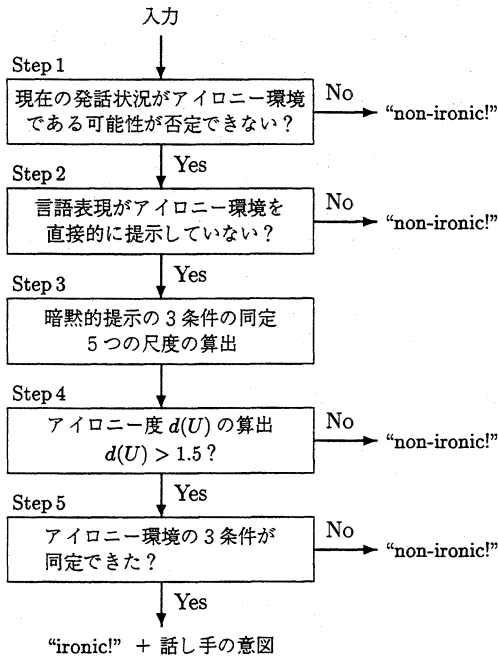


図 2: アイロニー解釈アルゴリズム

断されれば、Step5で既知である情報を用いて既知でない条件の成立を推論することによって、アイロニー環境の3条件が成立していることを同定する。すべての条件が同定できれば、現在の発話状況がアイロニー環境であることを伝達するというアイロニーの遂行する発語内行為を出力する。なおStep5において話し手の期待を推測できなかった場合や、推測はできたけれども現実との不一致がなかったりそれに対して否定的な態度を持っていない場合には、アイロニーではないと判断する。

3.2 アイロニーの解釈アルゴリズム

図2の解釈アルゴリズムが入力として受け取るのは1) 言語表現 U の命題内容 P , 2) U の字義的な発語内行為 F , 3) U の意味表現 M , 4) 発話状況 W と共有知識 K から成る聞き手の信念の集合 C であり、これらのうち1)~3)は適当なパーサによって得られると仮定する。言語表現の命題内容 P や発話状況 W を構成する聞き手の信念は図3に示すような論理式で表現される。論理式 $F = (! = (B \ T) \ I)$ はできごとや状態の内容 I と状況 $(B \ T)$ の支持関係を表現する。状況はその信念の主体を表す信念空間 B とできごとの時間 T から構成される。信念空間としては“H”（聞き手の信念），“SH”（話し手の信念に関する聞き手の信念）や“MBH”（聞き手側の共有信念）などを用いる。また論理式 $(neg \ F)$ は $(! = (B \ T) \ (not \ I))$ を表し、 $\neg F$ は $(B \ T)$ が I で不成立であることを意味する。さらに $(ascribe \ ?P \ ?X)$ は信念 $?P$ の主体を $?X$ に帰した結果得られる論理式を表す。変

```

(! = (MBH ?T) (wife x y))
(! = (MBH ?T) (husband y x))
(! = (MBH ?T) (pizza a))
(! = (MBH ?T) (eatable a))
(! = (MBH ?T) (accessible x Lt))
(! = (MBH ?T) (accessible y Lt))
(! = (MBH ?T) T1>T0)
(! = (MBH T0) (located a Lt))
(! = (MBH T1) (did eat(y,a)))
(! = (MBH T1) (not (did eat(x,a))))
(! = (SH T1) (blameworthy eat(y,a) x))
(! = (H T0) (hungry x))
(! = (H T0) (hope x (! = (?* ?T:>T0)
(not (hungry x)))))
  
```

図 3: 状況1における発話状況

Domain Knowledge:

```

(=> (! = (?B ?T0) (and (accessible ?X ?L)
(located ?A ?L) (eatable ?A)))) eat(?X,?A)
(! = (?B ?T1) (and (not (hungry ?X))
(not (located ?A ?L)))))
  
```

Speech Act Schemes:

```

(=> (and (ascribe ?P ?S) (speaker ?S) (hearer ?H)
(propositional-content ?P)) Inform(?S,?H,?P)
(! = (H ?T1) (intend ?S Convince(?S,?H,?P))))
  
```

Emotion-Eliciting Rules:

```

(<=> (! = (?B ?T) (hope ?X ?I))
(! = (?B ?T) (and (want ?X ?I) (expect ?X ?I))))
(=> (and (! = (?B ?T0) (hope ?X (! = (?B1 ?T:>T0)
(not ?I))))
(! = (?B ?T1:>T0) ?I))
(! = (?B ?T1) (disappointed ?X (! = (?B1 ?T) ?I))))
(=> (! = (?B ?T) (and (did ?A) (blameworthy ?A ?X)))
(! = (?B ?T) (reproach ?X Agent(?A) ?A)))
  
```

図 4: 本モデルで用いられる共有知識の例

数は“?”を用いて表現し，“:”に続けて条件を指定することもできる。命題内容 P は行為を表す述語によっても表現される。例えば、(1a)と(1e)の命題内容はそれぞれ状態 $(! = (?* \ T1) \ (not \ (hungry \ x)))$ と行為 $eat(y,a)$ である。なお(1a)の命題内容は特定の主体の信念ではないので、信念空間は変数“?”になっている。

聞き手と話し手が共有していると仮定する知識 K は図4に示すように領域知識、言語行為に関する知識、感情導出規則から構成される。すべての知識はできごとや状態間の因果関係 $(=> (! = Sit1 \ I1) \ A \ (! = Sit2 \ I2))$ で表現され、 $(! = Sit1 \ I1)$ という前提のもとで行為 A を遂行すると $(! = Sit2 \ I2)$ が成立するという関係を意味する。行為 A を省略すると規則 $(=> (! = Sit1 \ I1) \ (! = Sit2 \ I2))$ となり、これは無意志的な因果関係と考える。言語行為に関する知識としては *Inform*, *Ask-if*, *Ask-ref*, *Request*, *Offer*, *Thank*, *Apologize* の7つが用意されている。感情導出規則は[5]で提案された規則を修正したものを用いている。本モデルでは“expect”, “want”, “hope”の感情要素で「期待」を表現し, “disappointment”, “angry”, “reproach”の3つの

感情要素で話し手の否定的な態度を表現する。

図2のStep1では、話し手の期待 E が W に含まれている場合にはアイロニー環境の他の2つの条件のいずれかが不成立であることが C から帰結されるかを調べ、 E が見つからない（つまり期待が未知である）場合にはそのままStep2に進む。例えば図3の状況1において $(! = (?* ?T: ?T > T0) (not (hungry x)))$ が C から帰結されれば期待と現実の不一致が成立していないことになる。Step2では命題内容 P が C から帰結されれば直接的な表現であると判断する。

Step3では婉曲的言及があるかどうかを探索手法によって調べる。この際に探索木の深さを5に制限しており、すべての P_i に対してこの制限内で解が見つからない場合には、 U は話し手の期待に婉曲的に言及していないと判断する。なお E が未知の場合には探索は行なわない。語用論的不誠実性についてはまず F のすべての適切性条件を調べ、(i) 話し手の信念 SH から $(not P)$ が帰結されるか、(ii) 話し手の信念からは P と $(not P)$ のいずれも帰結されないが聞き手自身の信念 H から $(not P)$ が帰結される場合には、 P が適切性条件を違反していると判断する。これらの違反が見つからない場合には関連性の原理などの他の語用論の原則の違反を調べる。なお関連性の原理の違反は関連性の定式化[6]に基づいているが、丁寧さの原理や量の公準については現在の時点では経験的に違反と思われるパターン（例：「申し訳ありませんが〜」）との照合しか行っていない。

またStep3ではアイロニー度の算出に必要な5つの尺度の値を以下のように計算する。

- 接続関係の連鎖の長さ n により $d_a = 1 - 0.1n$ とする。婉曲的言及がないときは、 $d_a = 0$ とする。
 - U が上記の条件 (i) により誠実性条件違反と判断されたときは $d_i = 1$ 、条件 (ii) の場合には $d_i = 0.8$ とする。 U が他の語用論的な原則を違反しているときには $d_i = 2/3$ 、違反が検出されないときは $d_i = 0$ とする。
 - 検出された標識の個数 n より $d_e = 1.0 (n > 1)$ 、 $d_e = 2/3 (n = 1)$ 、 $d_e = 0 (n = 0)$ とする。
 - E がStep1で見つかったときには、それが共有信念であれば $d_o = 1$ とし、それ以外は $d_o = 0.8$ とする。 E が見つからないときには $d_o = 0$ とする。
 - 意味表現 $M = (Rel: sem1, Theme: M1)$ に対して、 $d_d(M) = e(sem1) + 0.5 \times d_d(M1)$ とする。ただし $e(sem1)$ は $(sem1)$ の評価値/3を表し、評価値は3から-3までの整数である。また $d_d(Rel: not, Theme: M) = -d_d(M)$ とする。そして $d_d(M) > 1$ のとき $d_d = 1$ 、 $d_d(M) < 0$ のとき $d_d = 0$ 、それ以外は $d_d = d_d(M)$ とする。
- 例えば発話 (1a) を入力文とすると、図1より婉曲的言及は $d_a = 1 - 0.1 \cdot 0 = 1$ となる。また C は $(! = (SH T1) (not (hungry x)))$ と $(! = (SH T1) (hungry x))$ のどちらも帰結しないが $(! = (H T1) (hungry x))$ を帰結するので、条件 (ii) より $F = Inform$ の誠実性条件の違反があると判断され、 $d_i = 0.8$ となる。さらに U は標識を含まないので $d_e = 0$ 、Step1で見つかった

期待は共有信念ではないので $d_o = 0.8$ となる。尺度 d_d は $- \{e(hungry) + 0.5 \times M((Pro ?x (I, ?x)))\} = -(-2/3 + 0) = 0.67$ と算出される。

Step4で3.1節のアイロニー度を算出しアイロニーかどうかを判断してから、Step5でアイロニー環境の3条件を感情導出規則などを用いて推論することによってアイロニーの発語内行為を得る。特に話し手の期待が未知である場合には、Step3と同じ探索手法により P に最も近い肯定的な内容をもつできごと・状態を話し手の期待とする。この際に、あるできごと・状態が一般的に望ましいかどうかは d_d の値を計算することによって判断する。例えば (1a) の場合には、 $d(U) = 0.8 \cdot 1 + (1 - 0.8) \cdot 0.67 + 1 \cdot (0.8 + 0) = 1.73$ からStep5に進み、以下の発語内行為を出力する。

```
Inform(x,y,(and (! = (SH T0) (hope x (! =
    (?* ?T: ?T > T0) (not (hungry x))))))
    (! = (SH T1) (hungry x))
    (! = (SH T1) (angry-at x y eat(y,a))))
```

4 おわりに

本稿では心理学や言語学で得られている知見に基づいたアイロニー解釈の計算モデルを提案した。しかし現時点ではまだ改良の余地が多く残されている。主な課題としては、アイロニーのさまざまな伝達目標の解釈、韻律的な特徴や非言語的標識の考慮、人間の判断との比較による本モデルの評価などが挙げられる。今後はこれらをふまえて解釈モデルを改良していくつもりである。

参考文献

- [1] Barbe, K. *Irony in Context*. John Benjamins Publishing Company, 1995.
- [2] Gibbs, R. *The Poetics of Mind*. Cambridge University Press, 1994.
- [3] Grice, H. *Studies in the Way of Words*. Harvard University Press, 1989.
- [4] Kreuz, R.J., and Glucksberg, S. How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General*, 118(4):374-386, 1989.
- [5] O'Rourke, P., and Ortony, A. Explaining emotions. *Cognitive Science*, 18:283-323, 1994.
- [6] Sperber, D., and Wilson, D. *Relevance: Communication and Cognition*. Basil Blackwell, 1986.
- [7] 滝澤 修, 伊藤 昭. アイロニー表現検出の手法. 人工知能学会誌, 9(6):875-881, 1994.
- [8] Utsumi, A. A unified theory of irony and its computational formalization. In *Proc. of COLING-96*, pp.962-967, 1996.
- [9] Utsumi, A. Implicit display theory of verbal irony: Toward a computational model of irony. In *Proc. of IWCH'96*, pp.29-38, 1996.
- [10] Wilson, D., and Sperber, D. On verbal irony. *Lingua*, 87:53-76, 1992.