

日本人英語学習者の談話における共参照関係の記述方法の検討

谷村緑 (NICT/京都外大), 和泉絵美 (NICT), 竹内和広 (大阪電通大), 井佐原均 (NICT)

{mtanimura, emi, isahara} @nict.go.jp / takeuchi@isc.osakac.ac.jp

1 はじめに

本発表の目的は, 日本人英語学習者の談話における共参照関係の記述方法を既存のタグ体系から検討することである. 共参照を記述するタグ体系には, 例えば, MUC-7 (Hirschman & Chinchor 1997[1]), UCREL (Figelstone 1992[2]), MATE (Poesio 2000[3]), MEDSTRACT (Castaño et al. 2002[4]), MMAX (Muller & Strube 2001[5]) などがある. これらのタグ体系は, 心理言語学, テキスト言語学などにみられる従来の研究と同様に先行詞の同定を問題としており, 指示対象を表出した表現を利用して記述することを試みている. しかし, このような対象は原則として多様性があることが記述上の問題として残っている. また, 我々が対象とする英語学習者の指示の現象を記述するためには, これらの先行研究を参考に問題を再定義する必要がある. ここでは, the NICT JLE Corpus の対話データを利用して, 学習者データにあった記述方法を検討する.

2 共参照関係

2.1 定義

共参照関係(coreference)とは, 2つの名詞表現が言語外の同じ談話対象(discourse entity)を指すときのその2つの関係をいう. これは, ある文連続の中の2つの言語形式が同一の対象, 又は類似したものや出来事を指示している場合, つまり先行詞-照応詞の関係が生じる場合(anaphoric relation)とは区別される(Kibble & van Deemter 2000)[6].

発話の連鎖という観点から考えると, 対話者は

同じ対象を指し示す必要があるが, Grosz(1977)[7]やSidner(1983)[8]は焦点(focus)¹, という概念を取り入れている. トピックとして焦点化された対象は, それ以降に用いられる名詞表現の内容に対して制約として働き, 談話の流れを示す. 逆に, 共参照関係を示す名詞表現は, 対話者に対して焦点を記す役割を果たす.

2.2 先行詞の認定

先行詞を認定するには, 指示詞が何を指すのか, なぜそれを指すといつて良いのかが問題となる(Webber 1983[9]; Garnham, Oakhill & Cruttenden 1992[10]; Vonk, Hustinx & Simons 1992[11]; O’Braien, Raney, Albrecht & Rayner 1997[12]). 例えば Webber(1983)は, 指示詞が談話対象を引き出すとする. これは, 指示詞に出会うと先行詞が活性化され, 一貫性が維持されることを意味する. 対話者は与えられた表現から指示する対象の範囲を決め, 指示する対象候補の中から一つを選び出す. 先行詞が活性化された概念にないときには, 並行的に検索が生じ, そこで, 談話からの先行の情報が指示の手がかりに対してチェックされる. 与えられた表現に指示する対象が発見できない場合には, 外界や知識の中で検索が行なわれ, それと引き出された談話情報とが結合される.

しかし, 心理学での記憶の実験では, 似たような指示表現を含む文を提示されると, 正しい談話

¹ Focus “refers to the influence of a listener’s memory for the linguistic form of an utterance (the actual words and the syntactic structure) on his interpretation of a subsequent utterance” [Grosz, 1977:5]

対象を検索し、同定することは難しく、頻繁に誤って回答をすることが指摘されており(Clark and Clark 1977[13]), 先行詞の決定に困難が伴うことを示唆している。

3 記述の注意点

3.1 先行詞の認定が難しい例

2.2 節を踏まえて、ここでは、先行詞の認定に伴うタグ付けの注意点について検討する。以下の例は、先行詞の認定が難しいと思われるものである。

- 1) Max split a bucket of water. He split *it* all over the rug.
- 2) The man carrying a blue sports bag stole a car. *He* was arrested a week later.
- 3) Mary broke a Chinese vase. *That* was valuable.
- 4) John cried as if a lion were growling. I saw *him* in a dream last night.

これらの 4 つの例文の特徴は以下のようにまとめることができる。

- A) 先行詞が特定できない(例 1) a bucket of water か a bucket か, 例 2) the man か the man carrying a blue sports bag か, 例 3) a Chinese vase かそれとも 1 文全体か).
- B) 談話対象が同じかどうか特定できない(例 2) 一週間後の彼がスポーツバッグを背負っているとは考えにくい).
- C) 談話対象が 2 つの要素の部分的統合であるため特定できない(例 4) (Yasuhara 2006) [14].

3.2 既存のタグ体系

3.1 節で指摘したように、先行詞の認定は容易ではないが、既存のタグ体系がこれらの現象をどのように扱っているかを MEDSTRACT (Castaño et al. 2002)を例に概観する。彼らは、分子生物学

や医学の分野のテキストに、指示詞と先行詞のタグを付与している。

- 5) Co-culture experiments (using porous membrane insert wells) of <Entity id="40" Type="cell type">fibroblasts</Entity> with <Entity id="41" Type="cell type">carcinoma cells</Entity> demonstrated that growth of <Entity id="42" Antecedent="40, 41">both populations of cells</Entity> was increased compared with <Entity id="43" Antecedent="40, 41">either</Entity> grown in isolation

例 5)では、both populations of cells は、fibroblasts と carcinoma cells ということになる。このように、共参照のタグ付けは、コーパス作成者が当該の分野における、同一対象を指示すると判断する基準を用意する必要がある。本稿での議論は、日本人英語学習者と英語母語話者との同一指示における「ずれ」をコーパス中に記述することである。そこで本稿では、まず、対象を限定して、英語教育の立場から、学習者の発話例における共参照現象を記述することを試みたい。

4 記述方法と結果

4.1 共参照関係の記述対象

本研究では、初出の談話対象が唯一的に決定できるものだけを対象に、共参照関係を記述する。

対象：

- 初出：固有名詞・my+名詞・your+名詞
(修飾語句を含まないもの)
- 既出：人称代名詞(he, she)(単数形のみで、it は含めない)

4.2 分析対象

the NICT JLE Corpusを利用してタグ付与を行う。このコーパスには、幾つかのタスクをベースにした面接官と受験者との対話が収録されている。Level 2(poor)からLevel 9(good)までの各レベルからランダムに10発話を選定し²、また、英語母語話者(NS)の発話を同様に10発話選定して、タグ付与を行った。表1は、レベル別の共参照関係出現数を表している。発話総語数は、言語能力の向上とともに増えているが、今回の調査項目の出現は必ずしも同じように増えてはいない。全く出現しない対話もあり、2回以上出現した対話は4対話のみである。

表1 レベル別共参照関係出現数

	対話数	発話 総語数	共参照関係 出現対話数	共参照関係 出現対話数 (2個以上)
Level 2	10	9125	0	0
Level 3	10	12060	2	0
Level 4	10	14735	3	0
Level 5	10	17152	3	1
Level 6	10	17296	4	0
Level 7	10	20402	4	1
Level 8	10	21311	4	0
Level 9	10	21588	7	0
NS	10	58543	4	2

4.3 共参照関係の例

○ 共参照がモノローグに出現する場合：

例6)は、焦点化されている主人公が共参照関係で、談話対象として想起され続けることにより発話の一貫性が保たれている。“de”は談話対象(discourse entity)を示す。

6) One day last week, at eight o'clock in the morning <DE_anchor ID="1" de="1">Mr. Suzuki</DE> is on his way to work as usual. And <DE ID="2" de="1">he</DE> got to the station. And it is really crowded as the usual morning. And <DE ID="3" de="1">he</DE> was standing next to two gentlemen. And the two people were really engaged in a conversation. And one guy unintentionally hit <DE ID="4" de="1">Mr. Suzuki</DE>. (file 255, level 7)

○ 共参照が2者の中で共有される場合：

例7)は、友達からのパーティへの招待を断る場面であるが、焦点の当たっている談話対象を対話者間で共有することにより発話の一貫性が保たれている。

7) A: I love <DE_ID="2" de="1">my mother</DE>. So I can't go there. I'm so sorry.
B: Yeah, I understand your situation, but actually this is a party for you.
A: Oh I forgot it. What should I do? O K. Another day, I'm going to hold a party for them.
B: Um-hm.
A: And can you excuse for that?
B: O K. Anyway, is <DE_ID="3" de="1">your mother</DE> O K? (file 340, level 5)

○ 共参照関係が再度出現する場合：

例8)では、名詞表現が、記憶の片隅に残っている談話対象を活性化させ、検索を促し、発話の一貫性が保たれている。

8) A: But today, er of course, you know, I have a test you know, having test, and then er I go I have to pick up <DE_anchor ID="1" de="1">my son</DE> at two o'clock to a kindergarten.

² Level 1 は収録数が少ないため、分析対象から除外した。

B: So this interview's almost over. And you said you are going to pick up <DE ID="3" de="1">your son</DE>.

A: Yes.

B: And play with <DE ID="4" de="1">your son</DE>?

5 まとめ

本研究では、日本人英語学習者の談話における共参照関係の記述方法を既存のタグ体系から検討した。共参照関係の記述における注意点を指摘し、現象を限ってタグ付与を行った。学習者の場合、発話の意味が不明瞭なために、記述が困難になる場合がある。今後このような場合の記述方法やその他の現象の記述方法も検討し、学習者の指示表現の特徴づけを更に行いたい。

参考文献

- [1] Hirschman, L., & Chinchor, N. 1997. MUC-7 Coreference Task Definition, Version 3.0. In *Proceedings of MUC-7*.
http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/muc_7_toc.html
- [2] Fligelstone, S. 1992. Developing a Scheme for Annotating Text to Show Anaphoric Relations. In G. Leitner (ed.), *New Directions in Corpus Linguistics*. pp.153-170. Berlin: Mouton de Gruyter.
- [3] Poesio, M. 2000. "Coreference" in *MATE Annotation Guidelines*.
http://www.ims.uni-stuttgart.de/projekte/mate/mdag/cr/cr_1.html
- [4] Castaño, J., Zhang, J., & Pustejovsky, J. 2002. Anaphora Resolution in Biomedical Literature. In *International Symposium on Reference Resolution*.
<http://medstract.org/papers/coreference.pdf>
- [5] Muller, C., & Strube, M. 2001. Annotating Anaphoric

and Bridging Expressions with MMAX. In *Proceedings of the 2nd SIG dial Workshop on Discourse and Dialogue*. pp.90-95. Aalborg, Denmark.

- [6] Kibble, R., & van Deemter, K. 2000. Coreference Annotation: Whither? In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, Athens, Greece.
- [7] Grosz, B. J. 1977. *The Representation and Use of Focus in Dialogue Understanding*. Technical Report 151, Artificial Intelligence Center, SRI International, Menlo Park, California.
- [8] Sidner, C.L. 1983. Focusing for Interpretation of Pronouns. *American Journal of Computational Linguistics*, 7, 217-231.
- [9] Webber, B. L. 1983. So What We Talk about Now? In M. Brady & R. C. Berwick (eds.), *Computational Models of Discourse*. pp. 331-371. MIT Press, Cambridge, Mass. / London: The MIT Press Series in Artificial Intelligence.
- [10] Garnham, A., Oakhill, J., & Cruttenden, H. 1992. The Role of Implicit Causality and Gender Cue in the Interpretation of Pronouns. *Language and Cognitive processes* 7 (3), 231-255.
- [11] Vonk, W., Hustinx, L. G. M. M., & Simons, W. H. G. 1992. The Use of Referential Expressions in Structuring Discourse. *Language and Cognitive Processes*, 7 (3), 301-333.
- [12] O'Braien, E. J., Raney, G. E., Albrecht, J. E., & Rayner, K. 1997. Processes Involved in the Resolution of Explicit Anaphors. *Discourse Processes*, 23, 1-24.
- [13] Clark, H. H., & Clark, E. 1977. *The Psychology of language : An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- [14] Yasuhara, K. 2006. *On Blended Referent*. 関西言語学会第31回大会発表レジュメ. 甲南大学.