

英語学習者の話題結束性に留意した談話タグの検討

谷村 緑 和泉 絵美 竹内 和広 井佐原 均

独立行政法人 情報通信研究機構(NICT)

{mtanimura, emi, kazuh, isahara} @nict.go.jp

はじめに

英語教育では、学習者の実践的コミュニケーション能力の育成を図ることが求められており(学習指導要領)、授業活動内容だけでなく、スピーキング能力の評価に対する関心も高い。本稿では、円滑なコミュニケーションを図るために必要な、話題の中心(=話題の結束性)の観点からの学習者コーパス(The NICT JLE Corpus)の特徴付けについて考える。分析対象とする The NICT JLE Corpus は、英語学習者のスピーキングデータのコーパスとしては最大規模のもので、対話データも含む。また、能力別に初級から上級まで9段階の評定が付加されている。以下では、複数の談話タグ付け体系を調査し、我々の目的に合致するタグ付け方法を議論する。

1 既存の談話タグ付け体系

談話のつながりのよさを表す上でよく利用される概念に結束性(cohesion) (Halliday and Hasan [1]; Hasan [2])がある。これは、意味的な概念関係が文法的、語彙的つながりを通して具現化されたもので、文の単位を越えたテキストにテキスト性 [一貫性] (=テキストの自然さ)を与えるものとされる。

談話タグを実際にアノテーションする試みにはいくつかある。Garside, Fligelstone and Botley [3] の談話タグ体系は、結束性の分類(指示、代用、省略、語彙的結束)に依拠し、文間のつながりを示す結束表現のアノテーションを行っているが、そのアノテーションは内的表象の手がかりにすぎないため、記述には注意が必要である。

Carletta, Dahlback, Reithinger, and Walker (eds.) [4] は、結束関係だけでなく、*I, you* といった外界照応

や、コピュラ文などにみられる指示表現、また名詞以外の先行詞などに対してもタグ付与を試み、**precision** や **recall** などの量的研究を行っているが、同様の理由で記述には注意が必要である。

結束性の記述の難しさには、以下のようなものがあげられている。①結束性の分類は列挙に頼っており、特に一般性を意図したものではない(Stoddard [5]; 山中 [6])。②内的表象における対象と指示表現が照応するのであり、「先行詞」と「照応表現」の関係は表層の置き換えにすぎない(Brown and Yule [7]; 山中 [6]; Cornish [8])。③読み手は絶え間なくテキストの伝達する情報とその情報同士の関係のダイナミックな表象を更新する(Brown and Yule [7])。

本稿では、特に③の観点に留意して、結束関係を記述する立場からではなく、円滑なコミュニケーションを図るために必要な、一貫性を支える話題の中心の観点から、学習者コーパス(The NICT JLE Corpus)の特徴付けについて考える。

2 話題の中心と一貫性

話題の中心を客観的に選び出すための手法として、本稿ではセンタリング理論の概念を援用する。センタリング理論は、談話の処理過程において連続する二発話に見られる指示表現の表層の形式選択(名詞/(ゼロ)代名詞)と一貫性について記述するモデルである (Grosz, Joshi and Weinstein [9]; 須賀[10]; Walker, Joshi and Prince (eds.)[11]; 石崎・伝 [12])。この理論の主要概念である中心は、何について(aboutness)の発話であるかをモデル化するための Grosz のいう焦点に相当する概念と考えられる (Grosz, Joshi and Weinstein [9])。焦点は、その都度同

じ表層の形式で談話に再導入する必要はなく、例えば、談話における照応の形式は受容者に対して焦点を記す役割を果たすことがある。

この談話現象の記述の道具立てとして、各発話の焦点の候補に複数の名詞を考え、最も焦点の当たっている顕在性の高い名詞句を、照応の対象として使う。具体的には、前向き中心(forward-looking centers) Cfと後ろ向き中心(backward-looking center) Cbという中心を規定している。Cfは、談話単位中の発話に示される指示対象で、Cfの中で最上位に位置する要素を優先中心 (preferred center) Cpとよぶ¹。Cbは先行発話で最も序列が高かった要素で、現在話の中心になっている要素をいう。Cbが現在の発話の中心を表すのに対して、Cpは次の発話の中心に関する予測を表し、このCbとCpの関係が一貫性の高低を決定する。

(1) 中心の遷移パターン

	Cb(Ui) = Cb(Ui-1) Or Cb(Ui-1) = 不定	Cb(Ui) ≠ Cb(Ui-1)
Cb(Ui) = Cp(Ui) Cb(Ui) ≠ Cp(Ui)	CONTINUE RETAIN	SMOOTH-SHIFT ROUGH-SHIFT

このセンターの遷移には、CONTINUE > RETAIN > SMOOTH-SHIFT > ROUGH-SHIFT という優先順位が仮定されている。

一貫性は、談話の順序(sequence)を解釈する際の受容者の推論の負担を反映すると考えられている。ある談話の最小単位である発話²において代名詞化された指示対象(複数個あるときにはその中でも文法機能の面で上位に位置するもの)が、その直後の発話においても文法機能面で上位に位置すれば、それは受容者にとって最も予測可能な談話のつなが

¹ Cfの要素はCfランキングによって順序づけられている。subject > indirect object > direct object > object of preposition

² センタリング理論には、三つの異なる単位認定の方法がある。(1)定形(finite)、不定形(infinite)を含むすべての節を発話単位とする、(2)埋め込み節以外の定型の動詞を有する節を発話単位とする(Kameyama [13]; Walker et al. (eds.) [11])、(3)いわゆる文を発話単位とみなす(Miltsakaki and Kukich [14][15])。

りであるとされる。つまり、代名詞化された主語が後続の発話の主語と同一指示の場合に最も好ましい談話のつながりを生じるため、同じ指示対象を中心とする談話の断片は労力を要せずに理解できる。従って、ある中心から別の中心にシフトするよりも、より一貫するのである。

既に、谷村ら [16]は、学習者の発話を観察し、英語学習者には、以上で説明した中心の遷移パターンではなく、学習者特有の特徴的な遷移パターンがあることを報告した。しかし、学習者の遷移パタンの詳細な特徴付けのためには、さらに、中心の記述に工夫が必要であると考えている。

以下では、The NICT JLE Corpus を用いてタグ付与を行い、Cb と Cf の記述による遷移パタンの違いから、学習者の発話の特徴付けについて検討する。

3 The NICT JLE Corpus における遷移パターン記述

The NICT JLE Corpus は、面接官と受験者の面談によるスピーキングテストが収められており、幾つかのタスクからなる(1枚の絵の描写、ロールプレイ、連続する絵を用いたストーリーテリング、フォローアップインタビュー(=タスクではない雑談)など)。以下では、独話にあたる絵の描写とストーリーテリング、対話にあたるフォローアップインタビューを取り上げ、Cb と Cf のタグ付与の結果から中心の確立と展開についてみていく(なおタグセットは資料1を参照)。まず(2)と(3)で、独話について検討する。

(2) 絵の描写の発話例<file 01095: level 5>

Utterances	Cb	Cf
A girl sitting on a chair.	no Cb	a girl
And a computer is on the desk.	no Cb	a computer
A dog is sleeping on the floor.	no Cb	a dog
A cat is sleeping on the bed.	no Cb	a cat
The door is open.	no Cb	a door
Now the time is June.	no Cb	a time
And there is a window with the curtain.	no Cb	a window
And the dust is in the dust box.	no Cb	a dust
The time is nine o'clock.	no Cb	a time
And the stereo the center of the room facing board.	no Cb	a stereo

(2)の絵の描写は、様々な事物の記述が求められるため、事物を列挙するという描写独特の遷移パターンがみられたが、Cb と Cf の記述から学習者の遷移パターンを客観的に記述することができる。

(3) ストーリーテリングの発話例<file 01163: level 3>

Utterances	Cb	Cf
It's a father and son, or uncle and nephew, two person want to eat for a restaurant.	no Cb	two person
They look for the table,	two people	they(=two people)
And they sit the table.	two people	they(=two people)
And take a order.	two people	zero(=two people)
And they eats their own favorite eat.	two people	they(=two people)
And eat and drink.	two people	zero(=two people)
<?> That's all are tea for uncle</?> ³ .	?	that
The young boy is very content.	no Cb	the young boy
Finally, they said goodbye.	no Cb	they(=two people)

(3)のストーリーテリングでは、two people が中心として確立し展開しているが、絵の描写と同様に Cb と Cf の記述から学習者の遷移パターンをみることができる。

(4)(5)(6)は、対話の例であるが、インタビューなので、中心導入の主導権は基本的に面接官にある。典型的な導入としては、Please tell me ~、Can you tell me about ~? Can you compare ~to ...? What kind of ~? How do you~?などが挙げられる。

(4) フォローアップの発話例<file 01211 level 9>

Utterance	Cb	Cf
<A>Actually it's my first time coming to this university.	no Cb	it(=campus)
<u>Can you tell me</u> more about the campus ? ⁴	no Cb	the campus
Oh really? as you can see, the campus is really small	campus	the campus
because it's in the middle of <H pn="others1">XXX03</H> ⁵ ,	campus	it(=campus)
but it's really interesting to see all the people who are in the machinery department	campus	it(=campus)
<A>Mh-hmm.		

³ <?></?> 転記者が聞き取りに自信のない箇所

⁴ <A> 面接官の発話; 受験者の発話

⁵ <H pn="others1">XXX03</H> 固有名詞 (地名)

And chemistry and everything is inside one campus .	campus	chemistry and everything
So it's really well small in campus size, but big in what we do right now here,	campus	it(=campus)
<A>Mhm.		

(4)では、受験者が、面接官がトピックとして取り上げた campus を受けて中心を確立しており、Cb と Cf の記述からその遷移パターンをみることができる。つまり、対話においても面接官の質問に対して受験者が適切に答えているかどうか客観的に分かる。

(5) フォローアップの発話例<file 01187 level 6>

Utterance	Cb	Cf
<A> What's so great about Thailand?	no Cb	what (?)
Why do you like that country so much?	no Cb	why (?)
 Because of the food . 	food (?)	food
 Thai food is very, very delicious.	food	Thai food
It's my taste.	food	it(=food)
<A>I see.		

(5)では、学習者がタイを好む理由に焦点が当たっており、学習者は面接官の質問を受けて food を中心として話題を展開している。Cfの対象は、wh-疑問詞と考えられるが、記述方法はまだ解決されていないため、遷移パターンを記述することはできない。

(6) フォローアップの発話例<file 01189 level 6>

1. So **what** did you do in the wintertime in <H pn="others1">XXX03</H>?
2. **I help my host father to shovel the snow**
3. <A>Ah-huh.
4. **And I made one snowman**.
- 5. And my host mom and my host dad is not Christian.
- 6. <A>Mhm-huh.
- 7. So we didn't celebrate any Christmas.

(6)の場合も wh-疑問詞に焦点が当たっているが、Cfの候補となる発話番号の2や4の記述方法は、まだ解決されていないため、遷移パターンを記述することはできない。これら(5)(6)については今後の

課題としたいが、この記述が可能となれば、学習者の対話における一貫性の記述が可能となる。

4 結論

本稿では、センタリング理論の概念である Cf と Cb を用い、独話や、依頼文から始まる対話において、学習者の一貫性の高さ・低さを記述する方法を紹介した。これにより、独話だけでなく、対話でも面接官の質問に対して受験者が適切に返答しているかといった点から客観的な測定ができると考えている。Cf が wh-疑問文の場合や節や文の単位に及ぶ場合における記述法については今後の課題であるが、更にデータを観察し、談話レベルの能力判定の一指標として、定式化していきたい。

参考文献

- [1] Halliday, M. A. K. and Hasan, R. 1976. *Cohesion in English*. London: Longman.
- [2] Hasan, R. 1984. Coherence and cohesive harmony. In J. Flood (ed.), *Understanding Reading Comprehension: Cognition, Language and the Structure of Prose*, pp. 184-219. Newark, Delaware: International Reading Association.
- [3] Garside, R., Leech, G. and McEnery, T. (eds.) 1997. *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Addison-Wesley London.
- [4] Carletta, J., Dahlback, N., Reithinger, N. and Walker, M. A. (eds.) 1997. *Standards for Dialogue Coding in Natural Language Processing*. Dagstuhl – Seminar – Report 167.
- [5] Stoddard, S. 1991. *Text and Texture: Patterns of Cohesion*. Norwood, N. J.: Ablex Publishing Corporation.
- [6] 山中桂一 1998. 『日本語のかたち』 東京大学出版
- [7] Brown, G. and Yule, G. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.
- [8] Cornish, F. 1999. *Anaphora, Discourse, and Understanding. Evidence from English and French*.

Oxford: Oxford University Press.

- [9] Grosz, B. J., Joshi, A. K. and Weinstein, S. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21 (2), 203-225.
- [10] 須賀あゆみ 1997. 「センタリングと指示表現の談話機能—代名詞 it と指示代名詞 this/that を比較して—」 『奈良女子大学文学部研究年報』 41, 59-72.
- [11] Walker, M. A., Joshi, A. K. and Prince, E. F. 1998. *Centering Theory in Discourse*. Oxford: Clarendon Press.
- [12] 石崎雅人・伝康晴 2001. 『言語と計算—3 談話と対話』 東京大学出版会
- [13] Kameyama, M. 1998. Intra-sentential centering: A case study. In M. A. Walker, A. K. Joshi, and E. F. Prince, (eds.), *Centering Theory in Discourse*, pp. 89-112. Oxford. Clarendon Press.
- [14] Miltakaki, E. and Kukich, K. 2000a. The role of centering theory's ROUGH-SHIFT in the teaching and evaluation of writing skills. In *Proceedings of ACL 2000*.
- [15] Miltakaki, E. and Kukich, K. 2000b. Automated evaluation of coherence in student essays. In *Proceedings of the Workshop on Language Resources and Tools in Educational Applications, LREC 2000*.
- [16] 谷村緑・竹内和広・井佐原均 2005. 「英語教育のための照応詞使用についての調査」 言語処理学会第 11 回年次大会発表論文集, pp1111-1114.

資料1 タグセットの例

タグ	説明
<Cb></Cb>	後ろ向き中心(Cb)
<Cb ref=0></Cb>	Cb がない
<Cb ref=Taro>he</Cb>	Cb が代名詞で具現
<Cb ref=Taro>the boy</Cb>	Cb が名詞で具現
<Cb ref=-1>it</Cb>	Cb が不明
<Cf></Cf>	前向き中心(Cf)
<Cf rank=1>he</Cf>	Cf 候補の最上位に位置し代名詞で具現
<Cf rank=1>Taro</Cf>	Cf 候補の最上位に位置し名詞で具現
<Cf rank=-1>it</Cf>	Cf が不明で代名詞で具現。