

日本語の省略と母語話者の直感： センタリング理論による検証

竹井 光子
広島修道大学
法学部
takeim@shudo-u.ac.jp

藤原 美保
ウィラメット大学
日本語・中国語学科
mfujiwar@willamette.edu

1 はじめに

日本語談話では文の要素をしばしば省略する。冗長さのために不自然にならずかつ省略によって曖昧さが生じたり多大な推論量を課したりすることのないようバランスをとりながら、文脈などから推測が可能であると判断できる場合に省略を行う。母語話者はこの判断を即時的に行いながら、自然な談話生成を行うことができる。この判断能力は日本語母語話者に備わっている言語的直感とも言えるだろう。しかしながら、この自動的、無意識的と思える処理にも、何らかの判断材料となるメカニズムが存在しているはずである。

竹井ら(2006)では、省略に関する直感を説明するメカニズムとしてセンタリング理論 (Grosz et al. 1995) に注目し、コーパス分析と母語話者調査の結果から理論の主張の妥当性の検証を行なったが、本稿ではさらに発展させた母語話者調査によってその妥当性の再検証を行う。そして、調査結果の分析から、母語話者の省略に関する傾向が、センタリング理論が定式化する談話の一貫性によって説明できることを報告し、センタリング・メカニズムの日本語教育における活用の可能性について論じる。

2 日本語教育と省略

日本語教育の場で、教師が学習者の生成する談話（書き言葉、話し言葉）に接するとき、不自然であると感じたり、解釈に困難を感じたりすることがある。それには、省略過少による冗長さ、省略過多による曖昧さが影響している場合が少なくない。次の(1)と(2)は、日本語学習者の作文である¹。

- (1) 農民がいつも盗賊に米やお金を盗まれて、農民が盗賊を倒すために七人の侍をやといました。
- (2) 昔々ある村にある子どもがいた。狼が来たと叫んだ。村人が来たが、狼はなかった。何度も繰り返して、村人が怒った。来なくなった。

¹ カナは漢字表記に変えてある。

これらの作文を添削するとする。省略に関して違和感を覚える箇所はないだろうか。例えば、(1)では2つ目の「農民が」は必要だろうか。(2)では、誰が「繰り返した」のか、誰が「来なくなった」のかは明確だろうか。さらに、省略すべきか否かの判断がきわめて容易な場合もあれば、非常に迷う場合もある。教師によって判断が分かれる場合もある。

これらの判断の基準は何であるのか。判断の差はどうして生じるのか。残念ながら、この基準を体系的に説明している日本語指導書は筆者らの知る限り存在しない²。

3 センタリング理論

3.1 理論の概要

センタリング理論は、談話の一貫性を構成する一側面のモデル化を試みた理論である。一貫性へのアプローチ方法としては、「接続関係」または「参照関係」に注目した2つの流れがあるが、センタリング理論は後者に当たる。すなわち、談話に登場するDISCOURSE ENTITY と呼ばれる名詞句が発話ごとにどのように移り変わるかに注目したアプローチである。

具体的には、前の発話（節）から引き継がれたENTITYのうち、最も上位のものがCENTER (CB)となる。CENTER 候補となるENTITYの順位付けは、それが文法上果たす役割を基準としている³。この「話題の焦点」であるCENTER (CB)は発話ごとに更新され、その「更新のされ方」をTRANSITIONとして規定している。TRANSITIONにはCONTINUE (CON), RETAIN (RET), SHIFT (SHIFT)の3つがあり、その連続の仕方によって、談話の一貫性の度合、談話の解釈に必要な推論量が異なるとしている。

Groszらによると、CON-CONの並びはRET-RET, SHIFT-SHIFTよりも好まれる。また、RETAINは後続のCENTERの変化を予測させる状態であり、RET-SHIFT

² 最も一般的な記述は、「主語が連続する場合には省略できる」であろう。

³ 本稿では、主格[が] > 主格以外(対格[を], 与格[に], 連体格[の], 場所格[で]など)の順とする。本稿で採用した定義や分析基準の詳細については、Yamura-Takei (2005)にある。

は望ましい連続である。さらに、CENTERの移行が完了した状態から新たなCENTERが連続することは自然であり、SHIFT-CONも妥当な流れである。したがって、CON-CONは一貫性のきわめて高い連続、CON-RET-SHIFT-CONは理想的なCENTER移行の流れであると言える。一方で、CON-SHIFTは予測なしの突然の変化、RET-CONは、変化予測を裏切る流れであり、解釈に必要となる推論量が増すと考えられる。

本稿では、前回の母語話者調査と同様、理論の原典 (Grosz et al. 1983, 1986, 1995) に忠実に、TRANSITIONの連続と人間が知覚する一貫性との関係を検証することにする。

3.2 分析例

前節で、Groszらが主張するTRANSITIONの連続と一貫性の度合との関係について触れたが、ここでは、談話例を挙げて具体的に説明する。

まず、談話例(3)を見て欲しい。ここでは、各発話のCENTER(CB)に当たる要素をすべて省略してある。

- (3) あの家トムに住んでいます。意地悪なネコです。いつもジェリーを追いかけています。ジェリーは同じ家に住んでいます。頭の良いネズミです。飼い犬のスパイクと仲良しです。トムとは仲良しではありません。よくジェリーをからかいます。ジェリーはときどき仕返しをします。スパイクはよくおどかします。たいてい最後には負けてしまいます。

アニメ「トムとジェリー」に馴染みがあれば、背景知識からこの談話が伝えようとしていることを推測するのは不可能ではないだろう。しかし、自然な流れとを感じる部分もあれば、「誰が」「誰を」等が不明瞭で流れが円滑でないと感じる部分もあるのではないだろうか。この感覚的な差を説明してくれるのがセンタリング理論であると考え、(4)に分析例を示す。

各発話中、CENTER(CB)となっている要素を下線で示している。また、本稿ではGroszらの主張にしたがって、単独のTRANSITIONではなく、その連続に注目する。ここでは便宜上、各発話に2つのTRANSITIONの連続を示すラベルを付すことにする。下記(4)の談話例中、このラベルを括弧内太字で示し、以後これを「連続パターン」と呼ぶことにする。例えば、(CON-SHIFT)ラベルの連続パターンはCONTINUEの発話に続くSHIFTの発話であることを、(RET-SHIFT)はRETAINの発話に続くSHIFTの発話であることを示す。TRANSITIONの連続を見るのは、同じSHIFTの発話であっても、何の発話の後続であるかが一貫性の度合を測る上で重要だからである。

- (4)
- a. あの家トムに住んでいます。
CF: Tom, house
 - b. トムは意地悪なネコです。
CB: Tom CF: Tom, cat CON (NULL-CON)
 - c. トムはいつもジェリーを追いかけています。
CB: Tom CF: Tom CON (CON-CON)
 - d. ジェリーはトムと同じ家に住んでいます。
CB: Tom CF: Jerry, Tom RET (CON-RET)
 - e. ジェリーは、頭の良いネズミです。
CB: Jerry CF: Jerry, mouse SHIFT (RET-SHIFT)
 - f. ジェリーは飼い犬のスパイクと仲良しです。
CB: Jerry CF: Jerry, Spike CON (SHIFT-CON)
 - g. スパイクはトムとは仲良しではありません。
CB: Spike CF: Spike, Tom SHIFT (CON-SHIFT)
 - h. トムはよくジェリーをからかいます。
CB: Tom CF: Tom, Jerry SHIFT (SHIFT-SHIFT)
 - i. ジェリーはときどきトムに仕返しをします。
CB: Tom CF: Jerry, Tom RET (SHIFT-RET)
 - j. スパイクはよくトムをおどかします。
CB: Tom CF: Spike, Tom RET (RET-RET)
 - k. トムはたいてい最後には負けてしまいます。
CB: Tom CF: Tom CON (RET-CON)

この談話は、(4a-b)で「トム」について話し始め(CON-CON)、(4c-e)で徐々に「ジェリー」にCENTERをシフトさせた後に(CON-RET-SHIFT)、(4f)でも「ジェリー」の話題を継続している(SHIFT-CON)。その後、(4g)では、CENTERが唐突に「スパイク」に移る(CON-SHIFT)が、(4g-h)では再びCENTERが「トム」へと移行する(SHIFT-SHIFT)。次の(4i-j)では、「トム」をCENTERとして保ちながら、「ジェリー」、「スパイク」と新たなENTITYを上位の主語の位置に導入し続けている(RET-RET)。CENTERの変化を予測させるが、その後に「ジェリー」や「スパイク」の言及はなく、予測に反して(4k)では「トム」へとCENTERが戻っている(RET-CON)。

すなわち、この談話を構成する各発話は、TRANSITIONの「連続パターン」の観点(一貫性の度合の観点)から見ると、すべて異なる状況にあることがわかる。これが省略を自然と感じるか否かに影響を与えていると仮定し、その検証を行なうのが母語話者調査の目的である。

ここで言う「省略」とは、一般的に「ゼロ代名詞」と呼ばれるもので、文脈などから推測が可能であると判断できる場合に省略という形をとる項(連用項および連体項⁴)を指す。

⁴ 連用項は「ガ格」「ヲ格」「ニ格」、連体項は「ノ格」である。

4 母語話者調査

4.1 調査方法

日本語母語話者（大学生）87名に対して、アンケート調査を行った。調査用の談話サンプルは、前述の9つのTRANSITIONの連続パターンをそれぞれ含む9サンプルを3セット、計27サンプルを用意した。

サンプル作成の際には、知識、言語要素など一貫性の知覚や推論量に影響を与えると考えられる要因を出来るだけ排除するため、一般的に馴染みのある内容で⁵、自然さを損なわない程度に平易な表現となるよう留意した。サンプルごとにCENTER(CB)に当たる名詞句を明示したものと省略(ゼロ代名詞化)したものとを並列配置し、どちらがより自然であると感じるかの判断を求めた。ここでは、前回の調査と同様、省略を認める割合が高いほど、その談話状況(TRANSITIONの連続パターン)の一貫性が高い(よって曖昧さの発生などの危険が生じない)と母語話者が判断したと仮定する。

4.2 調査結果

図1に、各TRANSITIONの連続パターンにおいてCENTER(CB)の省略をより自然であると判断した率を降順に示す。数値は、各パターン3セットの平均値である。

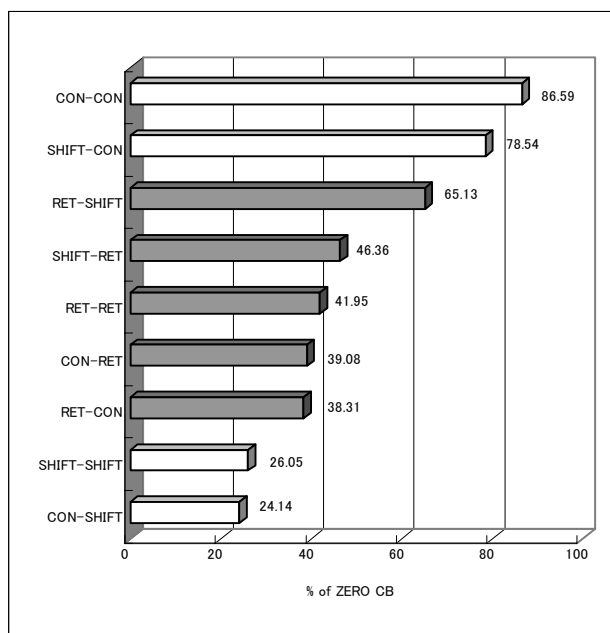


図1：母語話者調査結果

⁵ 「トムとジェリー」, 「ライオンとシマウマ」, 「石, はさみ, 紙」を題材とした3セットである。

まず、省略を自然とする率が、CON-CONの86.59%からCON-SHIFTの24.14%まで段階的な序列11になっていることが分かる。この率が一貫性の度合を示す目安であると仮定すると、一貫性とはあるかないかの二択の概念ではなく人が知覚する段階的な度合の問題であることを実証していることになる。

次に、「連続パターン」ごとの数値を見てみよう。最も一貫性が高いとされるCON-CONでは、86.59%の被験者が省略を認めている。また、CENTER移行後の連続であるSHIFT-CONにおいても78.54%という比較的高い割合で省略を認めている。一方で、唐突な流れであるCON-SHIFTでは24.14%、頻繁な話題の変化を示すSHIFT-SHIFTでは26.05%と、明らかに省略を許容する割合が低くなっており、センタリング理論の主張と一致している。これら4つのパターン(グラフの白色部分)では、省略が安全か危険かの選択に比較的強い傾向が出ていることがわかる。他の5つのパターン(グラフの灰色部分)は、約65%から38%のやや中間的な結果となっている⁶。この5つのパターンには、すべてCENTERの移行を予測させる状態にあたるRETAINが関わっているのは興味深い。

4.3 考察

調査結果をもとに、9つの連続パターンを一貫性の度合によってHIGH, MEDIUM, LOWの3つのグループに分類した(図2)。HIGHグループは、一貫性が極めて高く解釈に必要となる推論量も少なくすむため省略しても曖昧性が生じないという点で安全な談話状況である。LOWグループは、一貫性が極めて低く、必要とする推論量も多くを必要とするため省略は曖昧性を生じさせる可能性が高い点で避けられる傾向にある。MEDIUMグループは、一貫性、必要とする推論量ともに中間的で、省略するか否かの判断は、知識や他の言語要因の影響を受けやすい⁷。

HIGH	MEDIUM	LOW
CON-CON SHIFT-CON	RET-SHIFT SHIFT-RET RET-RET RET-CON CON-RET	SHIFT-SHIFT CON-SHIFT
推論量・少 ----- 多		
省略・安全 ----- 危険		

図2：TRANSITIONの連続パターンの分類

⁶ この結果は、「連続パターン」ごとにCENTERの省略が起きている割合を調査したコーパス分析結果(Yamura-Takei 2005)や前回の母語話者調査の結果(竹井ら2006)と、ほぼ一致している。

⁷ Yamura-Takei(2005)のコーパス分析の結果から、影響を与える言語要因としては、共感表現、選択制限、平行構造、主題化、倒置、接続表現などの有無が考えられる。

前述の談話例 (3) を、この分類にしたがって書き直したものを(5) に示す。HIGH グループの CENTER は省略したまま、LOW グループの CENTER は省略せず下線で強調してある。MEDIUM グループの CENTER は括弧内に置き場合によっては省略も可能であることを示している。

(5)

あの家にトムが住んでいます。トムは意地悪なネコです。いつもジェリーを追いかけています。ジェリーは(トムと)同じ家に住んでいます。(ジェリーは)頭の良いネズミです。飼い犬のスパイクと仲良しです。スパイクはトムとは仲良しではありません。トムはよくジェリーをからかいます。ジェリーはときどき(トムに)仕返しをします。スパイクはよく(トムを)おどかします。(トムは)たいてい最後には負けます。

すべての CENTER を省略していた談話例 (3) よりも不自然さや不明瞭さが減少しただろうか。LOW グループの CENTER の省略、HIGH グループの CENTER の明示が妥当であると感じる一方、MEDIUM グループの CENTER (括弧内) の省略の可否については、判断に迷ったり「どちらでもよい」と感じたりするのではないだろうか。例えば、誰が「仕返しする」かは、「からかう - 仕返しする」の意味関係に関する知識を判断材料に取り入れるかどうかで知覚する一貫性の度合が変わるであろう。また、誰を「スパイクがおどかさ」のか、「最後に負ける」のは誰なのかの判断は、「トムとジェリー」のストーリー展開に関する知識が影響するかもしれない。

5 まとめ

センタリング理論の主張の妥当性を母語話者による直感的判断調査によって検証した。最も一貫性が高いとされる CON-CON, 自然な流れである SHIFT-CON から、唐突な流れである RET-CON, SHIFT-SHIFT, CON-SHIFT まで、一貫性の度合により段階的な序列ができており、Grosz らの主張とも一致する結果となった。

母語話者は、センタリングが説明する一貫性の度合の差を直感的に感じとって省略に関する判断を下しているのではないだろうか。少なくとも判断材料の一つとして重要な役割を果たしていると言っても過言ではないだろう。

さらに、調査結果にもとづく HIGH, MEDIUM, LOW のグループ分けでは、「省略すべき」「省略するかどうかは談話状況から判断すべき」「省略すべきでない」のガイドラインを示すことが出来た。段階的な一貫性の度合を便宜上 3 つのグループに分類したわけだが、直感的判断の裏づけ、省略指導に

おける説明材料として利用することが出来るのではないだろうか。

センタリング理論は計算機的談話理論であり、その柔軟さゆえに様々な改変、拡張、応用が行われてきた。初期に BFP アルゴリズム(Brennan et al, 1987) が提案されたため、照応解析のための理論であるという印象が強い。また、テキスト生成の理論としても活用されている(Karamanis, 2003)。

省略についての直感的判断を説明することは難しい。この自然言語処理モデルが、判断の裏にある認知的メカニズムに理論的説明を加えることで、言語教育の場でも活用されることを願っている。「どこで何を省略すべきか、省略しても安全なのか」についての直感的認知過程に、実証的データに基づく理論的説明を与えることで、省略を指導する日本語教員の一助になればと考える。

参考文献

- Brennan, Suzan E., Marilyn Walker Friedman, and Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics, Stanford, CA*, 155-162.
- Grosz, Barbara, Aravind Joshi, and Scott Weinstein. 1983. Providing a unified account of noun phrase in discourse. In *Proceedings of the 21th Annual Meeting of the Association for Computational Linguistics, Cambridge, MA*, 44-50.
- Grosz, Barbara, Aravind Joshi, and Scott Weinstein. 1986. Towards a computational theory of discourse interpretation. Unpublished manuscript.
- Grosz, Barbara, Aravind Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21/2, 203-225.
- Kameyama, Megumi. 1985. *Zero Anaphora: The Case of Japanese*. PhD dissertation, Stanford University.
- Karamanis, Nikiforos. (2003). *Entity Coherence for Descriptive Text Structuring*. PhD dissertation, University of Edinburgh.
- Yamura-Takei, Mitsuko. 2005. *Theoretical, Technological and Pedagogical Approaches to Zero Arguments in Japanese Discourse: Making the Invisible Visible*. Doctoral thesis, Hiroshima City University.
- 竹井光子・藤原美保・相沢輝昭. 2006. センタリング理論とゼロ代名詞：日本語コーパス分析と母語話者調査の結果から. 言語処理学会第 12 回年次大会発表論文集, 横浜.