

自然会話における名詞句省略とその復元

石田 邦子 藤井 聖子

東京大学大学院総合文化研究科言語情報科学専攻

1 はじめに

本研究では、日本語の自然会話を分析対象として名詞句省略の復元メカニズムを検証した。データは Fujii コーパス [1] から、大学生の女性同士の会話 (J5F) と男性同士の会話 (J6M) を 5 分ずつ計 10 分を抜粋した。

まず、前方照応型および会話参加者の省略に関して、構文の人称選択制限などの言語学的知識を補助的に利用しつつ顕著性と距離をパラメータとしているアルゴリズム (Nariyama [6]) を拡張し、その拡張版アルゴリズムを用いて復元を行った。次に、そのアルゴリズムでは解決できなかった省略について、実際の復元の有様を分析した。

2 アルゴリズム拡張版による復元

2.1 Nariyama アルゴリズム

Nariyama [6] のアルゴリズム (以下 Nariyama アルゴリズム) は、前に出て来た明示的な項を顕著性に基づく順位でリストの形式で保持し、その上位から省略の復元候補としていくという、基本的には前方照応型の省略の解決を目指したアルゴリズムである。しかし、共感度や敬語、心的述語といった人称制限に関わる言語学的な知識も確認の段階で用い、更にそれでも復元ができない場合は世界知識を考慮するといった手順も組み込まれており、総合的な省略の復元を目指している。

明示的な名詞句は 1 文を処理するごとに顕著性のヒエラルキー¹に基づいて、有生・無生 (animate/inanimate) という情報とともに、同じ文法役割のものが出てくるまでリストに保持される。リストの第 1 位にある項目はその文、あるいはそれ以降の文に出てくる省略を復元するための第一候補指示対象 (T1) となり、文に省略があると判断されると、その復元に用いられる。複数

¹ 主題化された主語 (TOP.SB) > 主題化された目的語 (TOP.OB) > 主語 (SB) > 間接目的語 (IO) > 直接目的語 (OB) > その他 (OBL (斜格) など)

の省略がある際には、T1、T2、T3 の順で、SB 省略、IO 省略、OB 省略の復元に使われる。

ただし、以下のような構文はリストによる復元の前に検討する。

1. 第 1 番目の文であって、リストに「は」標示の指示対象が入っていない場合は、一人称か総称を当てはめる。
2. もしその文が、内部に独自の主格も主題もとれない従属節²を伴う複文で、主節の主語が「が」で標示されている場合、従属節で省略されている主語は、主節の主語を指示対象とする。
3. 関係詞節で欠けている項は関係詞の主要部と同じ指示対象である。

次に、この復元の結果を言語学的知識に基づいて確認 (verify) する。具体的には、

1. 有生・無生を確認する。
2. 接続助詞のタイプを確認する³。
3. 敬語を確認する。
4. 「思う」「悲しい」などの心的述語を確認する⁴。

この段階を経ても復元ができない場合は、リストの下位の項目を用いてもう一度指示対象の同定を試みる。それでも正しい結果が得られない場合は、以前のリストで T1 であったものを用いる。更にそれでも復元できない場合は、世界知識を使うことになる。

2.2 アルゴリズム拡張版

本研究は Nariyama アルゴリズムが書き言葉テキスト上で良い結果を出していることを踏まえ、分析対象

² 南 [5] の分類で「A 類の従属句」とされるもの。「～ながら」「～つつ」など

³ 南 [5] の分類で B 類とされる、従属節内に独自の主題はとれないが主格はとれるタイプの接続助詞 (「～たら」「～ので」など) の前後では主語が変わる傾向がある。逆に、A 類の従属節の主語は主節のものと同じである。ただし、これはあくまで傾向であって、確実に主語を予測するものではない。

⁴ 内的状態を表す述語は肯定文では一人称、疑問文の場合は二人称になるという人称の選択制限がある。

を自然会話に移した。それと同時に、いくつかの部分で Nariyama アルゴリズムを拡張した。主だった拡張規則を以下に示す。

1. Kameyama [4] や Walker et al. [7] で、直前の省略が次の省略になりやすいことがモデルに組み込まれていたことを考慮し、省略も基本的には主題と同等の顕著性をもつとして⁵指示対象の同定が終わったあとリストに組み込み、T1 として次の省略の復元第一候補とする。
2. 1. の規則を踏まえ、顕著性のヒエラルキーを次のようにした。「/」は同順位を表す。
 1. 主題 (TOP⁶ > TOP.SB / SBelli⁷ > TOP.IO / IOelli > TOP.OB / OBelli > TOP.OBL)
 2. 主語 (SB)
 3. 間接目的語 (IO)
 4. 直接目的語 (OB)
 5. 斜格 (OBL) / 属格 (GEN)
3. Nariyama アルゴリズムでは以前の T1 は、現在のリストの要素全てで復元ができなかった場合に用いる補助的な手段であったが、拡張版では以前の主題を T1 に次ぐ順位の T1' とする。
4. リストによる復元の段階で有生・無生も考慮する。
5. リストに有生の主題がなく、述語が有生の主語を要求する場合、話し手の省略であると考え⁸。
6. 話し手・聞き手はリストに入れない⁹。
7. 明示的な名詞句はその文を含み連続で 3 文までしか保持しない¹⁰。

2.3 結果と考察

J5F では全 103 文について、J6M は全 104 文についてアルゴリズム拡張版を実行した。表 1 は省略の指示

⁵主節と同じ指示対象をもつ従属節中の省略についてはその限りではない。

⁶述語の項にはなっていない主題

⁷elli=ellipsis (省略)

⁸人称の選択制限から一人称が不可能である場合は聞き手の省略と考える。

⁹ただし、明示的に現れた場合はその文の処理時のみリストに入れる。

話し手・聞き手の復元はほとんど全てが人称の選択制限または、5. の規則で復元が可能であった一方、話し手・聞き手がリストに入っていることが正しい省略の復元の妨害になっている場合があったためこの規則を設けた。(石田 [3])

¹⁰ただし、あいづちや聞き返し、言い直し、言い終わっていないことが明らかな発話などは文としてカウントしない。

出現頻度の低い文法役割の要素は 20 文もリストに保持されることもあったが、実際の復元に使われるのはせいぜい 2 文後までであることが分かったためこの規則を設けた。ただし、3 文保持されている要素による正しい復元の妨げ (例えば、SB2 を選びたいが、T1 に TOP.SB3 がある場合など) もあり、判断が難しい (石田 [3])。

対象の内訳を示したものである。前方照応型の省略の場合はその先行詞の文法役割、それ以外の場合については省略の指示対象 (話し手・聞き手・その他) が表されている。「復元回数」はそのうちアルゴリズムでやり直しをせずに復元できたものの回数である。なお、文法役割の末尾の数字は何文保持しているかを表している。1 は同文、2 は直前の文、3 はその前の文、「*」はそれ以上前に出て来たものである。回数が (0) となっているものはそのような文法役割の要素がそもそもデータに現れていないことを示している。

表 1 を見ると、前方照応型では直前の文に登場した SBelli を先行詞とする省略がもっとも多い。拡張版規則 1. により、そのほとんどがアルゴリズムによって復元できる。話し手の省略は SBelli2 と並んで多いが、これは拡張版規則 5. と人称の選択制限によってほとんどが復元できた。聞き手についても同様である。全体では J5F で 64.7%、J6M では 70.7%の省略が拡張版によって復元できた。前方照応型の省略に限定すると、J5F で 70.1%、J6M では 76.9%の復元率となった。

拡張版で新たに設けた T1' という順位を用いた復元は 5 個あった。そのうちのひとつでは 1 回のやり直しを経て選ばれている。ここは Nariyama アルゴリズムでは 2 回のやり直しが必要であった箇所である。逆にこの T1' という順位によって正しい復元が妨げられたという事例はなかった。

3 アルゴリズムによらない復元

表 1 で「その他」となっているもの、あるいは「*」がついた 3 文以上前に出て来た要素を指示対象としている省略は、直近の前方照応の解決を目指したアルゴリズム拡張版では解決できない。この節では、アルゴリズムによらない省略の復元を 7 つの場合に大別した。なお、() 内の数字は該当する省略の回数である。

1. 「なぜ」「なに」「どこ」など疑問詞を使った疑問文の後で「それ=質問の答え」が省略されている場合 (6)。これらの質問に対しては「それは～だ」の形式で答えたり、「それは教えない」などと対応するが、その「それ」が省略される場合である。この場合、リストでは復元できない。ただし、「それ=質問の答え」を省略するタイプは疑問詞を用いた疑問文のあとに出現するものであり、そのような条件付きでアルゴリズムに組み込むことは可能ではある。

2. 前に出て来たものと関係はあるが、全く同じ指示

先行詞の 文法役割	J5F		J6M		先行詞の 文法役割	J5F		J6M	
	個数	復元個数	個数	復元個数		個数	復元個数	個数	復元個数
SBelli2	22	19	27	26	SB2	3	1	2	1
SBelli3	1	1	3	3	SB*	1	0	2	0
SBelli*	3	0	8	0	IO2	2	2	0	0
IOelli2	11	10	3	3	OB2	5	2	1	1
IOelli*	3	0	2	0	OB*	0	0	1	0
OBelli2	2	2	7	7	OBL2	1	0	0	0
OBelli*	1	0	0	0	OBL*	0	0	1	0
TOP2	2	1	0	0	GEN1	1	1	0	0
TOP.SB1	0	0	1	1	GEN2	0	0	1	0
TOP.SB2	8	8	8	8	GEN*	0	0	1	0
TOP.IO	(0)	(0)	(0)	(0)	話し手	27	25	20	20
TOP.OB	0	0	(0)	(0)	聞き手	5	5	7	7
TOP.OBL2	(0)	(0)	1	1	その他	20	0	18	0
SB1	1	0	9	9	合計	119	77	123	87

表 1: 省略の指示対象の内訳

対象ではないものが省略されている場合 (11)。これは二つのタイプに分けられる。

第一のタイプは、前文、あるいはそれまでの内容を漠然と受けるものである。アルゴリズム拡張版は引用節などの名詞節も OB としてリストに入れているので、省略が引用節の内容を指す場合は復元が可能である。しかし、それはまれな場合であって、実際には省略が前文を丸ごと受けていたり、特定の文を指示対象として選べないということが多い。(1) 中 [S39] の ϕ_{SB1} は直前 [S38] の「話が弾んでそうだった」ことを指しているとも考えられるし、それを含んだ「楽しそうだった」こと全体を指しているとも考えられる。いずれにしてもアルゴリズムでは復元できない。

(1) [J6M]

[S36] Y: 楽しそうだった。

[S37] O: $\phi_{SB(女の人)}$ 誰だろうな。

[S38] Y: 弾んでそうだった話が。

[S39] Y: あ ϕ_{SB1} いいなと $\phi_{SB2(Y)}$ 思ったもん。

節を項にとる「～(が)いい」「～(が)楽しい」「～(と)思う」といった述語の項が省略となっている場合に、前文全体を省略の復元候補にする、という手続きを追加すれば、多少は解決が望める。しかしそれでも、直前の文だけではなく、前の内容を漠然と省略で受けている場合にはリストでの復元は不可能である。

第二のタイプとして、前に出て来た要素と同一の意味フレームにあって容易に活性化されうるような指示対象が省略されていることがある。例えば、直前に「公演」という語が出てくると、この話し手が所属している劇団に関するフレームが活性化され、直後でそれまで出て来ていない「劇団」という語が省略されても、聞き手は理解が可能であった。

3. 話し手を含む一人称複数省略されている場合 (1)。これは話し手が突然省略の形で登場することの延長にあると考えられる。ただし、復元のためには話し手以外のメンバーが直前で言及されているか、推測が可能であることが必要であると考えられる。(2) では、[S43] で Y と O の共通知識から「女の子」が省略されている (以下 5. 参照)。それを受けて続く [S44-45] では話し手 (Y) を含む複数省略されていると聞き手は理解ができる。

(2) [J6M]

[S42] Y: 去年ちょっとね楽しかったんだけどな最初俺。

O: ああそっかそっか。

[S43] Y: うん。ちゃんと $\phi_{SB(女の子)}$ いたから。

[S44] Y: $\phi_{SB(Yと女の子たち)}$ 三対三だった。

[S45] Y: $\phi_{SB1(Yと女の子たち)}$ 仲いいねとかよく ϕ_{SB2} 言われてたんだけどね。

4. ジェスチャーによる直示の場合 (1)。データ中では、「ほら」と言いながら話し手が自分のほおを軽く叩いて自分を指していた。その動作により、言語的に話し手が省略されていても理解が可能である。

5. 話者間の共通知識や一般的事項など、文脈外の知識によって省略の理解が可能になる場合 (9)。話者同士が既に共有している知識を前提として省略が使われる場合、単に共有しているだけでは聞き手は話し手の意図どおりの正しい復元をするとは限らない。発話時点で活性化していることが省略の理解のために必要である。先の (2)[S43] では突然「女の子」が省略されているが、[S42] のあとの聞き手 (O) のあいづちは、そ

れを理解するだけの知識が活性化していたことを伺わせる。

また、一般的な知識から省略が復元できる場合もある。例えば、特定の人ではなく、一般の人が省略されていると解釈するのが妥当な事例がいくつかあった。他にも、話者間に限定されない広く一般的な知識から当然の事項として省略が起きているような事例もあった。

6. 明示されている述語が手がかりになって省略の復元が可能になる場合(13)。これには二つのタイプがある。

まず第一のタイプとして、動詞の意味特性によって指示対象が選ばれることがある。これは、5. の一般知識や共通知識と組み合わせたり、正しい復元をもたらす。(3)は(2)に続く抜粋である。

(3) [J6M]

[S46] O: あっという間に、 ϕ_{SB} (女の子たち) 消えて
行きましたよね。

[S46] の ϕ_{SB} は [S45] でリストに入った SBelli である「Y と女の子たち」ではなく、「女の子たち」だけを考えていると考えられる。それは、「消えて行く」という述語から「Y」は含まれないと判断できるからである。

第二のタイプは、前と同じ述語や似た述語を用いることで、以前の発話部分との結束性が生じ、省略がそのときと同一の指示対象をもつと理解できる場合である。

(4) [J5F]

[S78] T: うちのお母さん[SB] がさ。後楽園遊園地
で。M(芸能人) 見たんだって。

S: あうそ。

[S79] T: うん。 ϕ_{SB} (芸能人 M) 結構もう知られてる
ときで。あでももう人気なかったかな。

S: あでも人気なくてもね。一般に。

[S80] T: あの人[TOP.SB] 出たじゃん結構?
ドラマ一人で。

[S81] S: うん。 ϕ_{SB} (芸能人 M) ドラマとか出た。

[S82] T: それで $\phi_{SB(T)}$ の母親) ϕ_{IO} 会ったんだって。

S: うん。

[S83] T: で ϕ_{SB} 女の子と歩いてたのなにか知らないけど。

(4) の [S82] を処理する段階では前文 [S81] で SBelli であった「芸能人 M」が ϕ_{SB} の指示対象であると予測されるが、実際には「T の母親」が指示対象である。これは [S78] の「(芸能人 M)を見た」と似た「会った」という述語の使用によって復元が可能になる(また以下 7. で見るように談話構造も関わっている)。

7. 一般化に対して具体化、本筋に対して補足など、談話の従属部分を飛び越え、以前の主要部分に出て来た指示対象を指して省略が用いられる場合(17)。正しい復元のためには話し手だけでなく、聞き手も談話構造(例えば、Grosz et al. [2])を把握していることが必要である。先の(4)では[S79-S81]で芸能人 M にあった時期に関する補足説明のための従属的階層に入った談話が[S82]冒頭「それで」を手がかりとして、Tの母親が M に会ったエピソードという主要部分に戻ったことが聞き手にも了解される。このような談話構造の把握も述語の意味特性と合わせて[S82]の ϕ_{SB} の指示対象同様に役立っていると考えられる。

4 おわりに

本研究では、Nariyama アルゴリズム拡張版が自然会話において直近の前方照応型の省略を復元するのに有効であることを示した。さらに、アルゴリズムでは解決できない省略についても考察し、その7つのパターンを明らかにした。

参考文献

- [1] Seiko Fujii. *The spoken Japanese corpus: Dyadic casual conversation by young adults*. University of Illinois at Urbana-Champaign, 1995.
- [2] Barbara J. Grosz and Candace L. Sidner. Attention, intention, and the structure of discourse. *Computational Linguistics*, Vol. 12, No. 3, pp. 175–204, 1986.
- [3] 石田邦子. 日本語会話における名詞句省略の復元メカニズム. 修士学位論文, 東京大学, 2005.
- [4] Megumi Kameyama. *Zero anaphora: the case of Japanese*. Ph.D. dissertation, Stanford University, 1985. [Published by UMI Ann Arbor].
- [5] 南不二男. 現代日本語文法の輪郭. 大修館書店, 1993.
- [6] Shigeko Nariyama. *Ellipsis and reference tracking in Japanese*. John Benjamins, Amsterdam, 2003.
- [7] Marilyn Walker, Masayo Iida, and Sharon Cote. Japanese discourse and the process of centering. *Computational Linguistics*, Vol. 20, No. 2, pp. 193–232, 1994.