

# 話者の認識を利用した主語の人称推定モデル

小比田 涼介<sup>†</sup>

宮本エジソン正<sup>‡</sup>

筑波大学 人文社会科学研究所

ryosuke.kohita@gmail.com<sup>†</sup> ; miyamoto@alum.mit.edu<sup>‡</sup>

## 1 はじめに

日本語にはヨーロッパ言語に見られるような動詞の屈折による主語の束縛といった文法現象がなく、主語の推定に役立つ語彙的な情報が少ないと考えられている [4, 7]。しかし、実際には述部の要素によって人称制限が生じる場合がある (1)。

- (1) a. (私 / \*彼) は寒いです。  
b. (\*私 / 彼) は寒いみたいです。

このような現象は言語学において詳細に記述されてきたが、その詳細さゆえに機械処理に耐え難く、自然言語処理分野で利用されることは少なくなっている。本論では認識論 (epistemology) に根ざした言語理論によって、日本語での人称制限を機械処理可能な状態にまとめなおす。これによって主語の人称推定において、談話内での焦点や格フレームだけではなく [7]、語彙レベルの情報も活用しやすくなる。

具体的には、2 節で理論的背景を記述した後、3 節にて自然会話コーパスを用いた人称推定モデルの構築、及び、モデル検証を行った。結果として、データの量や範囲といった根本的な課題は残りつつも、話者の認識と言語表現の相関を想定することが人称推定に役立つことを示した。特に一人称の省略といった外界照応に対して有効であり、文脈照応を得意とする既存の手法と相補的な役割を担える可能性がある。また話者の認識という情報は、ゼロ主語補完に限らず、係り受け解析や無助詞に対するケースマーキングといった他の課題に対しても役立つことも考えられた。

## 2 言語に現れる話者の認識

証拠性 (evidentiality) [6]、なわ張り論 [12]、共感性 (empathy) [9] といった認識論 (epistemology) に分類される言語理論は、話し手の命題に対する認識が言語形式に現れることを主張している。Plungian [6] によれば、命題に対する認識は直接経験 (心的態度や自身の五感による体験) と間接経験 (伝聞や推論など)

に分類され、それぞれが直接形、間接形という文形に対応する (直接形、間接形というのは断定・不断定という対立に置き換えることができる [12])。例えば、冒頭の例文における命題は「X は寒いコト」であり、X が「私」の場合は、話者の直接経験にあたるため直接形が取られる (1a)。反対に X が「彼」である場合は、それは話者にとって間接経験であるため (彼が寒いかどうかは本人にしかわからない) 間接形が用いられる (1b)。このように話者の認識と文形はリンクしており、その一貫性の規則が日本語の人称制限として現れている。そして、自身のことは直接的に、他者のことは間接的に経験することの方が多という原理的な在り方から、直接形が一人称優勢、間接形が二・三人称優勢の素性と考えられる。ただし、自身のことであっても間接的に経験した場合は間接形で表現され (例 < 酩酊状態で何も覚えていない > 「俺は昨日道で寝てたらしい。」)、他者のことであっても直接的に経験した場合は直接形で表現される (例 < 太郎が家に入っていくところをはっきり見た > 「太郎はさっき家に入っていったよ。」)。

また、命題のタイプによっては人称制限の現れ方が変化する。その一つが話者の心的態度であり、Plungian [6] の分類では endophoric と呼ばれるタイプの命題である。感情・感覚形容詞 (例 寒い、辛い)、願望表現 (例 ~たい、欲しい)、思考動詞 (例 思う、考える)、意志動詞 (例 つもり、動詞+てやる) であり、これらは当人しか直接経験できないため、直接形においてははっきりと一人称制限が生じる。認識と文形の相関を仮定すると、世界知識から人称制限の分布の変化を予想できる。例えば、未来のことは自分のことであっても分からないため、比較的一人称と間接形が共起しやすいが、現在や過去のことについては、既に経験していることであるため間接形で表すのは難しいだろう。また、自身のことを間接的に経験することは一般的に不可能であるため、間接形と一人称の共起は先述したような限られた文脈以外では不可能である。

間接形に含まれるのは、疑問形、evidentials (例 みたい、ようだ)<sup>1</sup>、その他非断定形式 (例 ~かなあ、~じゃないですか) である。

その他、受動態は一人称優勢 [9]、終助詞「ね」は二・三人称優勢 [12] の素性として考えられ、また、人称制限が主文に限定されること [13]、一人称の方が省略されやすいこと [10] などを踏まえると (紙幅の都合上、それぞれの詳細は先行研究参照) 図 1 のようにまとめることができる。

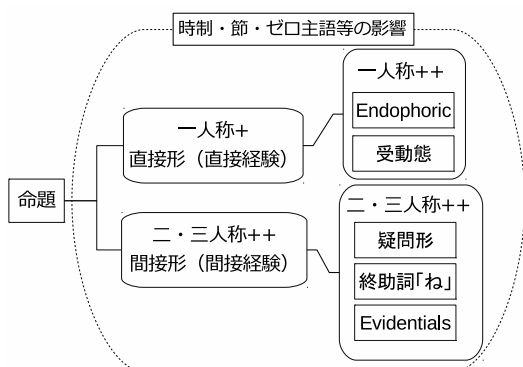


図 1: 文形及び語彙素性と人称の関係性「+」は当該の環境下でその人称が優勢であることを示している。例えば、直接形は一人称優勢であるが、endophoric 要素があればさらにその優勢性が高まる。

### 3 実験

#### 3.1 方法

##### 3.1.1 コーパス

BTSJ (Basic Transcription System for Japanese) による日本語話し言葉コーパス ([8], 以下 BTSJ) の初対面 (女性同士) の会話を対象とし、モデル構築用の学習データ (498 述語) と検証用のテストデータ (200 述語) を用意した。テストデータは 2 セット (各 100 述語) あり、一つは学習データにも含まれているペア (テストデータ A、ただし発話は異なる) もう一つは学習データに含まれていないペア (テストデータ B) である。

アノテーション作業においては、各ペアごとでランダムに一つの発話文を選択後、そこから順に述語を取り出していき、50 述語に達した時点でそのペアからの収集を終了とした。一発話文に複数の述語が含まれていた場合、その全てを対象とした。述語が省略されているもの、相槌や応答詞とみなされるもの (例 すごい、そうなんだ) 定型表現 (例 何て言うんだろう) については対象外とした。学習データが最終的に 498

<sup>1</sup>Evidentials とは、話し手が命題に関する情報をどのように得たかを示す語彙であり、「らしい」「ようだ」「そうだ」「みたい」「動詞 + がる」などが含まれる [1]

述語となっているのは、アノテーション作業中对象基準の変更があったためである。

##### 3.1.2 解析

一般化線形混合モデル (logit ; [5], 及び論文内の引用を参照) を用いて人称推定モデルを構築した。従属変数は、一人称、または、二・三人称のバイナリー値である<sup>2</sup>。独立変数は基礎素性と語彙素性の 2 つに分かれ (表 1)、基礎素性という大まかな枠組みで構築したモデル (ベースモデル) でどの程度推定できるのか、また、そのモデルに対して語彙素性を加えることでどの程度モデルが良くなるか検討する。

また、人によって一人称が多く出現したり (自分の話をよくする人)、三人称が多く出現したり (人の話をよくする人) といった個人差が考えられるため、発話者の切片をランダム要因として投入した。

モデル比較時には、計算を収束させやすくするため適宜センタリングを行ったが [2]、センタリングの有無によって結果が変わらないことはその都度確認した。以下で述べるモデルにおいては、全て文形のみセンタリングした。これは今回採用した最も複雑なモデルの計算を収束させるためであり、比較対象も同条件とした。

#### 3.2 結果：モデル構築

モデル検証における比較のために、4 つのモデルを構築した。表 2 が示しているのは、各モデルにおける要因とその係数である<sup>3</sup>。

係数が正であるものが一人称素性で、負であるものが二・三人称素性である。全体として、節 (補文)、endophoric、受動態、過去時制が一人称素性で、その他が二・三人称素性となっている。ミニムモデル (MIN) は述部に関する素性を含まないモデルで、ベースモデル (BAS) はそれに文形と文形と節の交互作用を加えたものである。また、BAS は基礎素性のモデル比較において最終的に残ったものでもある。マクスモデル (MAX) は、BAS に対して語彙素性を追加していき、計算が収束する範囲で最良と判断されたモデルである。evidentials モデル (EVI) は、MAX の文形の代わりに evidentials を含めたモデルである。

<sup>2</sup>一人称と非一人称というバイナリー値としたのは、データ数が少ないのと最初の課題として話者が否かを取り上げるためである。

<sup>3</sup>例えば、マクスモデルは以下の式 1 のように表せる。

$$\begin{aligned} \text{logit}[P(1stperson)] = & - .58 - 1.09 \times form + .67 \times clause \\ & - 2.03 \times omit + 2.62 \times endophoric \\ & - 3.00 \times interrogative - 2.03 \times neparticle \\ & + 3.04 \times passive + .83 \times past \\ & - 2.71 \times endophoric \times clause - 2.99 \times endophoric \times form \end{aligned} \quad (1)$$

表 1: 使用素性

素性の種類	素性	水準とコーディング
基礎素性	文形 (form)	直接形 (-0.5) か間接形 (0.5)
	節 (clause)	主文 (0) か補文 (1)
	主語省略 (omit)	主語省略 (0) か主語明示 (1)
語彙素性	endophoric	述部に当該素性がなければ 0、あれば 1
	evidentials	
	疑問形 (interrogative)	
	終助詞「ね」(neparticle)	
	受動態 (passive)	
	過去時制 (past)	

表 2: 各モデルにおける推定値

モデル	切片	文形	節	省略	end	evi	疑問	「ね」	受動	過去	交互作用
ミニム (MIN)	-.18		.90	-1.93							
ベース (BAS)	-.86	-2.72	1.39	-2.18							節:文形 1.94
evi(EVI)	-.14		.71	-1.97	3.65	-1.54	-4.81	-1.81	3.34	.85	end:節 -2.21 end:evid -.84
マクス (MAX)	-.58	-1.09	.67	-2.03	2.62		-3.00	-2.03	3.04	.83	end:節 -2.71 end:文形 -2.99

### 3.3 結果：モデル検証

前述した 4 つのモデルによる人称推定の結果が表 3 である。全体としては、テストデータ A の方がテストデータ B よりも正答率が良かった。原因としては、テストデータ A は学習データにも含まれていた発話者によるものであり、その重なりがうまく働いた可能性が考えられる。また単純にテストデータ B において推定の難しい環境が多かったかもしれず、この点についてはデータ数を今後増やしていく必要があるだろう。さて、全体の推定結果(表 3 上 3 段)から見ていく。MIN と BAS を比べると、両データにおいて MIN が二・三人称のみ多く正解していたのが、BAS では一人称の正答率を向上させ、かつ、二・三人称の正答率を 60%以上の水準に保っている。これは文形が人称との関係性のある程度持っていることを示唆している。BAS に語彙素性を加えた EVI と MAX は、二つとも精度が向上しており、特に二・三人称をより正しく推定できるようになっている。ただし、MAX は一人称の正答率が悪くなっており、過度に二・三人称に偏った推定を行ってしまった。一方 EVI は、両データで一人称の正答率をほぼ保ったまま、二・三人称の正答率を 15~20%引き上げている。MAX の過度の偏りは間接形と疑問形の重なりが引き起こしたものと考えられる。EVI では間接形という大きな括りではなく、

evidentials を分けたことが、丁度良いバランスとなったのだろう。次にゼロ主語だけの結果(表 3 下 3 段)と比較すると、どのモデルも一人称の正答率は上がっているものの、二・三人称の正答率が落ちている。省略されやすいのは一人称で、モデルもそれを反映しているため、二・三人称が省略された際に一人称と誤ってしまうのだろう。ただし、EVI と MAX はテストデータ A では 70%以上を保ったが、テストデータ B では約 50%と極端に下がった。先述したように、この差異がモデル構築時の問題なのか、データの偏りの問題なのかは、データ数を増やして確認する必要がある。

表 4 は一人称を正解とした時の適合率、再現率、F-値である。BAS は再現率は高くなっているものの、適合率は低く、一人称に偏った推定を行っている。EVI と MAX は比較的バランスよく推定を行っており、特に EVI はテストデータ B (全体)を除いた 3 つの環境で F-値が 0.75 を超えており、今回構築したモデルの中では最も性能の良いモデルといえるだろう。

## 4 おわりに

今回、比較的容易に言語学の知見を機械処理による人称推定に役立てることができた。そして、言語学の記述は一対一対応の規則ベースではなく、大きな体系に従った確率ベースでも表現しうることを示した。

表 3: 各モデルの推定結果 (%)

		テストデータ A				テストデータ B			
		MIN	BAS	EVI	MAX	MIN	BAS	EVI	MAX
全体	全体	70.0	70.0	83.0	82.0	66.0	68.0	75.0	73.0
	一人称	30.3	75.8	75.8	66.7	47.4	78.9	76.3	68.4
	二・三人称	89.6	67.2	86.6	89.6	77.4	61.3	74.2	75.8
ゼロ主語	省略時全体	60.0	60.0	77.4	75.8	55.6	58.7	69.8	66.7
	一人称	35.7	89.3	82.1	71.4	56.3	93.8	90.6	81.3
	二・三人称	79.4	35.3	73.5	79.4	54.8	22.6	48.4	51.6

<sup>a</sup> テストデータ A 全体: 計 100 述語中、一人称 33 述語、二・三人称 67 述語    <sup>b</sup> テストデータ A ゼロ主語: 計 62 述語中、一人称 28 述語、二・三人称 34 述語

<sup>c</sup> テストデータ B 全体: 計 100 述語中、一人称 38 述語、二・三人称 62 述語    <sup>d</sup> テストデータ B ゼロ主語: 計 63 述語中、一人称 32 述語、二・三人称 31 述語

表 4: 1 人称を正当とした時の各モデルの適合率、再現率、F-値

		テストデータ A				テストデータ B			
全体		MIN	BAS	EVI	MAX	MIN	BAS	EVI	MAX
全体	適合率	.588	.532	.735	.758	.563	.555	.644	.634
	再現率	.303	.758	.758	.667	.474	.789	.763	.684
	F-値	.400	.625	.746	.701	.514	.652	.699	.658
ゼロ主語	適合率	.588	.532	.719	.741	.563	.555	.644	.634
	再現率	.357	.893	.821	.714	.563	.938	.906	.813
	F-値	.444	.667	.767	.727	.563	.698	.753	.712

本論のモデルは主に外界照応を対象としている。これまでの手法では、照応詞に対して一定の確率以上の候補が文脈上存在しない場合に外界照応と判定していたが [4, 7]、このモデルでは言語特徴を用いて積極的に一人称を推定することができる。この点で、センタリング理論をはじめとする従来の自然言語処理における文脈照応技術とは相補的な関係になれる可能性がある。

述部の語彙情報を用いたことは他にも利点があり、まずは倒置や無助詞といった自然会話で現れやすい非標準的なパターンに対して強みがある。加えて、主語の人称推定に留まらず、係り受け解析や無助詞のケースマーキングに役立つ情報となる可能性がある。例えば、「太郎、もう早く家帰りたいよー俺」といった発話があったとき、「帰る」の主語は「たい」という endophoric 素性によって一人称と推定されており、この情報が後方に無格で出現した「俺」を主格として扱うために効くかもしれない。

また、FactBank[3] や拡張モダリティ [11] など、文の付加的な意味情報 (命題の事実性や発話者の態度など) の解析に取り組む研究もある。本論でモデル化した直接経験が間接経験かという話者の命題に対する認識も、それらの意味情報の一つとして考えることが可能であり、評価抽出や質疑応答といった他の機械処理

において有益な情報となりうるだろう。

しかし、今回のデータに関して、量、範囲、著者一人によるアノテーションといった課題が多く残っており、この点は改善していく必要がある。

## 参考文献

- [1] McCready Eric and Ogata Norry. Evidentiality, modality and probability. *Linguistics and Philosophy*, Vol. 30, No. 2, pp. 147–206, 2007.
- [2] Baayen R.H. *Analyzing linguistic data: A practical Introduction to statistics*. Cambridge University Press, 2008.
- [3] Sauri Roser and Pustejovsky James. Factbank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 2009.
- [4] Iida Ryu, Inui Kentaro, and Matsumoto Yuji. On the issue of combining anaphoricity determination and antecedent identification in anaphora resolution. In *International Conference on Natural Language Processing and Knowledge Engineering*, pp. 244–249, 2005.
- [5] Jaeger T. Florian. Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of memory and language*, Vol. 59, No. 4, pp. 434–446, 2008.
- [6] Plungian Vladimir A. The place of evidentiality within the universal grammatical space. *Journal of Pragmatics*, pp. 349–357, 2001.
- [7] Hayashibe Yuta, Komachi Mamoru, and Matsumoto Yuji. Japanese predicate argument structure analysis exploiting argument position and type. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pp. 201–209, 2011.
- [8] 宇佐美まゆみ監修. BTSJ による日本語話し言葉コーパス (2011 年版). 人間の相互作用研究のための多言語会話コーパスの構築とその語用論的分析方法の開発 平成 20-22 年度科学研究費補助金基盤研究 B (課題番号 20320072) 研究成果, 2011.
- [9] 久野峰. 談話の文法. 大修館書店, 1978.
- [10] 久野峰. 新日本文法研究. 大修館書店, 1983.
- [11] 松吉俊, 佐尾とせ, 乾健太郎, 松本裕二. 拡張モダリティタグ付与コーパスの設計と構築. 言語処理学会第 17 回年次大会, pp. 147–150, 2011.
- [12] 神尾昭雄. 情報のなかり理論—言語の機能的分析—. 大修館書店, 1990.
- [13] 長谷川信子. 一人称の省略: モダリティと授与構文. *Scientific Approaches to Language*, Vol. 5, pp. 127–147, 2006.