

## 決定木による日本語ゼロ代名詞補完の性能評価

山本和英

隅田英一郎

E-mail: {yamamoto, sumita}@itl.atr.co.jp

ATR 音声翻訳通信研究所

## 1 はじめに

日本語対話文における格要素の省略補完について述べる。主語や目的語などの表示が義務的でない日本語の言語処理においては、これら省略される(非明示の)格要素を補う処理が重要である。これら省略のある諸言語から英独など必須格を持つ言語への翻訳処理を行なう際には、補完処理は必須の処理となる。

格要素の省略補完には、多要素性と相互依存性を同時に考慮することが可能な枠組みが必要である。日本語を対象にした省略補完研究は、村田ら[村田97]、江原ら[江原96]、Nakaiwa et al.[Nak96]の研究などがある。これに対し我々は、省略補完知識を決定木による表現、及び省略情報の正解付きコーパスから言語現象と補完すべき省略の関係を帰納的に機械学習し、これによって日本語対話文の格省略を補完する手法を提案した[Yam97]。本稿では提案手法の有効性を確認するために行なった諸実験とその結果について報告する。なお、省略の検出処理は本稿の対象外とする。

## 2 決定木を用いた省略補完

## 2.1 補完情報の付与

決定木学習による学習、並びにテストを行なうことを目的に、対話コーパスに補完内容の情報を付与した。今回付与したタグの種類を表1に示す。

本稿では、表に示すように6種類のタグを設定した。タグの付与に関して、補完内容は日本語のみを考慮して付与した。日本語においては、特定されない人称を省略要素とする文、つまり一般的な「人」を念頭において発話していると考えられる文がしばしば見受けられるが、このような省略に対してはタグ(g)を設定した。

省略位置以前の要素に照応先がある場合は、これを一括して(a)のタグを付与した。本稿では、対話文に頻出する言語外省略の補完に主眼を置き、文脈処理を要する省略に関してはその認知のみを行ない、先行詞の特定は別処理で行なうと仮定した。

## 2.2 使用属性

省略された格要素を補完するためには、種々の情報を考慮して行なわなければならない。本研究では以下に述べる属性に対して決定木学習を行なった。

表 1: 付与したタグの種類

タグ	意味
{1sg}{1pl}	一人称単数、複数
{2sg}{2pl}	二人称単数、複数
{g}	一般
{a}	照応的

表 2: 使用属性とその要素数

属性	属性値	要素数
内容語(用言)	意味属性	100
内容語(格要素)	意味属性	100
機能語	助詞、助動詞など	166
言語外知識	話者情報	1
合計		367

**内容語の意味属性** 省略の対象となる文において、どのような内容語が含まれているかに関する情報。内容語は大きく、用言に関する情報と格要素に関する情報に分かれる。内容語の意味属性としては角川類語新辞典における中分類(100属性)を使用した。

**機能語の出現** 用言に後接する付属語群、及び助詞などの機能語の出現に関する情報。尊敬(召し上がる)動詞や疑問詞(どこ)の集合もまとめて機能語とした。また、受給表現(やる)や動詞「する」「なる」なども特殊な機能語と見なした。

**言語外情報** 言語外情報としては、発話された文の話者が情報提供者か情報授受者か、という属性を使用した。

以上のように、本研究では計367の属性を使用した。その内訳を表2に示す。

属性の照合方法は、形態素列とのマッチングで行なった。すなわち、補完対象の用言を中心にして、前、直前、当該位置、直後、後の5種類のうちどの位置に出現するかという情報をすべての属性に予め与えた。複文や重文などの、文が複数の単文からなる場合には、接続助詞を境界にして近似的に単文に分割した。

表 3: コーパスの話題分類

記号	話題
$H_1$	ホテルにおける部屋の予約、変更、解約
$H_2$	ホテル利用に関する問い合わせ、苦情
$H_R$	その他のホテル関連の話題
$R$	ホテル関係以外の旅行会話

### 2.3 コーパスと話題

本研究で使用したコーパスは、チケット予約、観光案内などにおける二者の会話を収録した ATR 旅行会話コーパス (以下、「コーパス」と呼ぶ) である。おおよそ 1 対話は 20 文から 40 文で構成され、平均は 26 文である。

コーパスは、ホテルにおける旅行者とフロントの対話を中心に、広範な範囲の対話を収録している。本稿ではこれを大きく表 3 に示す 4 つの話題に分類した。なお、話題  $H$  を  $H = H_1 + H_2 + H_R$ 、つまりコーパス中のホテルに関する全話題、話題  $T$  を  $T = H_1 + H_2 + H_R + R$ 、つまり旅行会話の全話題と定義する。

### 3 実験

本節では、学習された決定木による省略補完の有効性を検証する。まず、格の違いによる影響についての検証を行なった後、学習量、決定木学習の話題依存性、使用属性による相違の三点から検討を行なう。本稿では、性能評価尺度として F 値 (F-measure) を用いる。F 値は、再現率 (recall) と適合率 (precision) を一つの尺度として表現するために使用される尺度で、 $R$  を再現率、 $P$  を適合率としたとき、以下の式で定義する。

$$F = \frac{(\beta^2 + 1) \times P \times R}{\beta^2 \times P + R} \quad (1)$$

ここで、パラメータ  $\beta$  は適合率の再現率に対する相対的な重要性である。本稿ではこのパラメータを  $\beta = 1$  とし、F 値はパーセントを単位として表現する。

#### 3.1 表層格との関係

ここでは、日本語の主たる表層格「が」「を」「に」に対する補完精度の比較を行なうことによって、格との関係を考察する。本稿では省略された文の述部によって「が」格を二つに分離し、別個のものとして取り扱った。すなわち、一つは述部が動詞の場合、もう一つは述部が形容詞、形容動詞、あるいは「名詞+判定詞 (だ/である)」の場合である。以下では、前者を「が」格 (動)、後者を「が」格 (形) と表記する。

表 4: 格による補完精度の比較

格	$\langle 1sg \rangle$	$\langle 2sg \rangle$	$\langle a \rangle$	全省略
「が」(形)	58.3	68.1	85.9	79.7
「を」	66.7	—	97.7	95.6
「に」	95.2	95.7	81.9	91.7
「が」(動)	84.7	81.1	82.0	78.7

実験は、それぞれの格について 300 対話を学習対話とし、それらに含まれない 100 対話をテスト対話として未知文テストを行なった。また、すべての格で学習時に表 2 の 367 属性を用意した。実験の結果を表 4 に示す。

表からわかるように、「が」(動) と「が」(形) との比較では全体としての補完精度に大きな差異はないが、個別の人称に対する精度では両者に明確な差異が現れている。表には現れていないが、「が」(形) の補完人称に比較的多くの  $\langle a \rangle$  が含まれているため、 $\langle 1sg \rangle$  あるいは  $\langle 2sg \rangle$  に対する学習が十分に行なわれず、比較的低い精度になったと推察される。一方「を」格については、90% 以上の省略が照応的 ( $\langle a \rangle$ ) であり、言語外照応がほとんどないことから非常に高い数字となった。今後「を」格の補完処理には、特に文脈処理が必須であることを示す結果となった。

「に」格に関しては十分な性能が得られた。このように高い性能が得られた背景には、2 者対話を対象にしたテキストであること、話題が旅行会話に限定されているために使用される述語がある程度限定されることなどが考えられる。「に」格の多くは間接目的語で言語外照応が多かった<sup>1</sup>ため、限られた候補の中から最尤解を選択する問題に有効な本手法が有利に機能したものと考えられる。

#### 3.2 学習量との関係

学習量との関係を見るために以下の実験を行なった。学習量として、25、50、100、200、400 対話の 5 種類の集合を作成した。ここで、これらの集合は包含関係となるように作成した。テスト集合はこれらのいずれにも含まれない 100 会話 (「が」格の省略数:1685) を用意した。また、学習属性は表 2 のものを使用した。補完内容と学習量の差をグラフにしたものを片対数グラフで図 1 に示す。

図 1 によれば、ほぼすべての人称に関して学習量の増加と共に性能が単調に向上している。なお、再現率、適合率共に単調増加の傾向を示した。ただし、その増加の割合は徐々に鈍化し、 $\langle 1sg \rangle$  に関しては 400 対話で精度がわずかに減少していることがわかる。ま

<sup>1</sup>コーパスでは「に」格の省略の約 9 割が言語外省略で、照応的省略  $\langle a \rangle$  は 1 割前後であった。

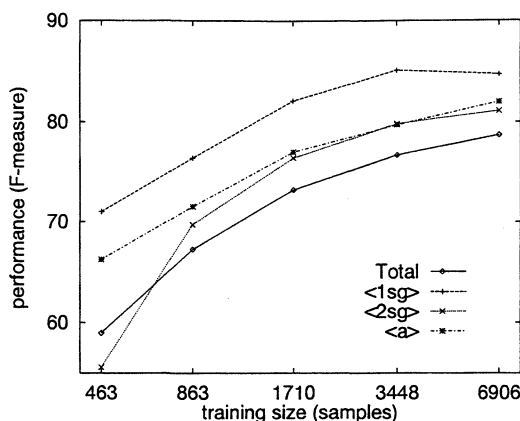


図 1: 学習量と補完性能

表 5: 決定木の話題依存性

学習 / テスト 構成比 (%)	$H_1$	$H_2$	$H_R$	$R$	合計
	20.1	27.7	11.2	40.9	100.0
$H_1$ /	78.1	55.9	65.3	61.6	63.7
$H_2$ /	71.3	67.0	62.6	62.6	65.6
$R$ /	75.1	61.7	61.1	75.4	69.9
$T$ /	73.4	62.5	62.6	66.2	66.2
$T - H_R$ /	73.7	61.9	59.5	63.9	64.8

たグラフにはないが、比較的学習サンプル数の少なかった  $\langle 2pl \rangle$  や  $\langle g \rangle$  が、学習量増加に伴い大きく精度が向上している。その様子から、 $\langle 1pl \rangle$  を含めたこれらの人称に関しては学習量の増加によって一層の精度向上が予想される。

一方、その他の人称並びに全体的な精度に関しては、全体として  $10^4 \sim 10^5$  サンプルの学習量で十分であると言える。また、この時の補完精度 (補完精度の上限) は 80% ~ 85% となると予想する。

### 3.3 話題依存性

ここでは、実験の結果と共に、決定木学習の話題依存性を議論する。学習用のテキストとして、四つの話題  $H_1$ 、 $H_2$ 、 $R$ 、 $T$  に属する対話を 50 対話無作為に抽出し、これによって決定木学習を行なった。テスト用の対話は前節と同一の未学習 100 対話を使用し、未知文テストを行なった。このとき、属性は表 2 の 367 属性を使用した。表 5 に、テスト対話 (= コーパス全体) の話題別構成比、並びに話題依存性を示す。表の縦は学習対話の話題、横はテスト対話の話題を示し、値は F 値で表現した。

表 6: 属性と補完率との関係

	$\langle 1sg \rangle$	$\langle 2sg \rangle$	$\langle a \rangle$	全省略
全属性	82.1	76.4	77.0	73.2
言語情報のみ	81.9	76.9	77.4	73.2
機能語のみ	75.4	68.0	67.2	65.3
内容語のみ	75.1	58.1	74.5	65.0
用言情報のみ	72.3	55.6	71.1	61.9

表に示すように、学習対話とテスト対話が一致している時に、話題に関係なく最も良好な性能となった。あらかじめテスト対話の話題が絞り込める問題に対しては、できるだけ同一の話題のみによって学習することが望ましい。学習文の話題別性能では、その構成比が最も高かったため話題  $R$  が最も高い性能を示した。

また表によると、広範な話題で学習を行なった場合 ( $T$ ) に、全体としても平均以上の補完精度を示した。学習文とテスト文の話題が同一の場合を除くと、 $T$  はすべての話題に対して良好な性能を示していることが観察される。このことから、テスト文の話題が未知の場合は、広範な話題に対して学習を行なうことが最も有効である。ただし表 5 の最下段に示すように、全く未知の話題 ( $H_R$ ) に対しては若干精度が低下する。たとえば少量の学習であっても、未学習よりはかなり優位であることがわかる。

### 3.4 属性との関係

本節では、格要素の省略補完の問題解決に、どの程度話題に依存した情報が必要かを議論する。ここでは、言語情報のみ (366 属性)、機能語のみ (166 属性)、内容語のみ (200 属性)、用言情報のみ (100 属性) の 4 種類の属性集合を用意した。これらはいずれも表 2 に示した属性の部分集合である。実験は 100 会話の学習、未学習 100 対話のテストにより行なった。実験結果を表 6 に示す。比較対象として、全属性に実験の結果を表の「全属性」欄に示す。

表より、言語情報のみを使用した学習では、言語外情報を加えた場合とほとんど同程度の精度が得られた。これは、言語外情報 (特に実験で用意した話者情報) がそれほど省略補完に重要でないことを示す。

機能語のみで決定木学習を行なった場合、全体で 8% 程度の精度低下が観察された。この結果は、話題に依存しない機能語のみで決定木学習した場合に、その精度に限界があることを示している。また、機能語のみの結果は照応的な省略 ( $\langle a \rangle$ ) の認知に対して大きな精度低下が見られることから、内容語は比較的照応関係の維持に寄与していることが予想される。

内容語のみの場合はさらに低い精度となった。日本

表 7: 学習量による包含率の変化

	が <sup>g</sup> /25	が <sup>g</sup> /100	が <sup>g</sup> /400
:here 43(意向)	100.0	100.0	100.0
:here 41(思考)	72.8	84.8	86.5
… か(終助詞)	53.1	83.2	66.3
… てくださる (尊敬語)	9.1	49.1	49.8
… ていただく	—	39.9	36.8
… する	—	33.2	33.9
:before 72(施設)	4.1	22.0	26.1
:before 94(建物)	55.1	0.5	3.8
:before 83(言語)	28.5	9.8	7.7
話者の役割	25.1	1.1	1.3
	11.7	9.1	20.5

語対話文においては、内容語よりも一部の機能語の存在によって省略が可能となる場合が多いということをこの結果は示している。さらに出現した用言のみで人称の決定を行なった場合は最も悪い精度を示した。

#### 4 議論

補完内容の決定に対する各属性の重要性を見る一つの尺度として、「包含率」を定義する。ある属性の属性包含率は、その属性が決定木の意志決定に使用されているサンプル数の、全サンプル数に対する割合である。

学習量との関係を議論した3.2節での実験における主な属性の包含率を表7に示す。包含率が100%である属性(=最上部で検査される属性)は「:here 43」つまり対象となる用言の意味コードが43(意向)であるかどうか、であった。この属性や「:here 41」(思考)には共に「思う/考える/願う」などの語が含まれていることから、この両属性は一種の機能語のような役割を果たしていると考えられる。

ただ、この両属性のように学習量に関係なく包含率の高い属性はむしろ少数で、同一の格、同一の話題であっても学習量の増加と共に多くの属性の包含率が変化していることが観察できる。表によると、学習量が少ない時は:before、つまり対象となる用言以前にどのような内容語が存在したかに関して多くの注意が注がれ、学習量の増加に伴って機能語、特に尊敬を示す語の存在によって人称を判断するようになることがわかる。

次に、格要素別の属性包含率を表8に示す。ここでも、「が」(動)と「が」(形)の明確な差異が見受けられる。「が」(形)の決定木は他の各要素の内容、例えば「で」などの格の存在とその格要素に興味を持ち、一方「が」(動)は述語と一部の重要な機能語に対して

表 8: 格による包含率の変化

	が <sup>g</sup> /400	が <sup>g</sup> (形)	に
… ございます	—	100.0	—
:before 16(状態)	5.1	68.5	0.5
:before 34(陳述)	5.3	59.0	11.2
で(格助詞)	5.2	23.9	1.9
お…/ご…	46.4	7.0	100.0
:here 43(意向)	100.0	—	49.8
:here 41(思考)	86.5	—	43.5
話者役割	20.5	33.1	28.0

質問を行なっている。また「に」格補完に作成した決定木は、一部の相違はあるものの「が」(動)と類似の傾向を示した。

#### 5 まとめ

日本語対話文の格要素省略に対して、決定木による補完処理の表現および機械学習によって補完知識を獲得する手法を提案した。本手法により獲得した決定木で未学習文に対してテストを行なった結果、「が」格と「に」格に対して本手法の有効性を確認することができた。また、決定木を用いた補完手法に関して、処理の有効性を学習量、話題依存性、使用属性との関係の3点から議論した。

#### 参考文献

- [Nak96] NAKAIWA, H. and SHIRAI, S.: Anaphora Resolution of Japanese Zero Pronouns with Deictic Reference, In *Proc. of COLING-96*, pp. 812-817 (1996).
- [Yam97] YAMAMOTO, K., SUMITA, E., FURUSE, O., and IIDA, H.: Ellipsis Resolution in Dialogues via Decision-Tree Learning, In *Proc. of Natural Language Processing Pacific-Rim Symposium (NLP RS'97)*, pp. 423-428 (1997).
- [江原96] 江原暉将, 金淵培: 確率モデルによるゼロ主語の補完, 自然言語処理, Vol. 3, No. 4, pp. 67-86 (1996).
- [村田97] 村田真樹, 長尾真: 用例や表層表現を用いた日本語文章中の指示詞・代名詞・ゼロ代名詞の指示対象の推定, 自然言語処理, Vol. 4, No. 1, pp. 87-109 (1997).