

英日機械翻訳における代名詞翻訳の改良

吉見 毅彦[†]

代名詞を含む英文を日本語として適格で自然な文に翻訳するためには、英語の代名詞を日本語の代名詞としてそのまま表現せず、ゼロ代名詞化したり他の表現に置き換えたりする必要がある。ゼロ代名詞化に関しては、人手で記述された規則による方法が既に提案されている。本稿では、1) ゼロ代名詞化に加え、他の表現に置き換えるべき場合も扱い、2) 規則を人手で記述するのではなく、決定木学習によって自動的に学習する方法を示す。学習に利用する属性は、ゼロ代名詞化に関してこれまでに解明されている言語学的制約や、ゼロ代名詞の復元に関する工学的研究で着目された手がかりを参考にして選択した。提案手法を我々の英日機械翻訳システム *Power E/J* による訳文に対して適用したところ、ゼロ代名詞化するか否かの判定を行なう場合の精度が 79.9%、ゼロ代名詞化するか否かに加え他の表現に書き換えるか否かの判定も行なう場合の精度が 72.2% となり、人手で記述された規則の精度に近い精度が得られた。また、選択した属性には、書き換え精度を低下させる属性は含まれておらず、ゼロ代名詞化に関する言語学的制約だけでなく、ゼロ代名詞の復元に関する手がかりも利用できることが明らかになった。

キーワード: 代名詞, ゼロ代名詞化, 機械学習, 決定木, 機械翻訳

Improvement of Translation Quality of Pronouns in an English-to-Japanese MT System

TAKEHIKO YOSHIMI[†]

In order to translate sentences of English containing pronouns into natural and suitable Japanese, it is frequently necessary either to eliminate pronouns or to turn them into some other expressions. As for eliminating unwanted pronouns, a set of manually-written rules has already been presented. In this article we propose to 1) offer a way of substituting unwanted pronouns for other expressions as well as eliminating them, and 2) use a decision tree learning algorithm to learn rules automatically from a corpus, without requiring human intervention. The features used for learning are selected from the linguistic constraints we have so far understood which apply on zero pronominalisation, and from the clues which have been used for anaphora resolution of zero pronouns in the engineering studies. Having applied the proposed method to the translation results of our English-to-Japanese machine translation system *Power E/J*, we found that in the cases where the judgement whether zero pronominalisation should be applied the accuracy of translation was 79.9%, where in addition to the above judgement the substitution of pronouns for other expressions was applied the accuracy was 72.2%. These results are well comparable with those obtained by hand-written rules. It also became clear that none of the selected features lowers the accuracy, which means we can use as features for our purpose not only the linguistic constraints on zero pronominalisation but also the clues for restoring zero pronouns.

KeyWords: *Pronoun, Zero Pronominalisation, Machine Learning, Decision Tree, Machine Translation*

1 はじめに

よく知られているように、人間が英語を日本語に翻訳するとき、英語の代名詞を日本語の代名詞としては表現せず、ゼロ代名詞化したり他の表現に置き換えたりすることが多い。これに対して、従来の英日機械翻訳システムでは、多くの場合、英語の代名詞はそのまま日本語の代名詞に訳される。このように代名詞を直訳すると、英文が伝えている意味と異なる意味を伝える訳文が生成されたり、文意は同じでも不自然で読みにくい訳文が生成されてしまうという問題が生じる。従って、品質の高い英日機械翻訳システムを実現するためには、ゼロ代名詞化する必要のある代名詞や他の表現に置き換えるべき代名詞を直訳しないようにすることが重要な課題となる。直訳すべきでない代名詞を認識することは一見単純であるように思えるが、ゼロ代名詞化や他の表現への書き換えには様々な要因が絡んでいるため、代名詞を直訳してもよい場合とそうでない場合をどのように区別すればよいかはそれほど自明なことではない。なお、本稿では、紛れない限り、人称代名詞と限定的機能を持つ所有代名詞 (Quirk, Greenbaum, Leech, and Svartvik 1985) を単に代名詞と呼ぶ。

ゼロ代名詞化に関する工学的な研究は、滑川ら (滑川裕樹, 乾健太郎, 徳永健伸, 田中穂積 1999) や宮ら (宮東城 永田守男 2000) による報告がある程度で、これまであまり行なわれていない。宮らの方法では、機械翻訳システムの出力文から代名詞を消すかそのまま残すかの二値の判定が、代名詞とそれに付属する助詞の表記に着目して人手で記述した規則に基づいて行なわれる。しかし、二値の判定では次の文 (J1) のような場合に適切に対処できない。文 (J1) には、“she” が “Mary” を指しているという文 (E1) の文意を伝えないという問題がある (神崎高明 1994)。この問題に対処するために、文 (J1) から「彼女」を消すと「家を出る」の主語が「メアリー」であるのか「ジョン」であるのかが曖昧になるという問題が新たに生じる。主語の曖昧さが生じるのを抑え、かつ文 (E1) と同じ文意を伝えるためには、文 (J1’) のように「彼女」を「自分」に置き換える必要がある。

(E1) Mary scolded John before *she* left home.

(J1) 彼女が家を出る前に、メアリーはジョンを叱った。

(J1’) ϕ_{she} 家を出る前に、メアリーはジョンを叱った。

(J1’’) 自分が家を出る前に、メアリーはジョンを叱った。

また、代名詞をどのように書き換えるかは様々な要因によって決まるため、複雑に関連し合う要因を人手で整理し、その結果に基づいて規則を記述するより、統計的帰納学習法を利用して

† シャープ (株) 技術本部 システム開発センター, Corporate Research and Development Group, System Technology Development Center, SHARP Corporation

事例集から規則を自動的に作成するほうが適切であると考えられる。

このようなことから本稿では、1) 代名詞を消すか残すかの二値の判定ではなく、消すか残すかあるいは他の表現に置き換えるかの多値の判定を行ない、2) 規則の記述を手で行なうのではなく、決定木学習アルゴリズムを利用して事例集から規則を自動的に作成する方法を示す。以下、代名詞を直訳するとどのような問題が生じるかを 2 節で整理する。次に 3 節で決定木学習に簡単に触れる。4 節では決定木学習に必要な正解付きコーパスの作成について述べ、5 節で決定木学習に使用した属性について説明する。6 節では、提案手法の有効性を検証するために行なった実験の結果について考察する。

2 代名詞の日英対照比較

英語の代名詞を日本語の代名詞に直訳したときに生じる問題として、英文では代名詞がある名詞句を指していないのに、訳文では指しているように解釈されてしまうという問題と、その逆に、英文では代名詞がある名詞句を指しているのに、訳文ではそのように解釈できないという問題がある。本節では、後者の問題について検討する。

2.1 後方照応 (語順逆転)

英語では代名詞による後方照応が可能であるのに対して、日本語では基本的に不可能である (神崎高明 1994)。このため、次の文 (E2) では “he” は “John” を指していると解釈されるのに対して、文 (J2) では「彼」が「ジョン」を指しているとは解釈されない。

(E2) When *he* arrived, John went straight to the bank.

(J2) 彼が着くと、ジョンは、銀行にまっすぐ行った。

また、英語では前方照応であっても、日本語に翻訳するとき主節と従属節の順序を逆転させると、文 (J3) のような不適格な訳文となる。文 (J2) や文 (J3) を適格文にするには、文 (J3') のように代名詞「彼」をゼロ代名詞化する必要がある。

(E3) John went straight to the bank when *he* arrived.

(J3) 彼が着くと、ジョンは、銀行にまっすぐ行った。

(J3') ϕ_{he} 着くと、ジョンは、銀行にまっすぐ行った。

後方照応は限られた条件の下でしか用いられず使用頻度も低いいため、あまり問題にならないともいえるが、従来の機械翻訳システムによる処理では主節と従属節の順序の逆転は頻繁に生じるため、代名詞に対する適切な書き換えが重要となる。

2.2 総称名詞句と不定代名詞

英語の代名詞は総称的に使われている名詞句を指すことができるのに対して、日本語の代名詞は総称的な名詞句を指すことはできない(神崎高明 1994)。例えば、次の文 (E4) を機械翻訳システムで処理すると、文 (J4) のように翻訳される。

(E4) By the time the average American reaches the age of 70, *he* consumes 13 tons of beef.

(J4) 平均的米国人は70歳に達する時までに、彼は、13トンの牛肉を消費する。

(J4') 平均的米国人は70歳に達する時までに、 ϕ_{he} 13トンの牛肉を消費する。

文 (E4) において、“the average American” は特定の人物を指しているのではなく総称的な意味で用いられているが、“he” は “the average American” を指していると解釈できる。これに対して、文 (J4) では「彼」が「平均的米国人」を指しているとは解釈されにくい。不適格文 (J4) を適格文にするには、文 (J4') のように代名詞「彼」をゼロ代名詞化しなければならない。

総称名詞句と同じく、“nobody” や “everyone” などの不定代名詞も英語では代名詞によって指示されるが日本語では指示されない。従って、次の文 (J5) は、文 (J5') のように、代名詞「彼」をゼロ代名詞化するかあるいは再帰代名詞「自分」に置き換える必要がある。

(E5) Everyone loves *his* mother.

(J5) 全ての人は、彼の母親を愛する。

(J5') 全ての人は、 $\{\phi_{his}, \text{自分の}\}$ 母親を愛する。

2.3 伝達文の話法

英語では話法が比較的確立されており直接話法と間接話法の対立が形式上明確であるのに対して、日本語では厳密な間接話法は事実上不可能である(安西徹雄 1983)。このため、英語の間接話法は、日本語では直接話法あるいはそれに近い形式で表現される。このとき、英語の代名詞を直訳すると、不適格なあるいは不自然な訳文になることがある。例えば、次の文 (E6) を機械翻訳システムで翻訳した文 (J6) は、引用符が存在しないだけで実質的には直接話法になっている。このとき、代名詞 “he” を直訳することは不適切である。代名詞「彼」をゼロ代名詞化するか「私」に置き換えると、日本語として自然な文 (J6') が得られる。

(E6) John said *he* usually gets up at 6 o'clock.

(J6) ジョンは、彼がいつも6時に起きると言った。

(J6') ジョンは、 $\{\phi_{he}, \text{私は}\}$ いつも6時に起きると言った。

話法の問題は、伝達文の出現頻度が比較的低い技術文書を対象としている限りあまり生じないが、本研究では、伝達文の出現頻度が比較的高い新聞記事を主な対象としているため、被伝達部に現れる代名詞を適切に翻訳することが重要な課題となる。

2.4 they の訳

三人称複数の代名詞 “they”, “their”, “them” を直訳するとき, 少なくとも「彼ら」か「それら」に訳し分けなければならない。この訳し分けの手がかりになるのは共起的意味制約や照応関係であるが, 従来の機械翻訳システムではこれらの手がかりが必ずしも有効に活用できていないため, 訳し分けの精度は十分高いとはいえない。実際, 実験に用いたシステムでは, 文 (E7) のように “they” が人間を表わす表現を指していると考えられるときでも「それら」と訳されることが多く, 次の文 (J7) のような不自然な訳文となる。文 (J7) のような翻訳では, おそらく先行文中に存在する人間を表わす表現を「それら」が指しているとは解釈されない。文 (J7) の主節の主語の「それら」はゼロ代名詞化すると主語の曖昧さが生じる可能性があるので, 「彼ら」に置き換えなければならない。

(E7) *They gathered to discuss what to do to prevent environmental destruction.*

(J7) それらは、公害を防ぐために何をするかを議論するために、集まった。

(J7') 彼らは、公害を防ぐために何をするかを議論するために、集まった。

3 決定木学習の利用

本研究では, 代名詞の書き換えを実現するために, 統計的帰納学習手法の一つである決定木学習を利用する。代名詞をどのようなときにどのように書き換えればよいかを決定する要因は, 多岐に渡っており複雑に関連し合っていると考えられる。書き換えに関連する様々な要因を人手で明示的にかつ統一的に記述することは容易ではない。これに対して, 決定木による方法では, 様々な要因の統合と重要な要因の選別が自動的に行なわれる。

決定木学習には C4.5 (Quinlan 1992) を利用する。C4.5 は, 事例を一般化することによって決定木の形式で分類モデルを帰納的に作成する。事例は, あらかじめ定められた属性とクラスによって表現される。決定木は, クラスを表わす終端節点と, 一つの属性を調べるテストに対応する非終端節点 (判別節点) から成り, 根節点から終端節点に向けて判別節点でのテストの結果に従って, 終端節点に対応するクラスに事例を分類する。C4.5 による決定木の作成は, 事例の集合 T を n 個の部分集合に分割するテスト X を利得比基準に従って選択することによって行なわれる。利得比基準では, T をテスト X で分割することによって得られる情報量を, T を n 個の部分集合に分割することによって得られる全情報量で割った値 (利得比) を最大にするようなテスト X が選ばれる。

4 正解付きコーパスの作成

4.1 作成の基本方針

不適切な代名詞をどのように書き換えるべきかを人手で判断し、書き換え方を示すラベルを付与したコーパスを次のような方針で作成した。ラベルは決定木学習のクラスに対応する。

- (1) 代名詞を書き換えるか否か、書き換える場合どのように書き換えるかは、文意の正しさはもちろん、文意の曖昧さや文の簡潔さ、自然さなどの条件も考慮に入れて決定しなければならない。これらすべての条件を満たす書き換え方が常に存在するとは限らず、いくつかの条件は満たすが他の条件は満たさないような書き換えしかできないこともある。このようなときには、文意の正しさが保たれる書き換えと文意の曖昧さが生じない書き換えを優先する。すなわち、原文の文意と訳文の文意が異なったり、原文には存在しない曖昧さが訳文で生じてしまったりするような書き換えは、たとえその書き換えによって文の簡潔さや自然さが増すとしても行なわない。
- (2) ある代名詞をどのように書き換えればよいかは、その代名詞を含む文がどのような文脈で用いられているかを参照しなければ確定できない場合と、その文の情報だけでは確実に判断できる場合がある。特に、代名詞を消すか残すかに関しては、ある特定の先行文が存在すれば消すことができる(消しても文意が曖昧にならない)が、そのような先行文が存在しなければ消すことはできない(消すと曖昧になる)ということが起こりやすい。今回のコーパス作成では、先行文が存在しないという条件の下で適切な書き換えになるようにする。

4.2 代名詞に付与するラベル

代名詞の書き換え方を示すラベルとして、[残]、[消]、[換(私)]、[換(我々)]、[換(彼ら)]、[換(自分)]の六種類を設けた。ラベル[残]は、着目している代名詞を書き換えると、意味的に不適格になるか、文意の曖昧さが増すか、文の不自然さが増すときに付与する。[消]は、代名詞を消しても意味的に適格であり、曖昧さが生じず、文の自然さが増すときに付与する。[換(私)]、[換(我々)]、[換(彼ら)]、[換(自分)]は、それぞれ、代名詞を括弧内の語(「私」など)に置き換えると意味的に適格であり、曖昧さが生じず、より自然であるときに付与する。[換]は、代名詞を消しても残しても不適切である場合に対処するために設けたラベルであり、[換(私)]と[換(我々)]は2.3節で述べた話法の問題の解決を、また[換(彼ら)]は2.4節で述べた“they”の訳語の問題の解決を主に目指している。

ある一つの代名詞に対して複数通りのラベル付けができることがある。例えば「私は、彼が私を助けてくれた恩を忘れはしない。」という文において、「私を」は消しても「自分を」に置き

換えてもよい．このようなときには，次の優先順位に従ってラベル付けを行なうことにした．

[消] > [残] > [換]

4.3 作成された正解付きコーパスの諸元

正解付きコーパス作成用の資料には英字新聞記事 300 記事を用いた．この 300 記事は 4240 文から成るが，このうち代名詞を含む文は 1845 文であった．この 1845 文を我々の機械翻訳システムで処理して得られた訳文から，代名詞の問題を除けば文法的にも意味的にも適格である文を選別し，1015 文を得た．

1015 文には合計で 1350 個の代名詞が出現していたが，1015 文のうち代名詞を一つだけ含む文が 734 文 (72.3%)，二つ含む文が 218 文 (21.5%) と，ほとんどの文が代名詞を一つか二つ含む文であった．最も多い場合，一文に五つ含まれていた．

1015 文における各代名詞の出現頻度と各代名詞に付与されたラベルの分布を表 1 に示す．「彼」、「それ」、「それら」で出現した全代名詞の 75.2% を占めている．また「あなた」は全く出現していなかった．設定したラベルのうち「我々」への置き換えを示すラベル [換 (我々)] は付与する必要がなかった．ラベル [消] や [残] の数に比べ [換] の件数が少ないが，これは 4.2 節で述べたラベルの優先順位を設けたことに一因がある．表 1 において「それら自身」、「彼自身」、「それ自身」は再帰代名詞 “*oneself*” の訳ではなく，“*one's own*” の訳であり，これらに付与されているラベル [消] は，全体を消すのではなく，それぞれ「それら」、「彼」、「それ」を消し「自身」は残すことを意味する．

表 1 各代名詞の出現頻度と付与したラベルの分布

	[消]	[残]	[換 (私)]	[換 (我々)]	[換 (彼ら)]	[換 (自分)]	計
彼	212	263	12	0	0	19	506
それ	161	109	0	0	0	0	270
それら	158	22	0	0	58	1	239
我々	28	99	0	0	0	1	128
私	54	56	0	0	0	1	111
彼女	21	21	1	0	0	2	45
彼ら	1	24	0	0	0	0	25
それら自身	7	0	0	0	3	0	10
自分	6	2	0	0	0	0	8
彼自身	3	0	0	0	0	0	3
それ自身	3	0	0	0	0	0	3
私自身	0	1	0	0	0	0	1
それら双方共	0	1	0	0	0	0	1
計	654 (48.4%)	598 (44.3%)	13 (1.0%)	0 (0.0%)	61 (4.5%)	24 (1.8%)	1350 (100%)

4.4 正解付きコーパス作成に要する労力

決定木学習による代名詞書き換えにおける課題の一つは、代名詞に関する英日比較対照コーパスをいかにして作成するかである。現在のところ、大規模で一貫性のある正解付きコーパスはほとんど整備されていない。正解付きコーパスの作成に必要な作業は、対訳コーパスの作成と、代名詞への正解ラベルの付与の二つの作業であるが、一般にこれらは両方とも労力を要する作業である。対訳コーパスには、本研究では、人手で作成したものではなく、機械翻訳システムで作成したものを利用している。なぜならば、機械翻訳システムからの出力文に含まれる不適切な代名詞を書き換えることが本研究の目的であるため、決定木学習に必要な属性は、人間による訳文から得られる属性ではなく、機械翻訳システムによる訳文から得られる属性であるからである。このように、人手による対訳コーパスを必要としない本研究では、正解ラベルの付与の労力だけで済む。正解付きコーパスの作成に今回要した労力は、およそ 1.6 人月であった。

5 着目した属性

決定木学習による代名詞の書き換えでは、学習に用いる属性としてどのような情報を取り入れるかが重要である。代名詞の書き換えには、文の情報構造(旧情報と新情報の区別)や結束性、視点なども関連している(久野 1978, 1983; 水谷静夫 1983; 神尾昭雄 1985; Hinds 1986; 高見健一 1997)。しかし、現状の技術レベルではこれらを直接利用することは容易でないで、これらに関連し現状の技術レベルで扱える情報を利用せざるを得ない。また、日英機械翻訳などにおいてゼロ代名詞をどのようにして復元するかが課題となっており、様々な手法が提案されている(藤澤伸二, 増山繁, 内藤昭三 1993; 工藤育男, 友清睦子 1993; 中岩浩巳, 池原悟 1993; 堂坂浩二 1994; 江原暉将, 金淵培 1996; Murata 1996; 中岩浩巳, 池原悟 1996)が、これらの手法で利用されている情報も属性として有効である可能性が高い。このようなことから本稿では、書き換えに影響しうる属性として、表 2 に示す 13 種類の属性に着目した。これらの属性には構文レベルのものや照応レベルのものも含まれているが、構文解析や照応解析は行わずに、「茶筌」(松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸 1999)による形態素解析の結果に基づいて、後述する単純な方法で正解付きコーパスから求めた。表 2 の「属性数」欄は作成した正解付きコーパスに実際に出現した属性値の数であり、「属性値」欄は値の例である。

以下、各属性について、その属性に着目した理由と、それを文から抽出する方法について述べる。

代名詞の表記 表 1 からわかるように、適切な書き換えの傾向は代名詞毎に異なる。例えば、「それら」はほとんどの場合書き換える必要があるのに対して、「我々」や「彼ら」は書き換えられにくい。従って、代名詞の表記は書き換え方を分類する手がかりとして利用できる可能性がある。

表 2 着目した属性の一覧

属性名	属性数	属性値 (一部)
代名詞の表記	14	それ, その, それら, それら自身, それら双方共, それ自身, 我々, 私, 私自身, 自分, 彼, 彼ら, 彼自身, 彼女
代名詞の付属語	25	から, が, と, と共に, に, において, にとつての, には, によって, に関する, に対して, の, のうちの, のために
等位の助詞	2	有, 無
係り先の意味コード	69	111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 130, 131, 132, 133, 未定義
係り先の付属語	45	から, が, で, と, と共に, に, において, における, について, にもかかわらず, 意志形, 基本形, 連用テ接続, 連用ニ接続, 連用形
節タイプ	4	主節, 伝達節, 従属節, 対象外
従属節の接続形 (文頭側)	48	<が>, <けれども>, <て>, <と, と>, <と, ので, て>, <と, ので>, <と, のに, と>, <と, 連用形>, <と>, <ので>, <ば, 連用形>, <連用形, て>, <連用形, ば>, <連用形, 連用形>, <連用形>, 無
従属節の接続形 (文末側)	65	<けれども, と>, <て>, <で, と>, <と, て>, <と, と, 連用形>, <と, と, ば>, <と, と>, <と, 連用形, と>, <と>, <ので, と>, <ので>, <ば>, <連用形, と>, <連用形, ば>, <連用形>, 無
不定代名詞	3	だれ, 全て, 無
同一代名詞 (文頭側)	40	<[消] 我々>, <[消] 我々, [消] 我々>, <[消] それ>, <[消] それら>, <[残] 私>, <[残] 我々, [消] 我々, [消] 我々>, <[残] 我々>, <[消] 彼>, <[消] 彼, [残] 彼>, <[消] 彼女>, <[残] 彼>, <[残] それ>, 無
同一代名詞 (文末側)	41	<[消] それ>, <[消] それら>, <[消] 私>, <[消] 我々>, <[消] 我々, [残] 我々, [換(自分)] 我々>, <[消] 彼>, <[消] 彼, [消] 彼>, <[消] 彼女>, <[換(自分)] 彼>, <[換(自分)] 我々>, <[残] それ>, <[残] 我々>, <[残] 我々, [換(自分)] 我々>, 無
照応	4	成立, 不成立, 先行名詞無, 文脈外
接続表現	15	けれども, しかし, しかしながら, そこで, そして, それでも, それに, だが, なぜなら, にもかかわらず, または, もしくは, 一方, 従って, 無

代名詞の付属語 一般に, 旧情報は省略されやすく, 新情報は省略されにくいことが知られている (久野 1978) . 旧情報と新情報を区別する手がかりの一つとして助詞がある . 助詞「は」には旧情報を示す一機能があり, 助詞「が」には新情報を示す機能がある . このようにゼロ代名詞化は助詞の影響を受けるため, 代名詞に付属する助詞の表記を属性として取り入れる .

等位の助詞 次の文 (J8) における「彼」や「私」のように代名詞が等位句の構成要素になっているとき, その代名詞をゼロ代名詞化してはならない . 従って, 代名詞が等位句の構成要素になっているかどうかを, 助詞「と」が直前に存在するかどうかで判断する . なお, 「彼」が等位句の構成要素になっているかどうかは, 属性「代名詞の付属語」によって判断できる .

(E8) He and I went there.

(J8) 彼と私は, そこに行った。

係り先の意味コードと係り先の付属語 代名詞の書き換えには、「代名詞の表記」や「代名詞の付属語」のような形態素語彙レベルの情報だけでなく、構文レベルの情報、例えば代名詞が係っている用言や体言に関する情報も重要である。特に用言には、待遇表現(水谷静夫 1983)や受給表現のようにゼロ代名詞を復元するために有効な情報が含まれている。これらの情報は、「伺う」や「いらっしゃる」のような動詞から得られる場合と、「致します」や「下さい」のような助動詞(相当表現)から得られる場合(工藤育男・友清睦子 1993; 中岩浩巳・池原悟 1996)があるが、属性「係り先の意味コード」は前者を対象とし、属性「係り先の付属語」は後者を対象とする。

ここで、代名詞がどの用言あるいは体言に係っているかを正確に認識するためには構文解析を行なう必要がある。しかし、本稿では、次のような簡単な手順で近似的に認識する。代名詞が体言に係りうるとき、最も近い体言に係るものとし、その体言の意味標識を属性「係り先の意味コード」の値とし、その体言に付属する助詞の表記を属性「係り先の付属語」の値とする。代名詞が用言に係りうるときは、代名詞に付属する助詞が「は」でありその直後に読点「、」が存在するとき最も遠い用言に係るものとし、それ以外は最も近い用言に係るものとする。意味標識としては、分類語彙表(国立国語研究所 1984)の意味コードの上位三桁を利用する。用言あるいは体言の意味コードが分類語彙表に記述されていないとき、属性「係り先の意味コード」の値は未定義とする。

節タイプ 2.3節で述べたように、英語の間接話法の被伝達節で使用されている代名詞は、日本語では「私」などに置き換える必要がある。従って、代名詞が被伝達節に属しているかどうかで代名詞書き換えの手がかりになる。また、代名詞が属している節が被伝達節でないときでも、主節か(被伝達節以外の)従属節かの区別は有用であると考えられる。

代名詞が属する節の種類についても正確な認識には構文解析が必要であるが、次のような簡単な手順で近似的に認識する。代名詞が属している節が被伝達節かどうかは、引用の助詞「と」や引用符が代名詞より文末側に存在するかどうかで判断する。代名詞が属している節が主節か従属節の判断は次のように行なう。もし、代名詞に付属する助詞が「は」でありその直後に読点「、」が存在すれば、主節に属するものとする。さもなければ、代名詞より文末側に存在する用言の数が一つだけならば主節に属するものとし、複数ならば従属節に属するものとする。ただし、この区別は代名詞が用言に係っているときのみ行ない、体言に係っているときは対象外とする。

従属節の接続形 久野の考察(久野 1983)によれば、次の文(J9)の適格性が低いのに対して文(J9')が適格である理由の一つは、「ので」と「間に」の違いにある。文(J9')では、因果関係を表わす「ので」の存在によって、「山田」を「批判した」の目的語とする蓋然性が高くなっている。これに対して、「間に」によって従属節と主節が結び付けられている文(J9)では、主節で「山田」を目的語とする蓋然性はそれほど高くない。

(J9) 山田がアメリカに行っている間に、私が学会で $\phi_{山田}$ 批判した。

(J9') 山田が久し振りで帰って来たので、 $\phi_{山田}$ 夕食に招いた。

このように従属節の接続形が代名詞の書き換えに影響を及ぼすことから、着目している代名詞より文頭側あるいは文末側にどのような従属節の接続形が存在するかを属性として取り入れる。

不定代名詞 2.2 節で述べたように、日本語で総称名詞句や不定代名詞を指せるのはゼロ代名詞か再帰代名詞「自分」である。このことを考慮して、着目している代名詞より文頭側に不定代名詞が存在するかどうか、存在する場合それはどのような不定代名詞かを属性として取り入れる。総称名詞句は、その認識が容易ではないため、本稿では対象外とする¹。

同一代名詞 ある代名詞をどのように書き換えるかには、それと同じ代名詞が文のどの位置に存在しており、かつそれらがどのように書き換えられるかも関連している可能性がある。次の文 (J10) において文頭の「彼」を消す理由の一つは、(主節の主語である)「彼」が後方に存在しており、それが残されることであろう。

(E10) *He does his work when he feels like doing it.*

(J10) 彼がそれをしたいと思うとき、彼は、彼の仕事をする。

(J10') $\phi_{he}\phi_{it}$ したいと思うとき、彼は、 ϕ_{his} 仕事をする。

表 2 の「同一代名詞 (文頭側)」と「同一代名詞 (文末側)」の属性値はそれぞれ、着目している代名詞より文頭側、文末側に存在する同一代名詞のリストを表わしている。各代名詞に付与されているラベルは、その代名詞に対する書き換えを示す。例えば<[残] 我々,[換 (自分)] 我々>は、残される「我々」と「自分」に置き換えられる「我々」がこの順に存在することを意味する。

照応 藤沢らの調査 (藤澤伸二他 1993) によれば、明示されている代名詞の先行詞は代名詞が含まれている文の直前の文に現れることが最も多いのに対して、ゼロ代名詞の先行詞はゼロ代名詞が存在する文と同じ文に現れることが最も多い。このことから、逆に、代名詞とその先行詞が同じ文に存在すれば、代名詞をゼロ代名詞化する必要性が高いとも考えられる。

照応に関する属性値は、照応成立、照応不成立、先行名詞無、文脈外照応とする。代名詞が一人称か二人称であるときは、先行名詞を指さない文脈外照応であるとみなす。代名詞が三人称であり先行名詞が存在するとき、分類語彙表の意味コードを利用して照応の成立、不成立を次のような簡単な手順で判定する。代名詞が人間を指せる「彼」、「彼女」、「彼ら」であるとき、人間を表わす分類語彙表の意味コード 120, 121, 123, 124 を持つ先行名詞が存在すれば照応成立とみなす。代名詞が人間以外を指せる「それ」、「それら」であるときは、上記以外の意味

¹ 総称名詞句の認識に村田の手法 (Murata 1996) を利用することが考えられる。しかし、村田の手法では構文構造が正しく得られることが前提になっているのに対して、本稿ではごく簡単な処理によって得られる属性を利用する方針であるため、総称名詞句は対象外とする。

コードを持つ先行名詞が存在すれば照応成立とみなす．先行名詞の意味コードが分類語彙表で未定義であるときは照応不成立とする．着目している代名詞と同一の代名詞が存在するときはそれとの間で照応が成立するとみなす．

接続表現 代名詞の書き換えには，同一文内の照応だけでなく，先行文との関係も影響を及ぼす．二つの文をつなぐ言語的手段の一つとして，先行文との修辭的關係 (Cohen 1984; Mann 1984; Knott and Dale 1994) を表わす接続表現の使用がある (Halliday and Hasan 1976)．このような先行文との修辭的關係を考慮に入れ，文頭に存在する接続詞の表記を属性として取り入れる．

6 実験と考察

代名詞書き換えの有効性を検証するために，次の五種類の実験を行なった．決定木学習アルゴリズムには C4.5 をオプションなしで利用した．いずれの実験でも十分割の交差検定を行なった．

実験 1 代名詞を消すか残すかの二クラスでの実験

実験 2 代名詞を消すか残すか他の表現に置き換えるかの五クラスでの実験

実験 3 各クラスの事例数をほぼ同数にしての実験 (五クラス)

実験 4 各属性の効果を確認するための実験 (五クラス)

実験 5 代名詞毎の実験 (五クラス)

6.1 [消]/[残] の二クラスでの実験

まず予備的な実験として，宮ら (宮東城・永田守男 2000) の設定と同じく，代名詞を消すか残すかを分類する場合について精度を調べる実験を行なった．実験に用いた事例は，作成した 1350 事例のうちラベル [消] または [残] が付与されている 1252 事例である．誤分類率は，十分割の交差検定の平均で，訓練事例に対して 6.4%，試験事例に対して 21.8%であった．

試験事例に対する正誤行列を表 3 に示す．表中の数値は，十分割の交差検定による結果の和である．表 3 より，正しく分類された事例は，消すべき代名詞が正しく消された 564 件と，残すべき代名詞が正しく残された 437 件である．他方，誤って分類された事例は，残すべき代名詞が誤って消された 161 件と，消すべき代名詞が誤って残された 90 件である．誤りのうち前者の場合は，曖昧さの増大につながり翻訳品質に悪影響を及ぼす．これに対して後者の場合は，翻訳品質は現状維持となり，実用上新たな悪影響は出ないと考えてよい．このことから，提案手法で [消]/[残] の二クラスの分類を行なった場合の真の精度は $79.9\%((564+437)/1252)$ ，実用精度は $87.1\%((564+437+90)/1252)$ となる．実験に用いた 1252 事例のうち，代名詞を消すべき事例は 654 件，残すべき事例は 598 件である (表 1 参照) ので，単純な多数決基準に従って分類

したときの誤分類率は 47.8%(598/1252) である．この値を基準誤分類率とみなせば，提案手法は有効であると考えられる．

実験条件が異なるので単純な比較はできないが，宮らの実験結果では人手で記述した規則の精度が 85.8% であることから，決定木学習によってほぼ同程度の精度が達成できているといえる．

表 3 [消]/[残] での分類精度

(a)		(b)		← classified as
564	(45.0%)	90	(7.2%)	(a): [消]
161	(12.9%)	437	(34.9%)	(b): [残]

6.2 [消]/[残]/[換(私)]/[換(彼ら)]/[換(自分)] の五クラスでの実験

本節では，代名詞を消しても残しておいても不適切な翻訳に対処するために他の表現に置き換えるというクラスを設定した場合の実験結果について考察する．実験対象は 1350 事例である．

誤分類率は，十分割の交差検定の平均で，訓練事例に対して 9.2%，試験事例に対して 29.2% であった．試験事例に対する正誤行列を表 4 に示す．表 4 より，提案手法で [消]/[残]/[換(私)]/[換(彼ら)]/[換(自分)] の五クラスの分類を行なった場合，真の精度は 72.2% ((566+409)/1350) である．また，[残] に分類された事例は翻訳品質に新たな悪影響を及ぼさないので，実用精度は 80.1% ((566+409+87+1+13+5)/1350) になる．

表 4 を見てまず気付くことは，新たに導入したクラス [換] に分類されるべき事例が一件も正しく分類されていないことである．クラス [換] の全事例が誤って分類された原因は，表 1 に示したように，このクラスの事例数がクラス [消] や [残] の事例数に比べて極端に少なかったため，適切な学習ができなかったことと考えられる．この点に関しては，クラス分布の偏りが少ない場合にどの程度の分類精度が得られるのかを確認するための実験を 6.3 節で別途行なう．

表 4 [消]/[残]/[換] での分類精度

(a)	(b)	(c)	(d)	(e)	← classified as
566	87	0	0	1	(a): [消]
187	409	1	1	0	(b): [残]
12	1	0	0	0	(c): [換(私)]
48	13	0	0	0	(d): [換(彼ら)]
19	5	0	0	0	(e): [換(自分)]

十分割の交差検定によって作成された十本の決定木のうちの一本の一部を図 1 に示す．図 1 の決定木では，属性「節タイプ」が最も重要であり，「代名詞の表記」，「代名詞の付属語」が次

に重要であるとみなされている。十本のうち七本の決定木で、このように、「節タイプ」が根節点に配置され、「代名詞の表記」、「代名詞の付属語」、「照応」、「同一代名詞(文頭側)」がその子節点に配置されていた。残り三本の決定木では、根節点に配置された属性は「代名詞の付属語」であり、その子節点に配置されたのは、「同一代名詞(文頭側)」、「従属節の接続形(文頭側)」、「従属節の接続形(文末側)」、「照応」、「代名詞の表記」、「係り先の意味コード」であった。宮らは代名詞と助詞の表記に着目して規則を記述しているが、これら形態素語彙レベルの属性以外に、「節タイプ」などの構文レベルの属性や「照応」などの照応レベルの属性も重要であることを、実験で得られた決定木の形状は示唆している。ただし、今回の実験では構文レベルの属性値と照応レベルの属性値は近似的な方法で決定しているが、これら重要視された属性の値をより正確に求め今後さらに検証を行なう必要がある。

5節で設定した属性のうち最も重要視されなかった属性は、「等位の助詞」と「不定代名詞」であった。「等位の助詞」と「不定代名詞」以外は全ての決定木に現れたが、「等位の助詞」は1本の決定木にしか現れず、「不定代名詞」はどの決定木にも現れなかった。これらが重要視されなかった理由は、等位句、不定代名詞がコーパス中にそれぞれ20個と少数しか存在しなかったことにあると考えられる。

6.3 クラス分布の偏りを排除した実験

6.2節の実験で見られたクラス分布の偏りの影響を抑えるために、各クラスの数をはほぼ同じになるように、クラス[消]が付与されている事例とクラス[残]が付与されている事例からそれぞれ50事例ずつを無作為に抽出し、これらの事例とクラス[換]の全事例を合わせた198事例を対象として実験を行なった。

この場合の誤分類率は、十分割の交差検定の平均で、訓練事例に対して7.6%、試験事例に対して35.3%であった。6.2節での実験結果(訓練事例:9.2%、試験事例:29.2%)に比べると試験事例に対する誤分類率が高くなっているが、これは事例数が少ないことが主な原因であろう。

試験事例に対する正誤行列を表5に示す。クラス[換(私)]と[換(自分)]の正解率は十分高いとはいえないが、6.2節での結果に比べると改善されている。また、クラス[換(彼ら)]は全事例が正しく分類されている。このことから、クラス[換]に関する事例を今後重点的に収集していくことが、全体の精度向上に有効である。

6.4 各属性の有効性を調べる実験

本節では、着目した個々の属性が分類精度にどの程度寄与しているかを調べる。各属性を利用しないときに誤分類率がどのように変化するかを表6に示す。表6より、精度に悪影響を及ぼしている属性は存在しないことがわかる。

特に分類に有効な属性は、「代名詞の表記」、「係り先の意味コード」、「係り先の付属語」、「従

```

節タイプ = 主節:
| 代名詞の表記 = 我々: [残] (15.0)
| 代名詞の表記 = 私: [残] (11.0/1.0)
| 代名詞の表記 = 彼ら: [残] (19.0)
| 代名詞の表記 = 彼女: [残] (7.0)
| 代名詞の表記 = それ:
| | 接続表現 = しかし: [消] (4.0)
| | 接続表現 = そして: [残] (7.0)
| | 接続表現 = なぜなら: [残] (1.0)
:   :
| | 接続表現 = 無:
| | | 係り先の意味コード = 114: [残] (1.0)
| | | 係り先の意味コード = 116: [消] (1.0)
:   :
| 代名詞の表記 = それら:
| | 従属節の接続形 (文頭側) = <ば>: [換 (彼ら)] (4.0)
| | 従属節の接続形 (文頭側) = <連用形>: [消] (1.0)
:   :
| 代名詞の表記 = 彼:
| | 係り先の意味コード = 231: [残] (106.0)
:   :
節タイプ = 従属節:
| 代名詞の付属語 = において: [消] (1.0)
| 代名詞の付属語 = に対して: [換 (自分)] (1.0)
| 代名詞の付属語 = のように: [残] (1.0)
| 代名詞の付属語 = の後: [消] (1.0)
| 代名詞の付属語 = は: [残] (3.0)
:   :
| 代名詞の付属語 = が:
| | 同一代名詞 (文末側) = <[消] それら>: [消] (3.0/1.0)
| | 同一代名詞 (文末側) = <[消] 彼>: [消] (8.0)
:   :
| 代名詞の付属語 = に:
| | 係り先の付属語 = こと: [換 (自分)] (1.0)
| | 係り先の付属語 = ば: [残] (5.0)
:   :
| 代名詞の付属語 = を:
| | 代名詞の表記 = それ: [消] (2.0/1.0)
| | 代名詞の表記 = それら: [換 (彼ら)] (1.0)
:   :

```

図 1 作成された決定木の一部

表 5 [消]/[残]/[換] での分類精度 (ほぼ同じ事例数)

(a)	(b)	(c)	(d)	(e)	← classified as
24	6	5	12	3	(a): [消]
9	32	3	3	3	(b): [残]
1	3	3	0	6	(c): [換 (私)]
0	0	0	61	0	(d): [換 (彼ら)]
4	5	5	1	9	(e): [換 (自分)]

属節の接続形(文末側)」である。属性「代名詞の表記」の有効性が高いことから、宮らが「代名詞の表記」を手がかりにしていることは適切であるといえる。属性「係り先の意味コード」の値は分類語彙表に基づいて与えたが、1350 事例のうち 220 事例 (16.3%) で、代名詞の係り先である用言や体言の意味コードが分類語彙表に記述されていないため、「係り先の意味コード」の値が未定義となっていた。これらに対して適切な意味コードを与えることができれば、精度向上につながる可能性が高い。

他方、分類精度の向上に寄与していない属性は、「等位の助詞」、「不定代名詞」、「同一代名詞(文頭側)」である。このうち、「等位の助詞」と「不定代名詞」が寄与しなかった理由は、事例数が少なかったことにある。「同一代名詞(文頭側)」が寄与しなかった理由は、4.3 節で示したように、代名詞を一つしか含まない文がコーパス全体の 72.3% を占めており、これらの文から作成される事例では「同一代名詞(文頭側)」の値がすべて同じ「無」になるため、有効に働かなかったのではないかと考えられる。しかし、「同一代名詞(文頭側)」が精度向上にまったく寄与していないのに対して、「同一代名詞(文末側)」は若干寄与している。この差がなぜ生じたかについては今後の検討課題である。

表 6 各属性の有効性

属性	誤分類率
全属性	9.2%
代名詞の表記	11.0%
代名詞の付属語	9.6%
等位の助詞	9.2%
係り先の意味コード	10.4%
係り先の付属語	10.0%
節タイプ	9.6%
従属節の接続形(文頭側)	9.9%
従属節の接続形(文末側)	10.2%
照応	9.3%
同一代名詞(文頭側)	9.2%
同一代名詞(文末側)	9.5%
接続表現	9.7%
不定代名詞	9.2%

6.5 代名詞毎の実験

分類精度は代名詞の種類によっても異なると予想される。そこで、表 1 に示した代名詞のうち、出現頻度が高い「彼」、「それ」、「それら」、「我々」、「私」の五種類の代名詞について、代名詞毎に実験を行なった。結果を表 7 に示す。基準誤分類率は、出現頻度が最も高いクラスを選んだときの誤分類率である。

表7を見ると、代名詞「それら」と「私」に対する分類精度が特に悪い。「それら」に関しては、適切に分類できないクラス [換] の多さが高い誤分類率の原因である。「私」に関しては、分類に有効であろうと当初考えていた手がかり (属性値) が適切に得られなかったことに原因がある。「私」に関する有力な手がかりの一つは受給表現や待遇表現であり、これらは「係り先の意味コード」や「係り先の付属語」の値として反映されるものと考えていた。ところが、実験に用いた機械翻訳システムからの出力文にはこれらの表現が含まれていなかった。今後、他のシステムからの出力文を用いてこれらの表現の有効性を検証していく必要がある。

表7 代名詞毎の誤分類率

代名詞	誤分類率		基準誤分類率
	訓練事例	試験事例	
彼	8.2%	24.7%	48.0%
それ	8.0%	28.5%	40.4%
それら	13.1%	35.1%	33.9%
我々	4.6%	14.8%	22.7%
私	13.0%	47.7%	49.5%

7 おわりに

本稿では、従来の英日機械翻訳システムにおける代名詞翻訳の問題点を挙げ、それらを解決する方法を提案した。すなわち、従来研究と異なり、1) 不適切な代名詞を消すか残すかあるいは他の表現に置き換えるかの判定を行ない、2) 決定木学習アルゴリズムを利用して事例集から規則を自動的に作成する方法を示した。評価実験を通じて明らかになった主要な点は、1) 人手で記述した規則の精度と同程度の精度が得られることと、2) ゼロ代名詞化に関する言語学的制約だけでなく、ゼロ代名詞の復元に関する手がかりも、ゼロ代名詞化を含む代名詞書き換えの可否を判断するための手がかりとして利用できることである。

今後の課題として、次のような点が残されている。

機械翻訳システムとの密な結合 提案手法は、既存の機械翻訳システムから独立した後編集と位置付けることができるが、既存システムの生成部に組み込むことも可能である。一般に生成部では目的言語の構文木を参照することができるので、提案手法を既存システムの内部に組み込むことによって正確な構文情報を得ることが可能になり、分類精度の向上が期待される。今後この点に関して検討を行なっていく。

用言属性の拡張 今回の実験では、代名詞が係っている用言に関する属性として用言の意味コードと用言に付属している助動詞 (相当表現) を用いたが、今後は結合価フレームも用言の属性として取り入れていく必要がある。なぜならば、代名詞がゼロ代名詞化されるかどうかは、

代名詞に付属している助詞の表記と用言の結合価フレームによって判断しなければならないこともあるからである。例えば、助詞「に」をとることが明らかな用言のとき、「に」格の代名詞をゼロ代名詞化しても、それを復元することは比較的容易であるので、ゼロ代名詞化されやすい。これに対して、「に」をとることが明らかでない用言のときは、「に」格の代名詞のゼロ代名詞化は起こりにくいものと考えられる。

テキスト属性の導入 今回の正解付きコーパス作成では、代名詞をどのように書き換えるかの判断を、先行文が存在しないという条件の下で行なった。しかし、一文単位では代名詞の存在が不自然でなくとも、代名詞を含む文が連続して出現すると、日本語テキストとして不自然になることもある。従って、テキストレベルでの属性を取り入れる必要がある。このとき、一文内で観察される現象に比べて、複数の文から成るテキストにおいて初めて観察される現象の頻度が低くなることは避けられない。このため、テキストレベルの属性の信頼性を保つためには、非常に大規模なコーパスが必要となる。

謝辞

議論に参加頂いた英日機械翻訳グループの諸氏に感謝します。また、本稿の改善に有益なコメントを下さった査読者の方に感謝いたします。

参考文献

- 安西徹雄 (1983). 英語の発想. 講談社.
- Cohen, R. (1984). "A Computational Theory of the Function of Clue Words in Argument Understanding." In *Proceedings of the 10th International Conference on Computational Linguistics (COLING84)*, pp. 251-258.
- 堂坂浩二 (1994). "語用論的条件の解釈に基づく日本語ゼロ代名詞の指示対象同定." 情報処理学会論文誌, 35 (5), 768-778.
- 江原暉将 金淵培 (1996). "確率モデルによるゼロ主語の補完." 自然言語処理, 3 (4), 67-86.
- 藤澤伸二, 増山繁, 内藤昭三 (1993). "日本語文章における照応・省略現象の基本的検討." 情報処理学会論文誌, 34 (9), 1909-1918.
- Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English*. English Language Series 9. Longman.
- Hinds, J. (1986). Situation vs. Person Focus 日本語らしさと英語らしさ. くろしお出版.
- 神尾昭雄 (1985). "談話における視点." 日本語学, 4 (12), 10-21.
- 神崎高明 (1994). 日英語代名詞の研究. 研究社出版.
- 国立国語研究所 (1984). 分類語彙表. 秀英出版.
- Knott, A. and Dale, R. (1994). "Using Linguistic Phenomena to Motivate a Set of Coherence Relations." *Discourse Processes*, 18 (1), 35-62.

- <http://www.cogsci.ed.ac.uk/hcrc/publications/rp-39>.
- 工藤育男 友清睦子 (1993). “日本語の述部の特性を用いた省略の補完機構について.” 電子情報通信学会論文誌, **J76-D-II** (3), 624-635.
- 久野 (1978). 談話の文法. 大修館書店.
- 久野 (1983). 新日本文法研究, 第 11 章 ゼロ代名詞, 主題, 予測性, pp. 175-191. 大修館書店.
- Mann, W. C. (1984). “Discourse Structure for Text Generation.” In *Proceedings of the 10th International Conference on Computational Linguistics (COLING84)*, pp. 367-375.
- 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸 (1999). “日本語形態素解析システム『茶筌』 version 2.0 使用説明書第二版.” NAIST Technical Report NAIST-IS-TR99012, 奈良先端科学技術大学院大学情報科学研究科. <http://cl.aist-nara.ac.jp/lab/nlt/chasen/>.
- 宮東城 永田守男 (2000). “形態素を利用した自然な和文訳のためのゼロ代名詞生成規則.” 技術研究報告 NLC2000-5, 電子情報通信学会.
- 水谷静夫 (1983). “待遇表現の仕組み.” 林四郎, 荻野綱男, 田中幸子, 樺島忠夫 (編), 運用 I, 朝倉日本語新講座 1, pp. 158-178. 朝倉書店.
- Murata, M. (1996). *Anaphora Resolution in Japanese Sentences Using Surface Expressions and Examples*. Doctoral Dissertation, School of Electrical and Electronic Engineering, Faculty of Engineering, Kyoto University. <http://www-lab25.kuee.kyoto-u.ac.jp/member/murata/publish.html/96-doctor.ps>.
- 中岩浩巳 池原悟 (1993). “日英翻訳システムにおける用言意味属性を用いたゼロ代名詞照応解析.” 情報処理学会論文誌, **34** (8), 1705-1715.
- 中岩浩巳 池原悟 (1996). “語用論的・意味論的制約を用いた日本語ゼロ代名詞の文内照応解析.” 自然言語処理, **3** (4), 49-65.
- 滑川裕樹, 乾健太郎, 徳永健伸, 田中穂積 (1999). “照応・省略を含む日本語論説文生成.” 言語処理学会第 5 回年次大会発表論文集, pp. 153-156.
- Quinlan, J. (1992). *C4.5: Programs for Machine Learning*. Morgan Kaufmann. 古川康一 訳. AI によるデータ解析, トッパン, 1995.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.
- 高見健一 (1997). “統語論 機能主義.” 西光義弘 (編), 日英語対照による英語学概論, pp. 137-183. くろしお出版.

略歴

吉見 毅彦: 1987 年電気通信大学大学院計算機科学専攻修士課程修了. 1987 年よりシャープ (株) にて機械翻訳システムの研究開発に従事. 1999 年神戸大学大学院自然科学研究科博士課程修了.

(2000年11月6日受付)

(2001年4月13日採録)