

日本語表層表現を手がかりとした名詞の指示性と数の推定

村田真樹 黒橋 禎夫 長尾 眞

京都大学工学部 電気工学第二教室

〒606 京都市左京区吉田本町

あらまし

日本語を英語に翻訳する時には、日本語にはないが英語では必要な冠詞や数の問題に直面する。この難しい問題を解決するために、われわれは文章における名詞の指示性と数をそれぞれ三種類に分類した。この論文では、名詞の指示性と数とその名詞の現れる文中の言葉によりかなりの程度推定できることを示した。その推定のための規則はエキスパートシステムの書き換え規則に類する形で、文法書などから得られる知識をもとに経験的に作成した。これらの規則を作るのに利用したテキストでの正解率は、指示性で85.5%、数で89.0%であった。規則を作るのに利用していないいくつかのテキストでの正解率は平均して指示性で68.9%、数で85.6%という結果となった。

和文キーワード 表層表現, 指示性, 数, 総称名詞, 定名詞, 不定名詞

An Estimate of Referential Property and Number of Nouns from Japanese Surface Expressions

Masaki Murata Sadao Kurohashi Makoto Nagao

Department of Electrical Engineering, Kyoto University

Yoshida-honmachi, Sakyo, Kyoto 606, Japan

Abstract

Translating Japanese nouns into English, we face the problem of articles and numbers which the Japanese language does not have, but which are necessary for the English composition. To solve this difficult problem we classified the referential property and the number of nouns into three types respectively. This paper shows that the referential property and the number of nouns can be estimated fairly reliably by the words in the sentence. Many rules for the estimation were written in forms similar to rewriting rules in expert systems. We obtained the correct recognition scores of 85.5% and 89.0% in the estimation of referential property and number respectively for the sentences which were used for the construction of our rules. We tested these rules for some other texts, and obtained the scores of 68.9% and 85.6% respectively.

英文 key words surface expression, referential property, number, generic noun, definite noun, indefinite noun

1 はじめに

機械翻訳システムには、少し微妙だが重要な問題として冠詞の問題がある。例えば、

本というのは人間の成長に欠かせません。 (1)

の「本」は総称的な使われ方で、英語では“a book”にも“books”にも“the book”にも訳される。これに対して、

昨日僕が貸した本は読みましたか。 (2)

の「本」は英語では“the book”と訳される。

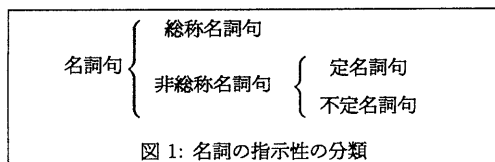
冠詞の問題は、多くの場合、名詞の指示性と数を明らかにすることによって解決できる。文(1)の「本」は総称名詞句で数は未定であり、“a book”にも“books”にも“the book”にも訳される。また、文(2)の「本」は定名詞句でほとんどの場合単数と解釈してよい。よって、英語では“the book”となる。

名詞の指示性と数は日本語の表層表現から得られることが多い。例えば、文(1)では「というのは」という表現から「本」が総称名詞句とわかる。文(2)では修飾節「昨日僕が貸した」が限定していることから「本」が定名詞句とわかる。そこで、本研究では名詞の指示性と数を日本語文中にあるこのような表層表現を手がかりとして推定することを試みた。

2 名詞の指示性と数の分類

2.1 名詞の指示性の分類

名詞の指示性とは名詞の対象への指示の仕方である。まず名詞句を、その名詞句の類の成員すべてか類自体を指示対象とする総称名詞句と、類の成員の一部を指示対象とする非総称名詞句に分ける。次に、非総称名詞句を指示対象が確定しているか否かで、定名詞句と不定名詞句に分ける(図1)¹。



総称名詞句 総称名詞句は、その名詞句が意味する類に属する任意の成員(単数でも、複数でも、不可算のもの

¹ 日本語の名詞に対して、この分類と同じような分類をしているものに文献⁽¹⁾が挙げられる。しかし、そこでは指示対象を持たない名詞が考えられている。例えば、「私は大学教師です」の「大学教師」は指示対象を持たないとしてあった。それに対し、本研究では「大学教師」は大学教師という類のある成員と考え「不定名詞」と考える。また、「定」「不定」の区別は聞き手の知識による分類となっており、ここでの分類とは異なる。

でもよい)のすべて、もしくはその名詞句が意味する類それ自身を指示する。例えば、次の文(3)の「犬」は総称名詞句である。

犬は役に立つ動物です。 (3)

ここでの「犬」は「犬」という類に属する成員のすべてを指示対象としている。

定名詞句 定名詞句は、その名詞句が意味する類に属する文脈上唯一の成員(単数でも複数でも不可算のものでよい)を指示する。例えば、次の文(4)の「犬」は定名詞句である。

その犬は役に立ちます。 (4)

ここでの「犬」は、「犬」という類に属する文脈上唯一の成員を指示対象としている。このことは、指示詞「その」によって表わされており、聞き手は「その犬」なるものを確定できる。

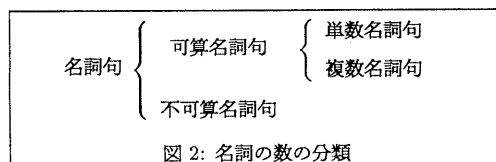
不定名詞句 不定名詞句は、その名詞句が意味する類に属するある不特定の成員(単数でも複数でも不可算のものでよい)を指示する。不特定の成員を指示するというのは、現時点での聞き手の情報ではその名詞句が成員のどれを指し示すのか確定していないという意味である。また、現時点での聞き手の情報では、その名詞句が成員のどれを指し示しているとしても、その文の解釈として間違っていないということでもある。不定名詞句は総称名詞句とは異なり、その名詞句の意味する類の成員のすべてを指示するのではなく、その名詞句の意味する類の成員の一部を指示する。次の文の「犬」は不定名詞句である。

犬が三匹います。 (5)

ここでの「犬」は犬という類に属する任意の三匹の成員を指示対象として持ちえる。これはどんな犬でも三匹いればこの文が使えるということである。

2.2 名詞の数の分類

名詞の数とはその名詞が指示する対象の数のことである。名詞句をその指示対象が数え上げられるか数え上げられないかに応じて、可算名詞句、不可算名詞句に分ける。次に、可算名詞句をその指示する対象が一個か複数個かに応じて単数名詞句、複数名詞句に分ける(図2)。



単数 名詞の指し示す対象が、話者の頭の中で一個のものとして他のものと区別して捉えることができる場合、その名詞の数は単数となる。例えば、次の文の「ケーキ」は単数である。

彼女はケーキを一個持って行きました。(6)

ここでの「ケーキ」は個々に区別して捉えることができ、一個である。

複数 名詞の指し示す対象が、話者の頭の中で個々に区別できるものとして複数個ある場合、その名詞の数は複数となる。例えば、次の文の「ケーキ」は複数である。

この店にはたくさんのケーキがあります。(7)

ここでの「ケーキ」は個々に区別して捉えることができ、またたくさんあるので複数個ある。

不可算 名詞の指し示す対象が、話者の頭の中で個々に区別できないものである場合、名詞の数は不可算となる。例えば、次の文の「銅」は不可算である。

銅はよく熱を伝導します。(8)

ここでの「銅」は個々に区別して捉えることができない。「銅」は「銅」という物質として使われており、不可算である。

3 名詞の指示性と数の推定方法

3.1 『可能性』と『得点』

名詞の指示性と数の推定は、表層の言語表現を手がかりにした規則を異なった種類の表現に応じて必要なだけ作り、入力文に対してそれらを適用することによって行なう。

ある表層表現を手がかりにして、そこにあらわれる名詞がある分類に属さないことがわかる場合がある。例えば、「ある犬」の「犬」は「不定名詞」であって、「総称名詞」「定名詞」になる可能性はないことがわかる。これを表現するために、『可能性』という評価値を導入する。『可能性』が1のときその分類に属する可能性があることを意味し、『可能性』が0のときその分類の可能性がないことを意味する。『可能性』が0となる規則が適用されれば、名詞がその分類に属する可能性はなくなることになる。

ある名詞がある分類に属するかどうかを一つの表層表現から推定するのではなく、複数の表層表現を手がかりに推定を行えば分類の精度はよくなると考えられる。そのため各規則に重要性を表わす『得点』という評価値を導入し、ある名詞に対して適用された規則の『得点』を合計することでその名詞がある分類に属する場合の評価値とする。『可能性』の情報だけではどの分類に属す

るのか一意に推定できない場合に、この評価値はいずれの分類が最も適当であるか推定する基準になる。

(規則の適用条件)

{不定	(可能性 得点)
定	(可能性 得点)
総称	(可能性 得点)}

図 3: 名詞の指示性を推定する規則

(規則の適用条件)

{単数	(可能性 得点)
複数	(可能性 得点)
不可算	(可能性 得点)}

図 4: 名詞の数を推定する規則

規則は指示性の場合に図 3、数の場合に図 4の構造をしている。図の「規則の適用条件」には、その規則が適用されるかどうかの条件として、後で説明する依存構造の表現の形で文中の手がかりとなる表現を記述する。各分類には『可能性』と『得点』を一つずつ与えている。『可能性』は1か0のみであり、『得点』は0から10の間の整数である。

『可能性』が1の分類がただ一つ求まった場合は、その分類を推定の結果とする。『可能性』が1の分類が複数ある場合は、その中で『得点』が最も大きい分類を推定の結果とする。同点の場合は同点の分類すべてを推定の結果とする。

3.2 システムの動作

文中の名詞の指示性と数の推定は次のようなステップで行なわれる。

- (1) 与えられた文の形態素解析、構文解析²が行なわれ、依存構造の表現に変換される。この依存構造の表現は図5のような形式のものであり、その例を図6に示す。
- (2) 依存構造の表現に変換された文の名詞を文頭から順に推定する³。各名詞に対しては指示性を先に数を後に推定する⁴。指示性の推定は、指示性の規則をすべて用いて各分類の『可能性』と『得点』を計算する。この『可能性』と『得点』から3.1節で述べたように指示性を推定する。数の推定も同様である。
- (3) 上記(2)で得られる推定の結果を図7に示す形で出力する。

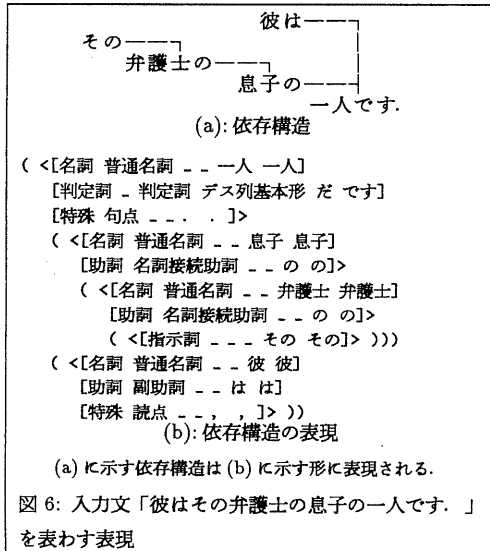
² 形態素解析、構文解析は参考文献⁽²⁾⁽³⁾のものをを用いた。

³ このため、既に推定された指示性と数は後に出てくる名詞の解析の時に手がかりとして使用できる(例: 4.1節の具体例(c)(d))。

⁴ このため、数の推定には指示性の解析結果を用いることができる(例: 4.2節の規則の例の3)。

依存構造の表現 ::= (文節 依存構造の表現
 依存構造の表現 ...)
 文節 ::= < 単語 単語 ... >
 単語 ::= [品詞 品詞細分類
 活用型 活用形 基本形 変化形]

図 5: 依存構造の表現の基本要素の形式



システムでは文章ごとに解析しており、文章全体の表層表現を利用できるようにしている。これは同一名詞が既出のとき適用される規則に用いられる。

3.3 解析の対象から除外した名詞

時間を表わす名詞、「中」「上」「左」「右」「下」「後」「前」「近く」「遠く」「別」「他」のような名詞、「本当の」「普通の」「役に立つ」などの連体詞の一部とみなせる名詞は対象から除外した。「～のまま」「～する程」,「～する訳でない」,「～する度」の下線部に当たる名詞なども除外した。「大部」「全員」「一緒に」などの副詞とみなせるものも除外した。以上の名詞以外は、対訳の英語文では名詞に訳されていない場合でもすべて解析の対象とした。

4 推定に用いる規則

規則は日本語、英語の文法書⁽⁴⁾⁽⁵⁾を参考として作ったが、実験対象テキストを見て独自に考えて作ったものもある。実験中に新たな規則を随時追加していったが、現時点ですべてを網羅できているとはいえない。現在の規則の数は、指示性が 86 個で、数が 48 個である。次に規則の例をあげる。

(< [名詞 普通名詞 -- 一人 一人 不定 単数]
 [判定詞 判定詞 デス列基本形 だ です]
 [特殊 句点 -- .] >
 (< [名詞 普通名詞 -- 息子 息子 定 複数]
 [助詞 名詞接続助詞 -- の の] >
 (< [名詞 普通名詞 -- 弁護士 弁護士 定 単数]
 [助詞 名詞接続助詞 -- の の] >
 (< [指示詞 -- その その] >)))
 (< [名詞 普通名詞 -- 彼 彼 定 単数]
 [助詞 副助詞 -- は は]
 [特殊 読点 -- , ,] >))

図 7: 図 6 の文に対する指示性と数の判定結果の表現

4.1 指示性の規則

1. 名詞が指示詞（「この」や「その」など）によって修飾される時、
 { 不定名詞 (0 0) 定名詞 (1 2) 総称名詞 (0 0) }⁵
 (例文) この本 はおもしろい。
 (訳文) This book is interesting.
2. 名詞につく助詞が「は」で述語が過去形の時、
 { 不定名詞 (1 0) 定名詞 (1 3) 総称名詞 (1 1) }
 (例文) 犬 は向うに行きました。
 (訳文) The dog went away.
3. 名詞につく助詞が「は」で述語が現在形の時、
 { 不定名詞 (1 0) 定名詞 (1 2) 総称名詞 (1 3) }
 (例文) 犬 は役に立つ動物です。
 (訳文) Dogs are useful animals.
4. 名詞につく助詞が「へ」「まで」「から」の時、
 { 不定名詞 (1 0) 定名詞 (1 2) 総称名詞 (1 0) }
 (例文) 彼を 空港 まで迎えに行きましょう。
 (訳文) Let us go to meet him at the airport.
 他にも、(i) 「地球」「宇宙」のような名詞自身から推定する規則、(ii) 名詞に数詞がかかることから推定する規則、(iii) 同一名詞の既出により推定する規則、(iv) 「いつも」「昔は」「～では」のような副詞が動詞にかかることから推定する規則、(v) 「～が好き」「～を楽しむ」のような動詞から推定する規則などがある。手が必要となる語がない時は不定名詞と推定するようにしている。

例として、次の文の中に現れる名詞「果物」に注目し、これにどのような規則が適用され得点がどのようになるか、具体的に説明する。

⁵ 各分類の可能性と得点を表わす。図 3 参照。

⁶ 主語が総称名詞になる場合であるので「a dog」でも「the dog」でもよい。

我々が昨日摘みとった果物は味がいいです。

The fruit that we picked yesterday tastes delicious.

以下のように七つの規則が適用され、この「果物」は定名詞と推定された。

- (a) 名詞につく助詞が「は」で述語が現在形の時、
(果物は味がいいです。)⁷
{ 不定名詞 (1 0) 定名詞 (1 2) 総称名詞 (1 3) }
- (b) 名詞に係る節の述部が過去形の時、(摘みとった)
{ 不定名詞 (1 0) 定名詞 (1 1) 総称名詞 (1 0) }
- (c) 名詞に係る節に「は」か「が」がついた定名詞がある時、(我々が)
{ 不定名詞 (1 0) 定名詞 (1 1) 総称名詞 (1 0) }
- (d) 名詞に係る節に助詞がついた定名詞がある時、
(我々が)
{ 不定名詞 (1 0) 定名詞 (1 1) 総称名詞 (1 0) }
- (e) 名詞に係る節に代名詞がある時、(我々が)
{ 不定名詞 (1 0) 定名詞 (1 1) 総称名詞 (1 0) }
- (f) 名詞につく助詞が「は」で述語が形容詞の時、
(果物は味がいいです。)
{ 不定名詞 (1 0) 定名詞 (1 3) 総称名詞 (1 4) }
- (g) 普通名詞の時、(果物)
{ 不定名詞 (1 1) 定名詞 (1 0) 総称名詞 (1 0) }

これらすべての規則の適用の結果として「果物」の最終の『可能性』と『得点』は、

{ 不定名詞 (1 1) 定名詞 (1 9) 総称名詞 (1 7) }

となり、定名詞と推定された。

4.2 数の規則

1. 「その」「この」「あの」によって修飾される時、
{ 単数 (1 3) 複数 (1 0) 不可算 (1 1) }
(例文) あの本をください。
(訳文) Give me that book.
2. 名詞につく助詞が「は」「が」「も」「を」で、述部に数詞に係る場合、
数詞の数が単数の時、
{ 単数 (1 2) 複数 (1 0) 不可算 (1 0) }
数詞の数が複数の時、
{ 単数 (1 0) 複数 (1 2) 不可算 (1 0) }
(例文) りんごを二個食べる。
(訳文) I eat two apples.
3. 名詞の指示性が総称名詞と推定されており、係り先の動詞が「が好きです」「を楽しむ」などのように、総称の名詞を格にとる場合、

⁷ 規則が適用される手がかりとなる表現。

{ 単数 (1 0) 複数 (1 2) 不可算 (1 0) }

(例文) 私は、りんごが好きです。

(訳文) I like apples.

他にも、(i) 数詞に係ることにより推定する規則、

(ii) 「集める」「溢れる」などの動詞から推定する規則、

(iii) 「いくらでも」「何度でも」のような副詞が動詞にかかることから推定する規則などがある。手がかりとなる語がない時は単数と推定するようにしている。

5 実験と考察

5.1 実験の準備

実験に用いたテキストは名詞の指示性と数の正解の分類がわかりやすいように日英の対訳がある文章に限った。実験対象のテキストの各名詞に対してあらかじめ正解の分類を手で決定した。正解の決定の際には対訳の英語文を見て行なったが、必ずしも冠詞にとらわれることなく2節で説明した分類の定義によって正解を決定した。指示性の分類については、総称名詞の判定は極めて困難であり、表1のようなものを総称名詞としたが、正解が間違っている可能性がある。以下、正解とはこの人手による分類のことをいう。

表 1: 総称名詞句とした名詞の例 (下線部の名詞)

(1) <u>ラクダ</u> は <u>水</u> を飲まなくても長い間歩くことができます。
(2) ワシントンスクールからークラスの学生たちが、昨日、 <u>見学</u> にいきました。
(3) 多くの若い <u>男</u> の <u>人</u> たちは <u>陸軍</u> に兵役します。
(4) <u>紳士</u> は普通 <u>淑女</u> のために <u>ドア</u> を開けます。
(5) 有名なシャーロックホームズ探偵の物語は大抵ロンドン地域を <u>背景</u> にしたものです。
(6) 彼はクリスマスの <u>贈り物</u> に本を買いました。
(7) ワールドカップ大会の決勝戦は、 <u>タンゴ</u> のアルゼンチンと <u>行進曲</u> の西ドイツとの勝負だといった人がいた。

数の分類については、対訳の英語文で名詞に訳されているものは冠詞に合わせて正解を設定した。明らかに複数とか不可算とわかるものはそのように設定し、それ以外は単数とした。総称名詞句が主語に来る場合は冠詞が何になるのかわからないので、数の分類は単数でも複数でも不可算でもいいことにして正解は「未定」⁸とした。

⁸ 各分類の得点が同点の時のみ正解とする分類。

表 2: 指示性の正解率 (学習サンプル)

評価	不定	定	総称	その他	総数
英語冠詞用法辞典					
正解	96	184	58	1	339
正解を含む	0	3	1	0	4
部分解	0	0	0	0	0
不正解	4	25	7	1	37
正解率 (%)	96.0	86.8	87.9	50.0	89.2
こぶとりじいさん					
正解	73	140	6	1	222
正解を含む	3	4	0	0	7
部分解	0	0	0	0	0
不正解	11	23	4	0	38
正解率 (%)	83.9	84.0	60.0	100.0	83.2
天声人語					
正解	25	35	16	0	76
正解を含む	0	4	2	0	6
部分解	0	0	0	0	0
不正解	5	10	1	0	16
正解率 (%)	83.3	71.4	84.2	—	77.6
出現率 (%)	29.1	57.7	12.8	0.4	100.0
全正解率 (%)	89.4	84.0	84.2	66.7	85.5

表 4: 指示性の正解率 (テストサンプル)

評価	不定	定	総称	その他	総数
つるのおんがえし					
正解	109	363	13	10	495
正解を含む	6	25	0	0	31
部分解	0	0	0	0	0
不正解	32	135	6	0	173
正解率 (%)	74.2	69.4	68.4	100.0	70.8
天声人語					
正解	75	81	16	0	172
正解を含む	8	9	1	0	18
部分解	0	0	0	0	0
不正解	33	51	9	0	93
正解率 (%)	64.7	57.5	61.5	—	60.8
冷戦後世界と太平洋アジア					
正解	21	108	11	2	142
正解を含む	6	7	0	0	13
部分解	0	0	0	0	0
不正解	11	24	2	0	37
正解率 (%)	55.3	77.7	84.6	100.0	74.0
出現率 (%)	25.6	68.4	4.9	1.0	100.0
全正解率 (%)	68.1	68.7	69.0	100.0	68.9

表 3: 数の正解率 (学習サンプル)

評価	単数	複数	不可算	その他	総数
英語冠詞用法辞典					
正解	274	32	18	25	349
正解を含む	1	1	1	0	3
部分解	0	0	0	11	11
不正解	3	10	0	4	17
正解率 (%)	98.6	74.4	94.7	62.5	91.8
こぶとりじいさん					
正解	205	24	5	0	234
正解を含む	2	0	0	0	2
部分解	0	0	0	7	7
不正解	1	22	1	0	24
正解率 (%)	98.7	52.2	83.3	0.0	87.6
天声人語					
正解	64	13	0	3	80
正解を含む	2	1	0	0	3
部分解	0	0	0	6	6
不正解	1	6	1	1	9
正解率 (%)	95.5	65.0	0.0	30.0	81.6
出現率 (%)	74.2	14.6	3.5	7.7	100.0
全正解率 (%)	98.2	63.3	88.5	49.1	89.0

表 5: 数の正解率 (テストサンプル)

評価	単数	複数	不可算	その他	総数
つるのおんがえし					
正解	610	13	1	1	625
正解を含む	12	2	0	0	14
部分解	0	0	0	1	1
不正解	2	20	37	0	59
正解率 (%)	97.8	37.1	2.6	50.0	89.4
天声人語					
正解	197	13	2	3	215
正解を含む	3	1	0	0	4
部分解	0	0	0	3	3
不正解	3	55	3	0	61
正解率 (%)	97.0	18.8	40.0	50.0	76.0
冷戦後世界と太平洋アジア					
正解	157	6	1	1	165
正解を含む	3	0	0	0	3
部分解	0	0	0	0	0
不正解	3	20	1	0	24
正解率 (%)	96.3	23.1	50.0	100.0	85.9
出現率 (%)	84.3	11.1	3.8	0.8	100.0
全正解率 (%)	97.4	24.6	8.9	55.6	85.6

5.2 実験

3.2節で述べたように指示性と数の推定の前に形態素・構文解析を行なうが、そこでの誤りは手で修正した。まず、三つの資料 { 英語冠詞用法辞典⁽⁴⁾ の典型的な用法の例文 (140 文, 解析した名詞 380 個), 物語の「こぶとりじいさん」⁽⁶⁾ 全文 (104 文, 解析した名詞 267 個), 86 年 7 月 1 日の天声人語 (23 文, 解析した名詞 98 個) } を学習サンプルとして実験を行なった。システムには正解を入力して自動的に正解率を出力できるようにして、正解率が向上するように規則を変更、追加した。最も正解率が良くなった規則の時の正解率を表 2, 表 3 に示す。

しかし、以上の実験では新しい文での正解率がわからない。そこで、以上のようにして作った規則を固定して、新たな三つの資料 { 物語の「つるのおんがえし」⁽⁶⁾ 全文 (263 文, 解析した名詞 699 個), 86 年 7 月 8, 9, 15 日の天声人語の三回分 (75 文, 解析した名詞 283 個), 冷戦後世界と太平洋アジア (国際文化会館会報 Vol.3 No.2 1992 年 4 月号) (22 文, 解析した名詞 192 個) } をテストサンプルとして実験を行なった。これらの正解率を表 4, 表 5 に示す。

表中の「正解」は推定の結果が正解と一致した場合である。「正解を含む」は推定の結果の中に正解の分類がある場合である。例えば、正解が定名詞で推定の結果が定名詞と不定名詞が同点で得られた場合、「正解を含む」となる。「部分解」は推定の結果が正解に含まれる場合である。「不正解」は以上の評価以外のものである。「正解率」は「正解」の個数を総数で割ったものである。「全正解率」は三つの資料全てにおける正解率である。「出現率」は各分類の個数を総数で割ったものである。「その他」は、単数と複数と不可算の得点が同点の時のみ正解とする「未定」のように、複数個の分類が正解になるものの個数である。

5.3 考察

5.3.1 指示性的実験に対する考察

正解率が上がるようにテキストの表現に対して規則を変更して実験した学習サンプルのテキスト全体での正解率は 85.5% であった。また、各分類に対する正解率も極端に悪いものはない。このことから表層表現を手がかりとした我々の方法で極めて多くの名詞の指示性が推定できることがわかった。

規則を固定して実験したテストサンプルのテキスト全体での正解率は、68.9% であった。また、各分類に対する正解率もほぼ均等に良くすべて 50% 以上である。つ

表 6: 解析を失敗した定名詞の例 (下線部の名詞)

(1) <u>息子</u> はアインシュタインのような天才ではないかも知れない。
(2) 彼は <u>社長</u> の兄さんです。
(3) ジョンは <u>クラス</u> の中で一番背が高い。
(4) 彼女は <u>テーブル</u> のほこりを取り除くためにふきんを使いました。
(5) <u>仕事</u> で難しい所がありましたが、克服しました。
(6) <u>お腹</u> が痛いです。
(7) 私は <u>先生</u> と同じ本を持っています。
(8) 車は <u>道</u> の <u>わき</u> に駐車してあります。
(9) ジョンソン教授は <u>学会</u> で <u>論文</u> を読みました。

表 7: 解析を失敗した総称名詞の例 (下線部の名詞)

(1) 修飾節で限定され定名詞になる例 それ自体を守ろうとしない <u>文化</u> は滅びます。
(2) 述語が過去形のために定名詞になる例 <u>中国人</u> は独自の文字を発明しました。
(3) 判定詞「だ」がつくために不定名詞になる例 日本の社会では父親は <u>家長</u> です。
(4) 手がかりがなく不定名詞になる例 <u>食物</u> がおいしければおいしいほど、たくさん食べます。 <u>普通肺炎</u> にかかると入院しなければなりません。

るのおんがえしの実験では、定名詞の出現率が 74.8% であったのですべての解析結果を定名詞にする規則を作ると正解率が 74.8% になり実験の正解率の 70.8% より高くなるが、不定名詞と総称名詞の正解率が 0% になるので意味がない。われわれはそれぞれの分類が均等に良い正解率が得られることに価値があると考えている。

テストサンプルのテキストでの正解率はあまり良くないが、規則を修正すれば容易に上がると考えられる。しかし、新しい文章に対する正解率を上げようとするどこまでも規則を増やしていかなければならないという危険性がある。

規則を変更しても解析が失敗する例として、表 6, 表 7 がある。表 6 は定名詞であるのに、解析が失敗したものである。例 (1) の「息子」は名詞につく助詞が「は」で述語が現在形なので総称名詞と解析された。しかし、人間が見ると明らかに定名詞である。他の例も定名詞となる手がかりがなく解析が失敗した。これを解決するには文脈や発話状況などの表層表現以外の情報が必要である。

表 7 は総称名詞であるのに、解析を失敗したものである。それぞれの例に対して失敗した理由を付けている。総称名詞は判断が難しく、成功はしているが正解がっているのか不確かな例もある。

ラクダは水を飲まなくても長い間歩けます。(9)
という文の「ラクダ」は明らかに総称名詞であるが、「水」は総称名詞でいいのだろうか。総称名詞は他と明確に区別して考えにくく種々の性質のものがありそうなので、新たに分類を考え直す必要はないだろう。

これ以外の成功した例でも深く考察すると表層表現で推定できるのか疑問なものがある。

これは私が彼から借りた辞書です。(10)
この文の「辞書」は修飾節により限定されていて定名詞と解析される。しかし、この「辞書」は、「私」が「彼」から複数の辞書を借りており、そのうちの一つを指す場合には不定名詞となる。つまり、ある程度の知識がなければ定名詞か不定名詞かの判断ができない。

5.3.2 数の実験に対する考察

正解率が上がるようにテキストの表現に対して規則を変更して実験した学習サンプルとしてのテキスト全体での正解率は、89.0%であった。しかし、「複数」の正解率が全体的に悪い。また、「不可算」は不可算名詞としてあらかじめ登録してあるので見かけ上正解率が高いが、問題がないわけではない。

規則を固定して実験したテストサンプルとしてのテキスト全体での正解率は、85.6%であった。しかし、「複数」「不可算」の正解率は悪く、「単数」の出現率と正解率が高いおかげで高い正解率が出ているにすぎない。また、その「単数」の正解率が高いのも、手がかりのないときは「単数」と推定させているからにすぎない。

正解が「複数」である名詞で、解析が失敗した例に以下の下線部の名詞がある。

あなたが注文した建築材料がきました。(11)
手がかりがなく、解析結果は「単数」となってしまう。これを「複数」と判定できるようにするには、「建築材料」という名詞自身から「複数」と判定できるようにしなければならない。しかし、いつでも「建築材料」が「複数」とは限らない。

数量表現以外から「複数」とわかった例として以下の下線部の名詞がある。

その事故が発生してから野次馬が集まってきました。(12)
「が集まる」から「野次馬」が「複数」と解析される。

このような規則を作っていけば、数が明示されていなくても名詞の数がわかる場合がある。しかし、以下の例のように間違える場合もある。

シュートを浴びたゴールキーパーが、右のてのひらを裂いた。(13)

「を浴びる」の規則は実験では「複数」と「不可算」が強いとしていたが、一発のシュートで手を裂いたので、上のシュートは「単数」である。動詞からの推定もそう安易なものではない。

6 おわりに

学習サンプルでの正解率は、指示性で 85.5%、数で 89.0% であり、テストサンプルでの正解率は、指示性で 68.9%、数で 85.6% であった。

指示性の推定における課題としては、次の二つのことが残っている。一つは、人間が見ると状況から定名詞であることが明らかであるのに推定できていない場合である。状況の情報をうまく使えるようになれば、推定できるようになる。しかし、このときも知識だけでなく表層表現と知識の連携が必要であろう。もう一つは、総称名詞に関することである。総称名詞は他の分類とはっきりと区別して定義することが難しいという性質をもっており、まだまだ考えてゆかなければならない問題である。

一方、数の方は数量詞のような表層表現があれば推定できるが、そのような表現がなければ推定できない。しかし、「(名詞)を集める」ならば「複数」とするといった常識的な知識を規則に取り込んでいけば、現時点で推定できていないものも少しは推定していくことができると考えられる。

参考文献

- (1) 金水敏, 連体修飾成分の機能, 松村明教授古希記念国語研究論文集, (明治書院, 1986), pp. 602-624.
- (2) 松本, 黒橋, 宇津呂, 妙木, 長尾, 日本語形態素解析システム JUMAN 使用説明書 version 1.0, (京都大学工学部 長尾研究室, 1992).
- (3) 黒橋, 長尾, 格構造解析への評価関数の導入による統語的曖昧性の解消, 自然言語処理研究会 92-NL-92, (1992), pp. 65-72.
- (4) 熊山晶久, 英語冠詞用法辞典, (大修館書店, 1985).
- (5) 池内正幸, 名詞句の限定表現, 新英文法選書, 第 6 巻, (大修館書店, 1985).
- (6) 中尾清秋, こぶとりじいさん 他 鶴の恩がえし, きき耳ずきん, 英訳「日本むかしばなし」シリーズ, 第 7 巻, (日本英語教育協会, 1985).