

共起関係を利用した日本語複合名詞の分割

内山将夫 板橋秀一

筑波大学

複合名詞を表す平仮名文字列を単語の列に変換する方法を述べる。変換の際に生じる誤りとしては、「選択誤り」と「分割誤り」、及び、それらが複合した誤りがある。これらの誤りは複合名詞中に現れる同音語を特定することにより解決できる。単独では同音語となる単語も、直後の単語を含めて考えれば同音語とならないことが多い。しかし、単語とそれに続く単語をすべて辞書に登録すると、辞書がかなり大きくなってしまう。そのため、単語にマークをつけ、マーク間に対応をつけることにより、単語の対に登録することと同様な効果を得た。これにより、辞書効率の向上、未学習の単語の対に対しての簡単な推論ができることを示す。

Division of Japanese compound nouns using co-occurrence relation

Masao Utiyama Shyuichi Itahashi

University of Tsukuba

1-1 Tennoudai 1-chome, Tukuba City, Ibaraki 305, Japan

This paper discusses a method of transforming hiragana sequences expressing a compound noun into component nouns. Errors which occurs in the transformation are "selection errors", "division errors", and their compound errors. These errors can be resolved by discriminating homonyms in the compound noun. Most homonyms including the following word will become nonhomonyms. But if we enroll a word with all successors in a dictionary, it will become quite large. So we mark words. The correspondence between the marks has the same effect as enrolling the word pair. This way makes it possible to improve the efficiency of a dictionary, and to predict an unenrolled pair.

1 はじめに

複合名詞の解析の研究には、その立場により様々なものがある。従来の複合名詞の解析法を解析のレベル、解析手法、入力される文字列の点から整理すると以下のようになるであろう。以下で、括弧内は文献番号を示す。

解析のレベル

1. 複合名詞を単語に分割する [4] [5] [16]
2. 複合名詞の構造、意味を推定する [6] [7] [8] [9] [11] [13] [14] [17]

解析手法

1. 単語をクラス分けして、それに関する文法規則を設定し、解析を行なう [4] [7] [8] [9] [11] [14]
2. 統計的手法により造語モデルを作り、そのモデルに基づいて解析を行なう [5] [6]
3. 複合名詞を効率的に登録し、それを元にして解析を行なう [13] [17]
4. ヒューリスティックスに基づいて解析を行なう [16]

入力文字列

1. 複合名詞の読みである平仮名文字列 [7] [8] [9] [11] [13] [14] [17]
2. 複合名詞自体 [4] [5] [6] [16]

本稿は応用として、かな漢字変換を念頭においている。また、複合名詞解析の際の条件として、

1. 学習の際に人手がかからないこと。
2. 学習の結果が迅速に解析結果に反映すること。
3. 不十分な学習でも、ある程度の結果が得られること。

を設定した。そのため、解析のレベルとして複合名詞の分割を設定し、解析手法として複合名詞の登録を取り入れ、入力文字列は平仮名文字列とした。

2 同音語の識別

単独では同音語となるような語でも、次の語までを含めると同音語とならない場合が多い。語とそれに続く語を登録することにより、通常の文章においては95%弱のものについて、それを構成する語が一意に特定できることが報告されている [2]。複合名詞は文章よりも構造と構成物が限られてくるため、より高精度に特定できるものと考えられる。

しかし、単語(複合名詞を構成する単位の語を単語ということにする)とその後に続く単語群をすべて登録すると、複合名詞を直接登録するよりは効率的であるが、辞書がかなり大きくなってしまふ。そこで何らかの工夫が必要とされる。一つの方法として単語間にポインタを張る方法が挙げられる [15]。本稿では単語にマークを設定し、マーク間に対応をつける。これにより単語の対を登録することと同様な効果を得るとともに、簡単な推論ができることを示す。そのためにまず、単語の役割を二つに分ける。一つは係り元、つまり、対で登録されるべき単語の組の最初の単語としての役割であり、もう一つは係り先、つまり、単語の組の二番目のものとしての役割である。そのため、単語のデータ構造を次のように表現する。

単語のデータ構造 =

(読み 書き 係り元素性リスト 係り先素性リスト)

素性リスト = (x y z ...)

(x, y, z, ... は自然数)

(単に素性リストという場合は「係り元素性リスト」と「係り先素性リスト」の両方を指す。)

そして、ある単語の係り元素性リストと別の単語の係り先素性リストに次の対応があるとき、それらの単語は接続関係にあるということにする。

接続関係

単語 A の係り元素性リストを $M(A)$ 、係り先素性リストを $S(A)$ で表す。 $M(A)$ と $S(A)$ が次の関係のとき、単語 A と単語 B は接続関係にあるという。

$$M(A) = (m_1 m_2 \dots m_j)$$
$$S(A) = (s_1 s_2 \dots s_k) \quad \text{とすると、} (s_i, m_i \text{ は自然数})$$
$$(j > 0) \wedge (k > 0)$$
$$\wedge (0 < i \leq \min(j, k) \text{ について、} m_i \text{ と } s_i \text{ が同値})$$

同値: $(m_i = s_i) \vee (m_i \neq s_i \text{ だが、} m_i \text{ と } s_i \text{ が同値であると登録されている})$

つまり、単語の対を登録する代わりに、単語にマークをセットし、単語間のマークの対応により単語の対が接続関係にあることを示すのである。これにより、次のような簡単な推論ができる。

1. 係り元素性リストを持ち、係り先素性リストを持たないものは接頭語、係り先素性リストを持ち、係り元素性リストを持たないものは接尾語と考えることができる。これを利用して、単語間に接続関係がないような場合でも、複合名詞を構成する単語の特定において、ある程度の判断ができる。

例

辞書 = {(さい 再 (1) ())

(こうせい 構成 (1) ())

(さいこう 最高 (1) ())

(せい 生 (1) ())}

とする。

「さいこうせい」を分割することを考える。右方向最長一致法によると「最高 | 生」だが、

- 「再」が係り要素性リストしか持たないことから接頭語と判断できること、
- 「再」は素性リストを持つが「最高」は素性リストを持たないことから、「再」の方が使われるだろうということ、

などから、「再 | 構成」と判断できる。(ただし、縦棒 (|) は単語の区切れを示している。)

2. 接統関係が未登録の単語間の接統関係が判断できることがある

例

辞書 = {(ろく 六 (1) ())

(しち 七 (1) ())

(はち 八 (1) ())

(ねん 年 (1) ())

(がつ 月 (1) ())

(にち 日 (1) ())

(はち 鉢 (1) ())}

とする。

「六 | 年」、「六 | 月」、「七 | 月」、「七 | 日」、「八 | 日」を、この順番で学習することにより、単語の状態は変化する。

辞書 = {(ろく 六 (1) ())

(しち 七 (1) ())

(はち 八 (1) ())

(ねん 年 (1) ())

(がつ 月 (1) ())

(にち 日 (1) ())

(はち 鉢 (1) ())}

となる。

その結果、「はちねん」を分割する時に、「八」と「年」が接統関係にあり、「鉢」と「年」は接統関係にないので、「鉢 | 年」ではなく「八 | 年」が出力される。¹

3 学習

分割の際に曖昧²な部分があると、正しい複合名詞が第一候補にでないことがある。その際には学習を行なう。学習を受ける単語列は、正しい複合名詞において、曖昧な平仮名列から変換された単語列と、その前後の単語よりなる単語列である。ただし解析が失敗したときには、正しい複合名詞を構成する単語列全体に対して学習を行なう。

例

入力文字列 にせんどひゃくねん

解析結果 二 | 千 | 御 | 百 | 年

二 | 千 | 語 | 百 | 年

二 | 千 | 五 | 百 | 年

¹以上のことは双方向ポインタによっても実現できるが、探索空間が大きくなる。例えば、単語のデータ構造を

(id 読み 書き 係り先の id リスト 係り元の id リスト)

とすると、上例と同じ学習の結果、辞書は

辞書 = {(1 ろく 六 (4 5) ())

(2 しち 七 (5 6) ())

(3 はち 八 (6) ())

(4 ねん 年 (1) ())

(5 がつ 月 (1 2) ())

(6 にち 日 (2 3) ())

(7 はち 鉢 (1) ())}

となる。

この結果からでは、「八」と「年」の結び付きの方が、「鉢」と「年」のそれよりも強いと言うことが一目ではわからない。

²平仮名列から複数の漢字列の候補が出るとき、その平仮名列は曖昧であるとする。曖昧さを生じる原因として、分割の曖昧さと選択の曖昧さがある。

分割の曖昧さ さいこうせい → 最高 | 性、再 | 構成、...

選択の曖昧さ こうかんき → 交換 | 機、交換 | 木、交換 | 器、...

これらは複合して起こることもある。

の場合は曖昧な平仮名文字列「ご」の正しい変換である「五」と、その前後の単語からなる「千 | 五 | 百」について学習を行なう。

学習は以下の手順で行なう。

正しい複合名詞における部分単語列 $a(1) \dots a(n)$ が与えられたとする。そして、単語 $a(i)$ の係り元素性リスト、及び、単語 $a(i+1)$ の係り先素性リストを次のようなものとする。

$$M(a(i)) = (m_1 m_2 \dots m_p)$$

$$S(a(i+1)) = (s_1 s_2 \dots s_q)$$

段階 1.

$a(i)$ と $a(i+1)$ ($0 < i < n$) が 接続関係にないときは、次の操作によって、それらが接続関係にあるようにする。

1. $p = q = 0$ のとき

m_1 と s_1 に同一の新しい素性(今まで設定された素性と重複しない自然数のうちで最小のもの)を設定する。

2. $p = 0 \wedge q \neq 0$ のとき m_1 として s_1 を設定する。

3. $p \neq 0 \wedge q = 0$ のとき s_1 として m_1 を設定する。

4. その他

$0 < j \leq \min(p, q)$ について、 m_j と s_j が同値でないときは、同値であると登録する。

これだけの情報により、正しい複合名詞が第一候補となるとときには“段階 2”を行なわない。

段階 2.

$a(i)$ と $a(i+1)$ の接続関係の強さを増す。ただし、 $a(i)$ と $a(i+1)$ の接続関係の強さを次のように定義する。

単語 $a(i)$ と単語 $a(i+1)$ の接続関係の強さ
 $= M(a(i))$ と $S(a(i+1))$ において同値な素性の数。

1. $p = q$ のとき

m_{p+1} と s_{q+1} に同一の新しい素性を設定する。

2. $p \neq q$ のとき

³学習を行なったとすると

$p > q \Rightarrow s_{q+1}$ として m_{q+1} を設定する。

$p < q \Rightarrow m_{p+1}$ として s_{p+1} を設定する。

段階 3. 曖昧さが除去できるまで“段階 2”を繰り返す。

学習される単語列は、解析結果の複合名詞の列(以下では単に複合名詞列ということにする)から決定されるが、その複合名詞列からは期待した複合名詞ではないが文法的に誤りといえないものを除くことにする。また、複合名詞の正しいかの判別は、候補である複合名詞に対して、それが期待されたものか(1)、期待されたものではないが文法的に正しいものか(0)、全くの誤りか(-1)、をユーザに聞くことによって判断する。

例

「かがくじっけん」を解析した際に出力された複合名詞の列が、

((かがく 科学 () ()) (じっけん 実験 () ()) -1
 ((かがく 科学 () ()) (じっけん 実験 () ()) 0
 ((かがく 化学 () ()) (じっけん 実験 () ()) -1
 ((かがく 化学 () ()) (じっけん 実験 () ()) -1

であった場合、「科学実験」を除いたものについて学習を行なう。すると

((かがく 化学 (1) ()) (じっけん 実験 () (1)))

のようになる。

再び「かがくじっけん」を解析するとする。今度は「科学実験」が期待されているものであったとする。解析結果は以下のようになる。

((かがく 化学 (1) ()) (じっけん 実験 () (1))) 0
 失敗

これは、後で述べるように、解析において枝刈りを行なっているためである。解析が失敗したので学習を行なうと、次のようになる。

((かがく 科学 (1) ()) (じっけん 実験 () (1)))

再び「かがくじっけん」を解析する。そのとき、「科学実験」を期待していたにもかかわらず解析結果が

((かがく 化学 (1) ()) (じっけん 実験 () (1))) 0
 ((かがく 科学 (1) ()) (じっけん 実験 () (1))) 1

であるとする。このとき「化学実験」はユーザの答が“0”であるので、複合名詞列から取り除く。この結果、「科学実験」しか複合名詞列に残らなくなる。従って、学習は行なわれない。これにより、リストが際限なく伸びるを防ぐことができる。³

この例で、同音異義語である「科学」と「化学」が同一の係り元素性リストを持つようになってしまった

が、これは適当な例（化学変化、科学革命等）により分離できる。以下に、一連の学習例を示す。

学習例

入力文字列；出力列（最後尾の複合名詞が正解である）登録されている単語の集合をTで表す。

出力列中で、★がついている複合名詞は、単語間の接続関係の強さが他の複合名詞より強いため、解析の結果、必ず最初にその複合名詞が出力されることを示す。

T = {自由 群 軍 参加 生成 単位 解放}
解析) じゅうぐんさんか；自由 | 群 | 参加、
自由 | 軍 | 参加

学習)

段階 1

(じゅう 自由 () ()) → (じゅう 自由 (1) ())
(ぐん 軍 () ()) → (ぐん 軍 (2) (1))
(さんか 参加 () ()) → (さんか 参加 () (2))

解析) じゅうぐんせいせい；★自由 | 軍 | 生成、
自由 | 群 | 生成

学習)

段階 1

(じゅう 自由 (1) ()) → (じゅう 自由 (1) ())
(ぐん 群 () ()) → (ぐん 群 (3) (1))
(せいせい 生成 () ()) → (せいせい 生成 () (3))

解析) じゅうぐんせいせい；★自由 | 群 | 生成

解析) じゅうぐんさんか；★自由 | 軍 | 参加

「群」と「軍」の係り先素性リストが同一になった。

解析) たんいぐん；単位 | 軍、単位 | 群

学習)

段階 1

(たんい 単位 () ()) → (たんい 単位 (1) ())
(ぐん 群 (3) (1)) → (ぐん 群 (3) (1))

段階 2

(たんい 単位 (1) ()) → (たんい 単位 (1 4) ())
(ぐん 群 (3) (1)) → (ぐん 群 (3) (1 4))

解析) かいほうぐん；解放 | 群、解放 | 軍

学習)

段階 1

((かがく 科学 (1 2) ()) (じっけん 実験 () (1 2)))

のようになる。

次に「かがくじっけん」を解析したときには、「化学実験」が期待したものであるとすると、解析は失敗する。従って学習を行なうと、

((かがく 化学 (1 2 3) ()) (じっけん 実験 () (1 2 3)))

となる。この様に、共に意味的に正しいペアを区別するためだけにもかかわらず、リストが際限なく伸びる可能性がある。

⁴ 双方向ポインタでは「単位」と「軍」、および、「解放」と「群」が接続関係にないことはわからない。

(かいほう 解放 () ()) → (かいほう 解放 (1) ())
(ぐん 軍 (2) (1)) → (ぐん 軍 (2) (1))

段階 2

(かいほう 解放 (1) ()) → (かいほう 解放 (1 5) ())
(ぐん 軍 (2) (1)) → (ぐん 軍 (2) (1 5))

解析) たんいぐん；★単位 | 群

解析) かいほうぐん；★解放 | 軍

以上の学習により、同一の係り先素性リストを持っていた「群」と「軍」の係り先素性リストが区別された。⁴

4 解析

解析は深さ優先探索で行なう。その際、もっとも得点の高い枝以外は刈り取る。

4.1 候補の選び方

解析途中の文字列は一般に次の形をしている。

決定済みの単語列 候補 残り文字列

枝刈りはまず、score1 について最も得点の高い候補以外は捨てることにより行なう。続いて score2, score3 についても同様な処理をする、というように三段階に分けておこなう。得点の付け方は以下の通りである。

score1 決定済みの単語列と候補を見比べて、候補に得点をつける。

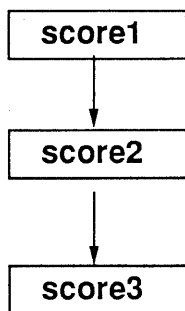
score1 = ((U 決定済みの各単語の係り元素性リスト) ∩ (候補の係り先素性リスト)) の要素数

score2 決定済みの単語列に候補を接続した単語列と残り文字列を見比べて得点をつける。

score2 = max (((U 決定済みの各単語及び候補の係り元素性リスト) ∩ (残り文字列の先頭に一致する単語の係り先素性リスト)) の要素の数)

score3 候補自身を見て得点をつける

score3 = (候補の読みの長さ) / (候補の読みの長さ + 残り文字列の長さ)



4.2 例外処理

- score1 に関する枝刈りで、残り文字列の先頭からの読みに適した単語が選べないような候補は得点に関わらず除去する。
- 残り文字列がないときは score2 に関する枝刈りを行なわない。
- score3 による枝刈りで、語頭のときは係り元素性リストのみをもつ単語、語尾のときは係り先素性リストのみを持つ単語については、得点が低くても残す。(接辞に関する処理)

残った候補の一つを選び、決定済みの単語列に付け加えて、解析を続行する。

score1, score2, score3 の総合点により枝刈りを行なうことも考えられるが、得点の正規化、及び例外処理に対する得点の付け方が不明なため行なわなかった。

5 実験結果

文献 [18] の目次に登録されている複合名詞について、約 1 万語の辞書（カタカナ、及び、1、2 文字の漢字からなる辞書）を元にして、実験を行なった。

比較対象として、ポインタを単語間に張る方法（単にポインタ方式ということにする、また、単語にマークを設定する方式をマーク方式ということにする）と右方向最長一致法を選んだ。（共に深さ優先探索である。）次にポインタ方式により複合名詞を分解する方法の要点を示す。

5.1 ポインタ方式の要点

5.1.1 単語の構造

単語のデータ構造は次のようなものである。

(id 読み 書き 係り先の id リスト)

「係り先の id リスト」は、今までに右方向に接続した単語の id を要素として持つ。

5.1.2 学習

ポインタ方式においての学習とは、解析結果から曖昧な部分とその前後の単語を抜き出して、それらの単語間にポインタを張ることをいう。

例

入力文字列	にせんどひゃくねん
解析結果	二 千 御 百 年
	二 千 語 百 年
	二 千 五 百 年

の場合は、「千 | 五 | 百」について学習を行なう。

その結果、辞書は次のようになる。

(1 せん 千 ())	→	(1 せん 千 (2))
(2 ご 五 ())	→	(2 ご 五 (3))
(3 ひゃく 百 ())	→	(3 ひゃく 百 ())

5.1.3 解析

解析は深さ優先探索で行なう。解析途中の文字列は以下の形をしている。

決定済みの単語列 候補 残り文字列

先と同様に、三段階の枝刈りにより候補を選ぶ。

score1 決定済みの単語の「係り先 id のリスト」の和集合の中に候補の id が含まれているような候補に得点 1 を与え、その他の候補には得点 0 を与える。

score2 決定済みの単語と候補の「係り先 id のリスト」の和集合の中に、残り文字列の先頭から一致するような単語の id が含まれているような候補に得点 1 を与え、その他の候補には得点 0 を与える。

score3 候補の読みの長さを候補の読みの長さとし残り文字列の長さを加えたもので割った値を与える。

例外処理については、score3 についての処理を除いて前と同様に行なう。

5.2 実験結果

まず、マーク方式とポインタ方式において、440 の複合名詞を一意に解析できるまで学習させた。

ただし、接辞の特徴づけを考えて、複合名詞の途中に接辞が含まれるようなものは除いた。

次に未学習の 286 の複合名詞について分割を行なった。こんどは、複合名詞の型に制限をつけなかった。

1. マーク方式による結果

成功 263 個

失敗 23 個

成功率 92 %

誤りの原因

原因	個数	例：正解 → 失敗
選択誤り	19	交換 機 → 交換 器
分割誤り	3	利用 者 団体 → 利用 団体 体
学習の影響	1	複製 記録 → 複製 機 録

単語一個当たりの素性の数 0.74 個

素性の種類 108 個

2. ポインタ方式による結果

成功 226 個

失敗 60 個

成功率 79 %

誤りの原因

選択誤り 54 個

分割誤り 3 個

学習の影響 1 個

係り先の id リストの平均の長さ 0.52

3. 右方向最長一致法による結果

成功 159 個

失敗 127 個

成功率 56 %

誤りの原因

選択誤り 120 個

分割誤り 7 個

次に 726 (= 440 + 286) の複合名詞を格納するために必要な辞書の大きさを示す。

- 単語のみを (読み書き係り元素性リスト係り先素性リスト) の形式で登録したもの。(マーク方式)

登録語数	432
メモリ	13KB

- 単語のみを (id 読み書き係り先の id リスト) の形式で登録したもの。(ポインタ方式)

登録語数	432
メモリ	14KB

(id の分だけマーク方式より容量をとっている)

- 解析された複合名詞とそれを構成する単語を (読み書き) の形式で登録したもの。(すべて登録する方法)

登録語数	1158
メモリ	34KB

今回の実験では、マーク方式は複合名詞をすべて辞書に登録する方式に比べて、辞書の大きさが半分以下になった。

6 考察

- 学習済みの複合名詞の一部あるいは全部が未学習の複合名詞と一致することがある。そのことにより、ポインタ方式の解析率が右方向最長一致法と比べて向上した。
- 2 章で示したように、マーク方式による推論 (特に、接辞に関する推論) の結果、マーク方式の解析率がポインタ方式に比べて向上した。
- 「複製 | 機 | 録」、「利用 | 団体 | 体」のようなものは、全候補を求めた後に、接続の強さや、単語の数、及び、統計的性質 (四字よりなる漢字列は二字と二字にわかれることが多いなど) などによる得点をつけることにより、除去することが可能である。だが、探索時間を考えて深さ優先探索を採用した。「分割誤り」と「学習の影響」による誤りが 4 つだったことを考えると妥当な選択だといえるだろう。

- 分野を限ることにより同音語が少なくなることが知られている [1]。(先の「ぐん」の例でも、数学ならば「群」であり、歴史ならば「軍」であろう。) 従って、専門分野毎に分けられた辞書がワープロで使われているのである。

ところで、分野を限ると、そこで使われる複合名詞の数には限りがある。よって、設定される素性の種類にも限りがある。従って、専門分野毎に設定される素性の最低値を決めておいて、分野毎の素性を付与することにより、一つの辞書がいくつかの専門語辞書を兼ねるようになる。先の例でいうと、

```
辞書 = {(じゆう 自由 (1 5001) ())
        (ぐん 軍 () (1))
        (ぐん 群 () (5001))}
```

素性の情報 = {1 以上 5000 以下の素性は歴史において使われ、5000 以上の素性は数学において使われる。}

などと設定しておけば、歴史の場合は「自由軍」がでて、数学の場合は「自由群」がでることになる。

また、辞書の使用段階においては、同音語の使用状況 (「軍」と「群」) などから求められている分野を推定し、それに合わせるような複合名詞をだすことも可能であろう。

5. 本稿では一種類の係り受け関係しか考えなかったが、マーク方式は複数の種類の係り受け関係についても適用できる[3]。例えば、「自動車事故」は「自動車に係わる事故」、「自動車工場」は「自動車を作る工場」などの文に展開することができる。これらを表す単語の構造として、

(じどうしゃ 自動車 (1 5001) ())

(じこ 事故 () (1))

(こうじょう 工場 () (5001))

などを登録して、さらに1以上5000以下の素性は「係わる」を示し5000以上は「作る」を示す、という情報を登録しておけば、複合名詞を構成する単語間の細かい係り受け関係を表すことができる。

7 むすび

単語にマークを設定し、マーク間に対応をつけることにより、複合名詞を効率的に辞書に登録することができることを示した。また、分野を限ることにより未学習の複合名詞の分解にも役に立つ。さらに、マークの種類により複数の種類の係り受け関係を表すこともできる。今後の課題としては、実際にかな漢字変換に応用することがあげられる。

参考文献

- [1] 田中、岡坂、長田：「同音異義語の解析 - 専門分野における -」、自然言語処理研究会 32-5、(1982)
- [2] 中野洋：「同音語の判別」、自然言語処理研究会 33-4、(1982)
- [3] 田中康仁：「語と語の関係について」、自然言語処理研究会 41-4、(1984)
- [4] 宮崎正弘：「係り受け解析を用いた複合語の自動分割法」、情報処理学会論文誌 Vol.25、No.6、pp970-979、(1984)
- [5] 武田、藤崎：「統計的手法を用いた漢字複合語の短単位分割」、自然言語処理研究会資料 48-2、(1985)
- [6] 西野、藤崎：「漢字複合語の確率的構造解析の試み」、情報処理学会第34回全国大会、pp1193-1194、(1987)
- [7] 松岡、菊池、藤原：「複合語と埋め込み文の格フレーム形式への一展開法」、情報処理学会第34回全国大会、pp1195-1196、(1987)
- [8] 石崎雅人：「日本語複合名詞の解析」、情報処理学会第35回全国大会、pp1315-1316、1987
- [9] 宮部、坂井、谷、川端：「PIVOT J-E：日本語複合語解析」、情報処理学会第37回全国大会、pp937-938、(1988)
- [10] 石崎雅人：「複合名詞における語と語の関係について」、情報処理学会第37回全国大会、pp1055-1056、(1988)
- [11] 坂井、宮部、村木：「PIVOT：日本語接辞解析」、情報処理学会第38回全国大会、pp388-389、(1989)
- [12] 亀井、村木：「日本語助数詞の分析」、情報処理学会第41回全国大会、No.3 pp155-156、(1990)
- [13] 福田、板橋：「係り受け結合頻度を用いた複合名詞解析の一方法」、情報処理学会第42回全国大会、No.3 pp9-10、(1991)
- [14] 植田、小松、横尾、宮崎：「部分複合語による複合名詞構造解析」、情報処理学会第43回全国大会、No.3 pp175-176、(1991)
- [15] 森本、青江：「トライ構造による複合語解析」、情報処理学会第43回全国大会、No.3 pp123-124、(1991)
- [16] 安原、窪：「双方向最長一致法による日本語漢字複合語自動分割」、情報処理学会第43回全国大会、No.3 pp125-126、(1991)
- [17] 内山、板橋：「係り受け共起頻度を利用した複合名詞の解析」、情報処理学会第44回全国大会、No.3 pp175-176、(1992)
- [18] A. チャンダー、J. ブラハム、R. ウィリアムソン 著 坂井利之監訳：「ブルーボックス コンピューター用語辞典 (第二版)」、講談社、(1988)