

概念解析における学習機能

高 橋 亘*

Machine Learning System for the Concept Analysis

Wataru Takahasi

Abstract : As is insisted by Ferdinand de Saussure in the case of many natural languages, a word of Japanese language as a linguistic sign has an arbitrary relationship between the signifier and the signified. Consequently, a word of Japanese language has a multiplicity of meaning. This fact means that a unit of perception may be corresponded to a collocation in which the polysemy of the entities is resolved uniquely. We call such a collocation which directly leads a real perception a perceptive collocation. We have already discussed a scheme in which the semantic elements are systematically derived from the perceptive collocations in the text paragraphs. In our scheme an aspect of the words which restrict their meanings in the process of their connection is represented with the linear operators.

In this article, a machine learning system is proposed. In our machine learning system the perceptive collocations are systematically learned from the text sentences. In order to raise the efficiency of learning our learning system possesses a wide variety of criterion with which text sentences are decomposed into the collocations. The decomposition method is closely related to the global variables in the M language. Some function words in the Japanese language play a crucial role in our method.

Our Japanese analyzing system with the new function can reasonably categorize the sentences into the fields of concept with the aid of the correlation functions which correlate between the perceptive collocations and the semantic elements.

Key words : 知覚連語 perceptive collocation 学習機能 machine learning 意味解析 semantic analysis 概念解析 concept analysis 日本語 Japanese コミュニケーション支援 computer aided communication ユニバーサル・インターフェイス universal interface

1 はじめに

この論文は、先に“M 言語による意味解析システムの学習機能”と題する論文¹⁾として、『Proceedings 2004 M Technology Association of Japan』に公表した内容に、その後の理論の進展と基礎付けを加筆し、理論の妥当性と整合性を強調するものである。

ソシュールの言うように、言語記号が概念に対して恣意的にふられた記号であるならば、単語が実体的な概念を表現することは不可能であり、人間の知覚に直接うったえるような表象は結合語によって与えられる。²⁾また、失文法性失語症患者の発話には、言葉が知覚と直接結びつくような連語として記憶されていると考えられる代償的発話や機能語の欠落が見られる。³⁾これらの知見から我々は、日本語文は、知覚に直接うったえるような結合語、もしくは連語を

*関西福祉科学大学社会福祉学部 教授

構成しながら叙述されることを仮説とし、このような連語を知覚連語とよんだ。^{2,4)}さらに知覚連語が日本語の文やパラグラフの解析の基礎を与えるとして、知覚連語とベクトル空間の線形演算子や基底ベクトルとを対応づける概念解析の方法を提案した。^{2,4)}

知覚連語を概念解析の基礎にとるということは、知覚連語の辞書を如何にして構成するかという問題をはらんでいる。仮に知覚連語の大半が2単語もしくは3単語から構成されているとしても、現実使用されている知覚連語の数は夥しいものになる。このような概念解析の方法を提案するという事は、それ自体、知覚連語の自動学習システムの方法が不可避であるということ、意味している。この論文ではM言語のデータの階層性を活用した自動学習システムの方法について述べる。

日本語の自然言語解析については、形態素解析が基礎であると考えの人が多く、既成の形態素解析システムとしては京都大学のJUMANや奈良先端科学技術大学院大学の『茶筌』がある。⁵⁾形態素解析という考え方は、言語をより基礎的な要素に還元して考える点ではより基礎的なものであるといえる。しかしながら、意味解析を考えたり、漢字の読みを決定する問題を考えたりするには、形態素を基礎におく考え方には、いくらか困難が伴う。この困難は、単語もしくは形態素の意味的多価性(形態を同じくする要素がいくつもの意味を持ったり、いくつもの品詞に対応したりする)に由来する。Support Vector Machineをもちいた修正学習法においても、この種が多価性は解消しそうにない。⁶⁾我々はこの多価性の問題を解決するために知覚連語を基礎に取るのである。我々の日本語解析システムは、日本語をある単位で切断していくという側面からみると形態素解析に似ており、辞書から豊富な知覚連語を削除し、形態素に対応する単語のみを残せば、単なる形態素解析システムとさほど異なることはない。

近年、意味解析の基礎を研究する人が増えて

いるが、多くは形態素解析の上に立って、統計的手法によって、属性としての単語と属性を説明する単語との相関関係をつける手法を用いており、語の結合規則から意味的に純粋な状態を取り出す議論が脱落している。⁷⁾単語の多義性が語の結合によって限定されたり、総合されたりする関係を無視して、精度の高い意味解析が行われるはずがない。

2 M言語による最小数連語切断の方法

我々の日本語解析システムが用いてきた、日本語文を連語に切断するシステムは、辞書登録されている連語についてあらかじめ、“連語を構成する文字についての階層構造をもった大域変数”(以下、「連語構成文字階層大域変数」と呼ぶ[図1])を定義しておくところに特徴があった。このシステムでは、開発の初期から、右方最大連語切断とも言うべき方法を用いてきた。我々の言う右方最大連語切断とは、既に連語切断が完了している文の位置から、文に沿って右方に一文字ずつたどりながら、それに呼応して上述の階層構造を手繰り、階層構造の末端に達したところで、当該の文を切断する方式である。

右方最大連語切断法は連語切断を高速に成し遂げる点で有効である反面、まだ登録されていない結合語を発見するには、見落としが生じるという欠点を持っている。未知の結合語の発見

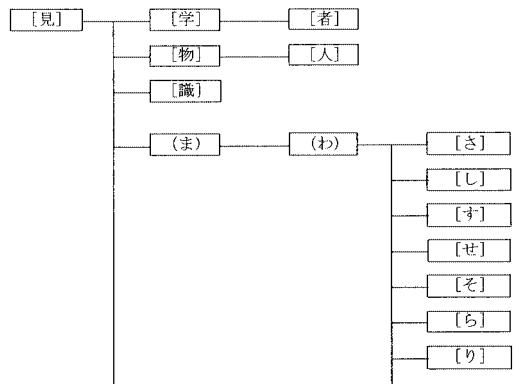


図1 見て始まる単語(連語)の構成文字階層構造



図2 ^CASC の階層構造

にはどうしても、試行的連語切断の方法が必要である。しかし、文を試行的に切断する方法では、試行回数の多さが常に問題となる。今我々が提案したいのは M 言語による効率的な試行的連語切断の方法である。

我々の試行的連語切断のアルゴリズムにおいても、上述の「連語構成文字階層大域変数」は重要である。ただし、今の場合、末端に限らず登録連語が存在する途中のノード（図1で言えば [] のついたノード）でも切断するところが右方最大連語切断の場合と異なる。

アルゴリズムの第一段階は、当該の文を右方に一文字ずつたどりながら、呼応して上述の階層構造を手繰り、階層構造の登録連語が存在するノード毎に切断することからはじまる。切りとられた連語を第1添字として大域変数（局所変数でもよい；仮に ^CASC とする）を定義し、文の残余を変数の値とする。この作業は文頭から最大連語（つまり連語構成文字階層の一つのリーフ）に達したところで終了させる。このことにより ^CASC の第1階層が総て定義し了わる。

次に ^CASC の第1階層を照合順位の後ろ（長い連語ほど照合順位は後になる）からたどって、^CASC の値である文の残余に対し同じ方法を適用して、切り取った連語を第2添字と

して ^CASC の第2階層とする。同様の操作を繰り返して、^CASC の階層を次第に増やしていくと、やがて ^CASC の値が NULL になるところが出来る。一度 NULL 値が達成されるとそのブランチは階層の増加は停止する。最初に NULL 値が実現したノードから、有効なノードの最大値を推定し、そこまでは、切断操作を行うが、それ以上のノードについては操作を打ち切ることとする。

以上の操作の後 NULL 値が実現した ^CASC をピックアップすれば、これらの中に有効な切断例の候補がある。もちろん、有効な候補の中に右方最大連語切断や、最小数連語切断（切断した連語の数が最小になる）も含まれることになる。このようにして得られる大域変数 ^CASC の例の階層構造を図2に示す。

3 知覚連語切断における機能語の活用と意味解析から見た機能語のアイデンティティー

第2節で述べたアルゴリズムに従うと、合理的に有効な連語切断の候補を高速にピックアップすることが出来る。しかし、文が長くなると ^CASC の階層が増え、階層が増えたと切断のために予想される所用時間は急速に大きくなる。この問題を解決することが出来るのは、句読点や機能語の存在である。句読点が文にお

る意味単位を意図的に切断することは論をまたないが、機能語もまた意味単位に抵触することなく文を切断することが出来る有力な候補である。

単語の中には、意味形成のために重要な役割を果たすものと、文の文法的構成のみに重要なものがある。知覚連語に寄与する単語が前者であり、助詞、助動詞、形式名詞、補助動詞など、機能語と呼ばれているものが後者である。しかし、句読点と異なり、同じ形態を持つ単語が、意味形成に一役買ったたり、非常に機能的に用いられたりするので、形態のみで意味形成に関与するのかわらないのかを決定することは難しい。つまり、機能語は意味形成に一切関与しないのかということ、知覚連語の中に入り込んでくる機能語はいくつもある。ただ、日本語解析の経験的規則から言えば、知覚連語を正確に切り分けていくと、真に機能的な単語は孤立してくるのである。したがって、文を一度、右方最大連語分割してみると、機能語が孤立する。孤立した機能語と句読点を頼りに文を切断し、切断された各部分に先に述べた試行的連語切断法を用いて再切断すると、高速性を保持しつつ、かつ学習対象の連語の見落としを少なくすることが出来る。

上に述べたように、我々の知覚連語切断において、機能語の存在はアルゴリズム上大変重要なものが、機能語をこのような形で活用できる基本原理は、機能語が内在的に文の構成上の機能的役割を担うために生成されるという本性を持っているからに他ならない。最初は語の結合による意味の構成や、語の意味的差異を機能的に表現するために生成された機能語が、次第に意味の形成という役割から自立して、構文を形成するための機能的な言葉として、その本性を確立してくるのである。通常機能語として扱われている語においても、機能的側面の強い語と弱い語が存在する。

我々は、副助詞の「は」、や「～と考える」、
「～と思う」、
「～と考える」などの構文で使用

される補文構造を作る助詞「と」、多くの形式名詞などは機能性が非常に強いと考えている。これらの語を含む多くの連語は、文の構文的構造をつき混ぜ、そのことによってかえって連語分解を混乱させるもととなる。補文や形式名詞にかかる修飾文には、文の形に関わりなく一般の文がくる。よく知られているように、日本語文の構文には大ざっぱな分け方をすると、文尾にくる品詞によって4種類に分類される。それらは名詞文、動詞文、形容詞文と形容動詞文である。

(i) 名詞文；

文尾が「名詞」、「名詞+助動詞」になる。

(ii) 動詞文；

文尾が「動詞」、「動詞+助動詞」になる。

(iii) 形容詞文；

文尾が「形容詞」、「形容詞+助動詞」になる。

(iv) 形容動詞文；

文尾が「形容動詞」、「形容動詞+助動詞」になる。

補文や形式名詞にかかる修飾文には、これらの4種類のうちいずれがくることも可能である。

これに対し屈折辞とみなされる、助動詞「た」、「だ」や助詞「て」（「で」）、「たら」や否定辞「ない」、「ず」、「ぬ」等は他の句（知覚連語）に結合して知覚連語の一部となると考える方が、品詞認識の精度が向上する。

4 知覚連語を学習するシステム

第2節および第3節で議論してきた試行的連語切断法では、いくつかの切断例を提示する。提示された例のうちいずれを採用するのかという問題があるが、意味解析や漢字の読み決定について、既知の知覚連語を用いた試行的連語切断が提示する切断例の中では、右方最大連語切断が最も正しいものを与えることがわかった。このことは、既知の知覚連語を用いる限り、右

方最大連語切断を適用するべきであるという、旧来の解析方法に誤りが無かったことが証明されたことになる。しかし、新しい知覚連語を発見し、これを学習した場合の結果を予測すると事態は変わる。

我々の知覚連語学習システムは、単語や連語の品詞（もしくは文法的役割）の並びを観測していて、知覚連語のプロトタイプを見つけ出し、これを知覚連語辞書の候補としてリストする。我々はこの学習候補をリストする機能のうち、知覚連語のプロトタイプを見つけ出す機能を独立させ、学習により連語分割数がどのように変化するかを予測する関数を構成した。したがって我々の知覚連語学習システムは、「学習の結果想定される連語分割数を与える関数」と、「知覚連語辞書の候補をリストする機能」の2つをあわせ持つことになる。前者の関数は試行的連語切断が提示する切断例の絞り込みに適用され、学習が実際に行われた場合に最小数連語切断となるものを抽出することに使用される。その結果、我々は文の連語切断のうち最も有望な候補を2つ得ることになる。右方最大連語切断を与えるものと学習後最小数連語切断となるものである。この両者は一致することもあるし、異なることもある。学習という観点から見た場合にはこの双方が学習の対象になり、現在の我々のシステムでは、双方ともに「知覚連語辞書の候補をリストする機能」が適用される。

知覚連語切断の結果としてどちらが有効かという問いについては、一般に日本語の意味解析もしくは漢字の読み決定のような目的のためには右方最大連語切断が適していることが多い。これは、日本語の意味は名詞が主となる知覚連語が担っていることが多く、動詞が後にくる文の構造上このような連語が左に来やすいためであると考えられる。

それでは最小数連語切断（学習完了後）が有効でないのかといえば、日本語を手話に変換する際などに有効だと考えられる切断が与えられ

ることが多い。これは日本語文では動詞が後にくることが多く、手話単語は「名詞」+「助詞」+「動詞」という3単語連語に対応しているため、右方最大連語切断で見落とされがちな右側にある大きな知覚連語を抽出しやすいためであると考えられる。

試行的連語切断が提示する切断例には、現在は捨象してしまっている切断例を含めて、知覚連語成立のクライテリオンの決め方や目的によっては、ピックアップ出来るものがあるかもしれないが、それは今後の課題である。

5 知覚連語切断と日本語文の構文

語が結合して意味的に純粋な状態を作るレベルという知覚連語の定義は、いくらか漠然としていて、厳密なものではない。それでも先の論文でみたように、二つの名詞が結合したり、もしくは助詞を挟んで「名詞+助詞+名詞」のような形の結合語が単語よりは意味を限定していて、より知覚に訴えてくることは容易に理解できる。いったいどの程度の語の結合が最も知覚連語と呼ぶのにふさわしいのかという問題は、いささか経験的な問題であり、初期の段階で厳密な定義を与えることは難しい。まずは、知覚連語の候補として適切だと考えられる品詞列のパターンをリストして、経験的に知覚連語の候補を挙げていき、その結果に、さらに経験則を追加することによって、候補を絞り込むアルゴリズムを確立していく方法をとってみたい。まず、現在我々が知覚連語のパターンと考えているものを Appendix にリストするが、これで完全であるというものではないことを、強調しておきたい。この節で述べることは、このような知覚連語のパターンを適用して、学習すべき知覚連語の候補から、実際に学習すべき連語を選択する際、（我々の学習システムでは、ここに人間が手で行う作業が入らざるを得ない）日本語文の構文的構造からくる制約についてである。

知覚連語で文を切断していく際に、動詞が原

型のまま、他の名詞を修飾しているような文では、少し注意がいる。たとえば、次のような例文をみよ。

その証拠は、この湖西を走る車の窓のそとをみよ。【司馬遼太郎(街道をゆく)】

例文の場合、局所的に見れば、“この湖西を走る”、“この湖西を走る車”、“湖西を走る”、“湖西を走る車”、“走る車”、“車の窓”、“車の窓のそと”などは、ある程度明確な意味を与えるという点では、知覚連語の候補と考えることができる。しかし文の構造から「車」は、“車の窓のそと”の構成子であると同時に“この湖西を走る”という動詞文の修飾を受けるという二重性を持つ。できるだけ長い知覚連語を取るという立場では、

“この湖西を走る車”
“車の窓のそと”

の二つを取りたいところであるが、「車」が二重に入るため、知覚連語分解になじまない。もっとも、“この湖西を走る車の窓”、“この湖西を走る車の窓のそと”、といった長い連語を知覚連語として採用することも考えられないわけではないが、長い連語は、一般的には、総合的な効果が働くために、意味を純粋に限定することからずれてしまうことが多い。さらに、文の構成が複雑になれば、修飾する動詞文、もしくは修飾される名詞句が、大きな構造をもって、一つの知覚連語でカバーしきれないことも起きる。このような矛盾を解決するために、我々は、「動詞文は名詞を修飾することができる」という可能性を残し、例の文を

- ①この湖西を走る>>{車}
- ②その証拠は、{車}の窓のそとをみよ

の二文のように理解するという立場をとりた

い。つまり、“この湖西を走る車”、“湖西を走る車”、“走る車”、などの動詞が名詞を修飾するような連語は、原則的には、知覚連語として採用しないことにするのである。この立場は、部分的な文によって語を修飾するという、いわゆる複文構造を分解して扱うという立場である。この立場を取る場合、「動詞+名詞」や「動詞+形式名詞」、あるいはこれに類する構文で「助動詞」や「助詞」を含むものの多くは、廃棄する必要がある。このパターン問題は、複文構造の意味決定問題として、今後の問題としたい。

文による語(句)の修飾の問題をもう少し一般的に考察してみよう。第3節で述べた4種類の文は、名詞や動詞、あるいは形容詞、形容動詞にかかって、これらを修飾するように用いることができるから、一般の文は、これらの文を枝葉に持つ文型となることが推察できる。例として、「長周期地震動」について言及した朝日新聞の記事(2004年9月11日)の冒頭からの3文について考察してみる。

- (1) 紀伊半島沖で5日にあった連続地震の際、関東や大阪、濃尾平野など広域で、ゆっくりとした揺れが長く続く長周期地震動が起きていたことが東京大学地震研究所の解析でわかった。
- (2) 高層ビルや橋など大きな構造物に被害が出る恐れのある地震動で、昨年の十勝沖地震では石油タンク火災の原因になった。
- (3) 今回は被害報告はないが、震源から遠く離れた場所で広く観測されたことで、東南海地震など巨大地震に備えた対策の必要性が裏づけられた。

(1)の文の中で上記の4文型が修飾的に使われている部分を<>でマークしてみると、次のようになる。

- (1)*<紀伊半島沖で5日にあった>連続

地震の際、＜関東や大阪、濃尾平野など広域で、ゆっくりとした揺れが長く続く長周期地震動が起きていた＞ことが＜東京大学地震研究所の解析で＞わかった。

したがって、この文は修飾関係に注目すると次の4つの部分に分解されることがわかる。

(1)**

- ①紀伊半島沖で5日にあった>>|連続地震|
- ②関東や大阪、濃尾平野など広域で、ゆっくりとした揺れが長く続く長周期地震動が起きていた>>|こと|
- ③東京大学地震研究所の解析で>>|わかった|
- ④|連続地震|の際、|こと|が|わかった|。

つまり①、②は、それぞれの動詞文が|連続地震|と|こと|を、③は、名詞文が|わかった|を、修飾していることがわかる。これらの三つの部分で|連続地震|、|こと|、|わかった|の意味規定がなされた後に、それぞれの意味を用いて、④が述べられているのである。

引き続き二つの文についても、同様に< >マークを付けると、(2)*、(3)*のようになり、それぞれ、(2)**、(3)**のような修飾関係が存在することがわかる。

(2)*<高層ビルや橋など大きな構造物に被害が出る恐れのある>地震動で、今年の十勝沖地震では石油タンク火災の原因になった。

(3)*今回は被害報告はないが、<震源から遠く離れた場所で広く観測された>ことで、<東南海地震など巨大地震に備えた>対策の必要性が裏づけられた。

(2)**

- ①高層ビルや橋など大きな構造物に被害が出る恐れのある>>|地震動|
- ②|地震動|で、今年の十勝沖地震では石油タンク火災の原因になった。

(3)**

- ①震源から遠く離れた場所で広く観測された>>|こと|
- ②東南海地震など巨大地震に備えた>>|対策|
- ③今回は被害報告はないが、|こと|で、|対策|の必要性が裏づけられた。

三つの文は、例文として多くはないが、一般の傾向をよく表している。つまり、一般的な文の中では、動詞文が名詞(名詞句)を修飾したり、名詞文が動詞(動詞句)を修飾したりするケースが圧倒的多数を占める。これに対して、形容詞文や形容動詞文が語や句を修飾する場合は、修飾文が長くなることはないという傾向がある。形容詞や形容動詞は単独で次にくる語を修飾していることが多い。上の例文中でも、“長く続く”、“大きな構造物”、“遠く離れる”、“巨大地震”などがそれである。

文学作品の中から“長い”が名詞を修飾している例を検索して、形容詞文として少し構造をもった形で次の名詞を修飾している例を抽出すると、

(4) まず、軸の長い鉤を買ってきて、その軸に赤や黄のモールを巻きつけ、金糸や銀糸で縛る。【開高健(夏の闇)】

(5) 四月に大雪が降るほど、冬の長い年でない、螢の奴は狂い咲いてくれんちゃ】【宮本輝(螢川)】

(6) じんじん端折りの頬冠りや、赤い腰巻の姉さんや、時には人間より顔の長い馬にまで逢う。【夏目漱石(草枕)】

(7) どうしても馬鹿で気の長い人種の発明

したものとはほか取れない。【夏目漱石(草枕)】

(8) 背のずんぐりした、色黒の、髯づらと、くっきり締った細面に、襟の長い、撫肩の、華奢姿。【夏目漱石(草枕)】

などの例文の中で、次のようなものが挙げられる。

軸の長い>>{鉤}
 冬の長い>>{年}
 人間より顔の長い>>{馬}
 馬鹿で気の長い>>{人種}
 襟の長い>>{華奢姿}

“軸の長い”，“冬の長い”，“人間より顔の長い”，“馬鹿で気の長い”，“襟の長い”などは形容詞型の知覚連語と考えてよく、最後の例のように修飾語と被修飾語が離れている場合を除いて、これらの知覚連語が形容する対象も直接的であるから、被修飾語も含めた，“軸の長い鉤”，“冬の長い年”，“人間より顔の長い馬”，“馬鹿で気の長い人種”なども知覚連語と考えられる。したがって、形容詞文や形容動詞文の修飾から生成される複文構造はないと考えてよい。

6 まとめと展望

日本語解析において、文の意味の大半は知覚連語が担っているとして、知覚連語を切り出すシステムを考案した。文の知覚連語を切り出していく作業は、文の機能語を自然に孤立させる。

我々の日本語解析システムは未知の知覚連語を発見し、学習すべき候補としてリストアップする機能を持っている。このような学習を可能にするには、右方最大連語切断のみならず、試行的連語切断を実行することが必要であるが、我々は、このために M 言語特有のアルゴリズムを用いた、試行的連語切断法を考案し

た。試行的連語切断の効率を向上させるには、まず右方最大連語切断を行って、文の機能語を孤立させ、機能語を目安に文を部分に分けたのち、部分に対して試行的連語切断を実行することが好ましい。

我々のシステムが知覚連語切断の最有力候補として提示するものは、右方最大連語切断の切断例と、学習後最小数連語切断となる切断例であるが、我々の学習機能はこの双方に適用される。切断結果には現在のところ、右方最大連語切断の結果を採用している。最小数連語切断の結果は、日本語・日本手話変換システムで重要になると考えられるが、これは将来の課題である。

機能語の認識や学習候補のリストの機能は人間の脳の中で行っている機能にも対比される。この点で我々の人工知能と人間の脳との対比はより密接なものになったといえる。我々の研究が、先に挙げた手話の問題を含めて、人間の脳が行っている言語と意味知覚の関係に関する認知科学的な理解への突破口にならないことを願うものである。

Appendix

我々が現在、知覚連語の候補として考えている品詞列のパターンは、次のとおりである。

I 人名を形成するパターン

- (1) 姓名 個名
- (2) 人名 称号
- (3) 姓名 称号
- (4) 個名 称号

II 連語名詞を形成するパターン

- (1) 名詞 名詞
- (2) 人名 名詞
- (3) 姓名 名詞
- (4) 個名 名詞
- (5) 連語名詞 名詞
- (6) 名詞 連語名詞
- (7) 名詞 接尾
- (8) 名詞 称号

- (9) 接頭 名詞
- (10) 名詞 助詞 名詞
ここに入る助詞は、「の」、「と」、「や」、「ヶ」
- (11) 名詞 助詞 連語名詞
ここに入る助詞は、「の」、「と」、「や」
- (12) 連語名詞 助詞 名詞
ここに入る助詞は、「の」、「と」、「や」
- (13) 個名 助詞 名詞
ここに入る助詞は、「の」
- (14) 姓名 助詞 名詞
ここに入る助詞は、「の」
- (15) 人名 助詞 名詞
ここに入る助詞は、「の」
- (16) 形容 名詞
- (17) 連語形容 名詞
- (18) 形容 連語名詞
- (19) 形動 名詞
- (20) 連語形動 名詞
- (21) 形動 助詞 名詞
ここに入る助詞は、「の」
- (22) 連体 名詞
- (23) 代名 助詞 名詞
ここに入る助詞は、「の」、「と」
- (24) 動詞 名詞 (大半は削除)
- (25) 動詞 助動詞 名詞 (大半は削除)
ここに入る助動詞は、「だ」、「た」
- (26) 動詞 形名 (大半は削除)
- (27) 連動詞 形名 (大半は削除)
- (28) 形容 形名 (大半は削除)
- (29) 連形容 形名 (大半は削除)
- (30) 形動 形名 (大半は削除)
- (31) 動詞 助動詞 形名 (大半は削除)
- (32) 名詞 助詞 形名 (大半は削除)
ここに入る助詞は、「の」
- (33) 連体 助詞 形名 (大半は削除)
ここに入る助詞は、「の」
- (34) 代名 助詞 形名 (大半は削除)
ここに入る助詞は、「の」
- (35) 副詞 助詞 形名 (大半は削除)
ここに入る助詞は、「の」
- (36) 連体 形名

Ⅲ 連語代名詞を形成するパターン

- (1) 代名 接尾

- (2) 名詞 助詞 代名
ここに入る助詞は、「の」

Ⅳ 連語動詞を形成するパターン

- (1) 動詞 動詞
- (2) 名詞 動詞
- (3) 副詞 動詞
- (4) 形容 動詞
- (5) 形動 動詞
- (6) 名詞 助詞 動詞
ここに入る助詞は、「は」、「が」、「の」、「に」、
「を」、「で」、「へ」、「と」
- (7) 代名 助詞 動詞
ここに入る助詞は、「は」、「が」、「に」、「を」
- (8) 動詞 助詞 動詞
ここに入る助詞は、「て」

Ⅴ 連語形容詞を形成するパターン

- (1) 名詞 助詞 形容
ここに入る助詞は、「が」、「の」、「に」、「より」
- (2) 動詞 形容
ここに入る形容詞は、「やすい」、「にくい」
- (3) 連動詞 形容
ここに入る形容詞は、「やすい」、「にくい」
- (4) 副詞 形容
ここに入る副詞は、「もっと」
- (5) 名詞 形容

Ⅵ 連語形容動詞を形成するパターン

- (1) 名詞 形動

Ⅶ 連語助動詞を形成するパターン

- (1) 動詞 助動詞
ここに入る助動詞は、「せる」、「させる」、「れる」、
「られる」、「ず」、「ぬ」、「じ」、「まじ」、「まい」、
「ます」、「ない」、「たい」、「らしい」
- (2) 連動詞 助動詞
ここに入る助動詞は、「せる」、「させる」、「れる」、
「られる」、「ず」、「ぬ」、「じ」、「まじ」、「まい」、
「ます」、「ない」、「たい」、「らしい」
- (3) 助動詞 助動詞
ここに入る助動詞は、「せる」、「させる」、「れる」、
「られる」、「ず」、「ぬ」、「じ」、「まじ」、「まい」、
「ない」、「たい」、「らしい」

- (4) 形動 助動詞
ここに入る助動詞は、「そうだ」、「ない」、「らしい」
- (5) 形容 助動詞
ここに入る助動詞は、「そうだ」、「ない」、「らしい」
- (6) 名詞 助動詞
ここに入る助動詞は、「ない」、「らしい」、「らし」
- (7) 名詞 助詞 助動詞
ここに入る助詞は、「の」
助動詞は、「ようだ」

Ⅷ 語尾変化のない連語を形成するパターン

- (1) 名詞 助詞
ここに入る助詞は、「だけ」
- (2) 動詞 助詞
ここに入る助詞は、「だけ」

参考文献

- 1) 高橋 亘, “M 言語による意味解析システムの学習機能”, 『Proceedings 2004 M Technology Association of Japan』, 49-52 (2004).
- 2) 高橋 亘, 渡邊大樹, “コンピュータによる概念解析の方法”, 『関西福祉科学大学研究紀要』, Vol. 7, 59-81 (2004).
- 3) 長谷川直子, 清藤秀樹, 高橋 亘, “日本語における失文法失語と言語知覚の単位”, 『電子情

- 報通信学会技術研究報告』SP 2001-76, WIT 2001-30 (2001-10) [音声・福祉情報工学], Vol. 101 No. 352, 23-30 (2001).
- 長谷川直子, 清藤秀樹, 高橋 亘, “日本語における失文法失語と言語知覚の階層構造”, 『関西福祉科学大学紀要』No. 5, 75-89 (2002).
- 4) 高橋 亘, 渡邊大樹, “M 言語による概念カテゴリー解析機能”, 『Proceedings 2003 M Technology Association of Japan』, 29-32 (2003).
- 5) 黒橋禎夫, 長尾 眞, “日本語形態素解析システム JUMAN version 3. 61, 京都大学大学院情報科学研究科 (1998).
- 黒橋禎夫, 河原大輔, “日本語形態素解析システム JUMAN version 4.0, 東京大学大学院情報情報理工学系研究科 (2003).
- 松本裕治, 北内 啓, 山下達雄, 平野善隆, 松田 寛, 高岡一馬, 浅原正幸, “形態素解析システム『茶筌』version 2. 3. 3 使用説明書”, 奈良先端科学技術大学院大学・情報科学研究科・自然言語処理学講座 (2003).
- 6) 中川哲治, 工藤 拓, 松本裕治, “Support Vector Machine を用いた形態素解析と修正学習法の提案”, 『情報処理学会論文誌』, Vol. 44, 1354-1367 (2003).
- 7) グエン・ベト・ハー, 帆苺 謙, 石川 勉, 笠原 要, “単語の意味の類似性判別のための大規模概念ベース”, 『情報処理学会論文誌』, Vol. 43, 3127-3135 (2003).