

専門用語抽出・解析処理を考慮したコーパスの作成

○吉岡 真治 岡田 真穂 影浦 峯 小山 照夫

学術情報センター 研究開発部

{yoshioka,okadama,kyo,koyama}@rd.nacsis.ac.jp

概要: 学術情報センター (NACSIS) では、NACSIS テストコレクションの一環として、専門用語抽出・解析処理を特に考慮した NACSIS コーパスを作成している。このコーパスでは、専門用語に関する処理の中心である語彙的処理を考慮する必要があり、語彙単位の明確化が重要な課題となる。そこで、本稿では、これらの問題に対応するための NACSIS コーパスの設計指針について述べる。さらに、本コーパスの作成状況と、その公開に関する状況について述べる。

Construction of Terminologically-motivated Corpus

Masaharu Yoshioka, Maho Okada, Kyo Kageura, Teruo Koyama
National Center for Science Information Systems (NACSIS)

Abstract: We are currently constructing the terminology-motivated corpus (the NACSIS corpus) as a part of the NACSIS test collection project. There are two issues to constructing this corpus. One is a definition of domain that makes the terminology research different from IR keyword one. The other is a definition of lexical units in the texts that gives basic unit for terminology operations. In this paper, we discuss the approach to design the NACSIS corpus for solving these two issues, and show the status of NACSIS.

1 緒言

1990年代に盛んになった、専門用語自動抽出を代表とする専門用語の自動処理の研究は、「専門用語とは何か」、「情報検索 (IR) におけるキーワードとどう異なるか」といった点に無頓着であるため、現在まで、IR の応用のない自動キーワード抽出になっている場合が多い。一方、情報検索研究においては、計算機能力の増大により、これまで考えられてきた自動インデクシングの諸手法が現実的に応用可能になり、また、全文検索システムの現実化で言語的なインデクシングに依存しない高度な検索が可能となってきたが、これらはむしろ工学的な展開であり、依然として「どのような単位の言葉がどのような面で有効であるか」という基本的な問題は未解決であると言ってよい。

また、近年の計算機の処理能力の向上にともない、言語に関する大量の情報を収集し、統計処理などを行う研究を支援するための大規模なテキストコーパスの作成に関する研究が数多く行われている [1]。しかし、これらのコーパスが対象としている文章は、新聞記事や文学に関するものが多く、学術論文を対象とした日本語のコーパスはほとんどない。また、これらのコーパスは文法的な操作を中心に考えて作られているために、語彙的な単位について明確な基準が存在せず、EDR コーパス [2] や京都大学テキストコーパス [3] のようにコーパスの作成時に依拠した辞書に依存した語彙の単位が用いられている。このようなコーパスは意味的知識の獲得や文法知識の獲得などの研究には有効であるが [4, 5]、先に述べた専門用語処理の観点からはその利用方法が制限されることになる。

そこで、このような背景を考慮し、学術情報センターでは、専門用語関連の自動処理、日本語 IR への語彙的な単位の問題の明確化のための基礎データを形成することを目的として、NACSIS コーパスの作成を行っている。本稿では、この NACSIS コーパスの設計指針とその現状について述べる。

2 NACSIS コーパスの設計方針

NACSIS コーパスの目的は、専門用語の抽出・解析などの処理を行うための基礎的な基盤を与える事である。このような基盤を実現するためには、専門用語とは何か、専門用語研究にとって適切な語彙単位はどのような単位か、といった点に考慮する必要がある。そのために、この NACSIS コーパスでは次の 2 点を考慮してコーパスの作成を行っている。

1. 専門用語の専門性を規定する専門領域に応じたコーパスの作成
2. 専門用語を構成する最小基本単位を規定し、解析者のニーズに応じた解析を許すコーパスの作成

以降の節では、この各々について述べる。

3 専門領域

専門用語とは、ある専門領域を仮定したときに、その専門領域において特有の意味を持ち、用いられる語である。そのため、多くの場合、抽出された用語の集合を見ると、その専門領域に関連する文献を特徴づける語彙と重なり、情報検索研究におけるキーワード抽出の研究と混同されてしまいがちである。しかし、この専門用語の抽出がキーワードの抽出と異なる点は、そこで得られた語彙の集合の妥当性が、情報検索の様な文書の分別に役立つかどうかで判断されるのではなく、実際にその専門領域において特有の意味を持っているかどうかで判断されるという点にある。

このような問題に対応するためには、専門用語の専門性を規定する専門領域を定義する必要がある。このような専門領域を定義する方法として、その専門領域を構成する中心的な考え方やそこで

用いられる専門用語などを用いて定義するなどといった様々な方法が考えられる。しかし、このような用語を中心として専門領域を定義する方法では、本コーパスの一番の目的である専門用語抽出などの処理がトートロジカルになる可能性があり、その位置づけが困難となる。そこで、本コーパス作成においては、実際にこのような専門領域を明確に定義することを行わず、実際に存在している学会が専門領域に対応するという実際的な定義を行うこととした。

このような学会と専門領域を対応づけるという定義は非常に操作的な定義であるため、様々な専門領域に関する定義が混ざりあった形の定義となっている。しかし、実際のテキストデータの獲得及び分類に関する操作を考慮すると実際的な定義であると考ええる。

4 専門用語を構成する語彙単位の取扱い

4.1 語彙の最小単位

専門用語の抽出や解析といった操作を考えると、専門用語の語としての情報の他に、その用語の構成要素の形態素に関する情報が必要である。例えば、次のような文とその要素間の依存関係について考える。

- 情報検索システムを構築する。

1. (情報検索システム) を) → (構築する)
2. (((情報検索) システム) を) → (構築する)

この文章に対し、文法的な解析のみを考えるのであれば、1の依存構造で十分である。しかし、専門用語の解析を行うためには、「情報検索システム」と「情報検索」の関係などを考える必要がある。この様に用語の中身の構造にまで踏み込んだ解析を行うためには、用語の中についても、その構成要素を細かく分解した 2の依存構造の情報が必要である。

ところが、従来の文法的な操作を考慮して作られたコーパスにおいては、JUMAN[7]などの形態素解析ツールを用いて作成されているために、このような形態素解析ツールの辞書に登録されている長い専門用語については、前者のように分割されない形となり、辞書に登録されていない場合

は後者のような形式になっている。この様な語彙単位の一貫性に関する問題は、文法情報を取扱うコーパスとしてはそれほど問題にならないが、専門用語を取扱うコーパスとしては大きな問題となる。

そこで、本 NACSIS コーパスでは、この様な語彙の最小単位として形態素レベルの一貫した分割を行うこととする。また、専門用語としての形態素のまとまりである語を取扱う必要性から、形態素を合成した語についての情報も記述することとする。

しかし、形態素と語という2つの単位だけでは、一貫性のある分割が困難となるような事例が存在する。以下に、その様な事例となる語を、分割の基準を示した英語による表記ならびに、派生語尾ならびに屈折語尾と対応づけて述べる。

1. 重み付き情報検索
weight attached information retrieval
2. 重要文抽出 key sentence extraction
3. 重要な importan- 形容詞 - 派生 / 屈折 (= important)
重要さ importan- 名詞 - 派生 (= importance)
4. 重さ heavy 名詞 - 派生 (= weight)
重み heavy 名詞 - 派生 (= weight)
5. 重み付き情報検索
heavy 名詞 - 派生 attached information retrieval

まず、1と2の例を比較すると、「重み」と「重要」が対応づけられることになる。一方、3と4を比較すると「重要」に対しては、「重」のみが対応することになり、これを考慮すると5の様な、1で述べた形態素の分割方法と異なる分割方法が適切であることになる。つまり、語幹としての整合性を考えた分割方法と、派生や屈折を考慮した整合性を考えた分割方法が相矛盾する場合があります。

日本語において、この様な矛盾の原因はその語彙要素の語源(和語、漢語、外来語)に由来していると考えられる。つまり、漢語と外来語においては、派生や屈折の単位となる要素と複合語における複合単位の要素がテキスト中での現れとして

は同じものだと考えられるのに対し、和語では派生や屈折の単位となる要素が複合語における複合単位の要素よりも小さい事が原因であると考えられる [6]。

そこで、本コーパスでは、強い区切りと弱い区切りという2つの種類の形態素の区切りを導入し、次のような形式で語の要素分割に適用する。

1. 語幹が漢語や外来語の場合は、派生や屈折の語尾を強い区切りで区切る。
2. 和語については、漢語や外来語と同レベルの複合の基本単位を強く区切り、漢語や外来語にも適用可能な派生や屈折の語尾については弱く区切る。
3. 漢語や外来語に対して適用できない和語の派生や屈折の語尾については区切りを入れない。

この方法を適用することにより、強い区切りにより複合語を構成する語幹としての区切りの一貫性が保たれ、弱い区切りにより屈折や派生を考慮した区切りの一貫性が保たれることになる。この結果、先ほどの矛盾した事例である「重み付き情報検索」は、強い区切りを s、弱い区切りを w とすると、「重 w み s 付き s 情報 s 検索」という形で分割される。

4.2 語彙単位に付与するカテゴリ情報

前小節で述べたように本コーパスでは、語彙単位の区切りに関して従来の形態素解析ツールで用いられているような定義と異なる定義を行ったのに伴い、従来の形態素解析ツールなどで用いられている品詞カテゴリーについても再考察する必要がある。

この考察をするにあたり、次のような事例となる語を、先ほどと同じように分割の基準を示した英語による表記ならびに、派生語尾ならびに屈折語尾と対応づけて述べる。

- 重要 さ ↔ 重要な
importan- 名詞 - 派生 ↔ importan- 形容詞
- 派生 / 屈折

普通の文法要素カテゴリーは「重要さ」や「重要な」といった単位の語に対して、前者を名詞、後

者を形容詞といったカテゴリーに分類するためのものである。しかし、本コーパスで行う単位分割を取扱うためには、これよりも細かい「重要(importan-)」という単位で取扱う必要がある。ここで、対応する英語表記から考えると、前者の「重要」は名詞派生の語幹であり、後者は形容詞派生の語幹であると共に原型でもある。

また、この「重要」という要素が複合語を構成する際には、その語幹からの派生の可能性を考慮して、場合に依りて名詞派生である「重要さ」という意味で用いられ、形容詞派生である「重要な」という意味で用いられる。そこで、このような語幹の役割をはっきりさせるためには、一つの語としての文法要素カテゴリーだけではなく、このような語幹に対しても文法要素と同様のカテゴリーを付ける必要がある。

しかし、現在のところ、これら形態素要素に対して付与することが可能な文法要素と同様のカテゴリーが存在しない。例えば、多くの国語辞書をもとに、「国際」の様なそれ自体では1語を構成しない形態素について「普通名詞」という語としての文法要素カテゴリーを割り当てている。また、日本語の文法体系を見てみると、語だけでなく、屈折である活用、さらには接辞などの情報を取扱っている[8]。これらの事は、日本語において語の境界に関する明確な定義が定まっておらず、その結果として、様々なレベルの文法要素に関する情報が単一のレベルに押し込められた形になっていると考えられる。

この問題に対し、本コーパスでは、形態素と語に同じカテゴリーを用いて分類することとした。ただし、語に対するカテゴリーと形態素に対するカテゴリーでは、そのカテゴリーを付ける基準が少し異なる。つまり、語に対するカテゴリーは普通の文法要素カテゴリーとして文の中の語の役割に注目してカテゴリーを付け、形態素に対するカテゴリーは複合語中における機能的な役割と、その語が単独で文中に現れたときのカテゴリーを考慮して決定する。このカテゴリーは益岡文法[8]及びそれを基本とするJUMAN[7]の文法体系に基づいて作成した(付録1参照)。

図1にNACSISコーパスの一部分と、そのコーパスの読み方を示す。また、この中の「文節間の結合の強さに影響されることが知られている」という文を例にとり、コーパスの読み方につ

いて、簡単に説明する。

この文において、「文節間」、「影響される」、「知られている」については、普通名詞(NN)、母音動詞(VA)、サ変動詞(VSH)といった品詞カテゴリーが付与されると共に、「文節」、「影響」、「知ら」という複合語の構成要素についても、「普通名詞」、「サ変名詞」、「ラ行5段活用動詞」という品詞カテゴリーが付与されている。この様に語と形態素に同じカテゴリーを適用することにより、「影響される」における形態素である「影響」と「影響力」における「影響」の関係が明確に付けられることになる。また、「強さ」については和語であるため、弱い区切りにより、その派生が区別されている。

現在のカテゴリー分けの問題点としては、形態素レベルのカテゴリーの分類が不十分である点が挙げられる。例えば、現状の文法要素カテゴリーの問題でも述べた「国際」といった複合語の構成要素にはなるものの単一の語として文中に現れないような語に関するカテゴリー付けが単一の語として文中に存在できる要素と同じカテゴリーに入っている。このような問題を解決するためには、形態素レベルにおけるカテゴリー分けに対するさらなる考察が必要である。

5 専門用語研究のためのNACSISコーパスの利用法

ここでは、専門用語研究における、本NACSISコーパスの利用法について考察する。本コーパスの特徴は、形態素、語といった語彙単位について一貫性を持って分割され、カテゴリーの情報が付与されているという点である。ここで、一般的な語彙研究との関係で考えると、本コーパスでは、次のような情報が提供可能である。

1. 語の単位

本コーパスでは、複合語単位の語は、複合された語として表現され、単一の語は、形態素による構成要素を持たない語として表現される。そこで、これらの情報を利用することにより、語の単位の情報が抽出できる。また、このレベルの語に付けられているカテゴリー情報は従来のテキストコーパスの情報と同じように利用可能である。

c 文節間 ぶんせつかん 文節間 NN
 文節 ぶんせつ 文節 NN K
 間 かん 間 TLNN K

の の の SCC W
 結合 けつごう 結合 NS K
 の の の SCC W
 強| さつよ| さ強| さ NN W
 に に SCA W

c 影響される えいきょうされる 影響される VAbs
 影響 えいきょう 影響 NS K
 ささする VSHmi W
 れる れる れる TLVAbs W

こと こと こと NF W
 が が が SCA W

c 知られている しられている 知られている VAbs
 知ら しら 知る VRAmi W
 れて れて れる TLVArytate W
 いる いる いる TLVAbs W

。。。 LSD
 EOS

しかし しかし しかし CC W
 、 、 、 LRD

c 文構造 ぶんこうぞう 文構造 NN
 文 ぶん 文 NN K
 構造 こうぞう 構造 NN K

の の の SCC W

c 厳密な げんみつな 厳密だ JNA_DArt
 厳密 げんみつ 厳密 JNAst K
 な なだ TLJJNA_DArt W

解析 かいせき 解析 NS K
 は は は SSB W

コーパスの読み方：複合語の場合、行の先頭にcの記述があり、その複合語の構成要素は、タブでインデントされている。その他の語は1単語1構成要素である。各行には、1. 文章中での表現、2. 読み、3. 辞書見出し、4. 文法カテゴリー、5. 語種(複合語の場合は省略)がスペースで区切られて左から順番に記述される。また、弱い区切りについては、「|」によって区切られる。

図 1: NACSIS コーパスの記述 (一部を抜粋)

2. 複合語の構成情報

本コーパスでは、各々の複合語について、その構成要素を対応付けて記述しており、形態素の構成情報は容易に獲得できる。また、この形態素要素のレベルは語のレベルと区別されているために、構成要素に対するカテゴリー情報は形態素のためのカテゴリー情報だと解釈され、語のためのカテゴリー情報と明確に区別できる。

3. 派生や屈折語尾の適用可能性

複合語の各構成要素には文法カテゴリー情報や語種の情報が記述されている。この記述と、実際の派生や屈折語尾の組み合わせを調べることにより、語種などの情報に対する派生や屈折語尾の適用可能性の検討が可能になる。

4. 語幹に関する情報

本コーパスでは、派生や屈折の語尾と語幹の区別が明確に行われているので、語幹に関する色々な情報が容易に獲得できる。

ここで述べたようないくつかのレベルでの語彙処理をそれぞれ一貫して取扱えるようにしたことで、次のように、これまでの専門用語処理(特に用語抽出)で提案されてきた手法の多くを同一の条件で比較検討することが可能となる。

1. 複合語中の構成要素が最小単位で区分されている。
 一貫した構成要素の区分により、異なる複合語中での構成要素の関係を扱うことができる。そのため、多くの研究者が行っているような用語に重みをつけるために複合語中の構成要素を利用する研究などの役に立つ[9]。
2. 複合語の構成要素カテゴリーと語のカテゴリーの関連を付けることができる。
 この分類を用いることにより、専門用語の言い換えなどに関する処理が可能になる。例えば、「重要な文を抽出する」と「重要文抽出」の関係についての対応関係をつけるための情報がコーパスから獲得できる[10]。

3. 語の単位を明確に利用した語の弁別が可能である。

このコーパスでは、複合語の単位が明確に規定されているために、「情報通信」という語と、「情報通信システム」という「情報通信」を含むような語との区別が可能であり、また、語内の共起と語にまたがる共起を区別することができる。このため、文中の共起関係などの情報がより正確に獲得できるようになり、語に埋め込まれた共起を考慮する手法の検証などに有益である [11]。

NACSIS コーパスではこれらの専門用語研究に対して必要となる情報を与えることが可能であり、専門用語研究に対して貢献することが期待される。

6 NACSIS コーパスの現状

以上に述べてきた設計方針に基づき学術情報センターでは NACSIS コーパスを作成中である。まず、最初に対象とした専門領域を規定する学会として、人工知能学会 (AI)、情報処理学会 (IP) の2つの学会を選択した。この各々の専門領域に対してについて学会発表抄録データ (AI:1870 抄録・IP:26252 抄録) と全文論文データ (AI: 300 論文・IP: 1000 論文) を機械可読化し、タグ付きコーパスの作成を開始している。

このタグ付きコーパスの作成を次のような手順で行っている。

1. テキストデータの機械可読化

学会発表抄録データについては NACSIS の学会発表データベース [12] のものを用い、全文論文データについては冊子体に OCR をかけたものに対し、人手による修正を行っている。

2. JUMAN による第1次処理

機械可読化したテキストデータについて、JUMAN による形態素解析を行う。

3. 人手による語彙単位切りの修正及びカテゴリーデータの修正

先にも述べたように JUMAN の語彙単位の大きさは辞書の登録項目に依存する。そこ

で、このような単位切りの修正およびカテゴリーデータの修正を行う。

また、この専門領域のテキストコーパスに加え、各々の専門領域における専門用語に対するリファレンスとして各々の学会が編集に関係している専門用語辞書から作成した見出し語情報である専門用語リスト (AI: 約 4000・IP: 約 30000 語) を同時に作成している [13]。

最後に、現在の NACSIS コーパスの作成状況及び公開状況について簡単に述べる。タグ付け作業については、現在までに、人工知能分野の学会発表抄録データ (1870 抄録) のみが終了している。また、このデータについては、1998 年 12 月 10 日に NTCIR ワークショップ [14] の一環としてワークショップ参加者に公開している。

7 結言

本稿では、専門用語抽出・解析処理を考慮した NACSIS コーパスについて、その設計指針、実現方法について述べると共に、その応用事例における役割を考察することによりその有効性を検討した。また、最後に現在の本コーパスの作成状況について簡単に述べたように、まだ、本コーパスは作成途上のものであり、対象とする分野を増やしていく必要がある。さらに、実際の利用者の意見のフィードバックを受けた形でのコーパスの改良についても検討を行っていきたいと考えている。

謝辞

本研究の一部は、日本学術振興会未来開拓学術研究推進事業「高度分散情報資源活用のためのユービキタス情報システムに関する研究」プロジェクト (JSPS-RFTF 96P00602) によるものである。

参考文献

- [1] 竹沢寿幸, 末松博. 音声・テキストコーパスとその構築技術, 標準化動向. 人工知能学会誌, Vol. 10, No. 2, pp. 168-180, 1995.
- [2] 日本電子化辞書研究所. EDR: EDR 電子化辞書仕様説明書, EDR TR-041, 1993.

- [3] 黒橋禎夫, 長尾眞. 京都大学テキストコーパス・プロジェクト. 1997年度人工知能学会全国大会(第11回)論文集, pp. 58-62, 1997.
- [4] 宇津呂武仁, 松本裕治. コーパスを用いた言語知識の獲得. 人工知能学会誌, Vol. 10, No. 2, pp. 197-204, 1995.
- [5] 工藤育男, 井ノ上直己. コーパスに基づく共起知識の獲得とその応用. 人工知能学会誌, Vol. 10, No. 2, pp. 205-212, 1995.
- [6] 野村雅昭, 石井正彦. 学術用語語基連接表. 国立国語研究所, 1988.
- [7] 松本裕治, 黒橋禎夫, 山地治, 妙木裕, 長尾眞. 日本語形態素解析システム JUMAN version3.3. 京都大学工学部, 1997.
- [8] 益岡隆志, 田窪行則. 基礎日本語文法 - 改訂版 -. くろしお出版, 東京, 1992.
- [9] Hiroshi Nakagawa and Tatsunori Mori. Nested Collocations and Compound Noun for Term Extraction. In *COMPUT-ERM'98*, pp. 64-70, 1998.
- [10] Christian Jacquemin, Judith Klavans, and Evelyne Tzoukermann. Expansion of Multi-Word Terms for Indexing and Retrieval Using Morphology and Syntax. In *ACL-EACL'97*, pp. 24-31, 1997.
- [11] K. T. Frantzi, S. Ananiadou and J. Tsujii. Statistical Measures for Terminological Extraction In *Proceedings of the Third International Conference on Statistical Analysis of Textual Data.*, pp. 297-308, 1995.
- [12] 学術情報センター. 学術情報センター要覧, 1998.
- [13] Kyo Kageura, Teruo Koyama, Masaharu Yoshioka, Atsuhiko Takasu, Toshihiko Nozue, and Keita Tsuji. NACSIS corpus project for IR and terminological research. In *NLPRS '97*, pp. 493-496, 1997.
- [14] Noriko Kando, Teruo Koyama, Keizo Oyama, Kyo Kageura, Masaharu Yoshioka, Toshihiko Nozue, Atsushi Matsumura, and Kazuko Kuriyama. NT-CIR : NACSIS test collection project. In *The 20th Annual Colloquium of BCS-IRSG*, 1998.

付録 1: NACSIS コーパスにおける品詞・形態素カテゴリ

品詞表		活用表	
品詞	記号	品詞	記号
名詞	N	接頭辞	PP
普通名詞	NN	名詞接頭辞	PPNN
サ変名詞	NS	ナ形容詞接頭辞	PPJNA
数詞	NC	動詞接頭辞	PPVB
数式	NM	接尾辞	TL
固有名詞	NP	名詞性名詞接尾辞	TLNN
人名	NPH	サ変名詞性名詞接尾辞	TLNS
地名	NPG	名詞性名詞助数辞	TLNC
形式名詞	NF	名詞性述語接尾辞	TLNPD
時相名詞	NT	サ変名詞性述語接尾辞	TLNSPD
副詞の名詞	NR	名詞性特殊接尾辞	TLNSP
動詞	V	形容詞性名詞接尾辞	
母音動詞	VA	イ形容詞アウオ段	TLJNAUO
子音動詞		ナ形容詞	TLJNNA
カ行 5 段	VKA	ナノ形容詞	TLJNNN
サ行 5 段	VSA	形容詞性述語接尾辞	
タ行 5 段	VTA	イ形容詞アウオ段	TLJJAUO
		ナ形容詞	TLJJNA
ワ行 5 段	VWA	ナノ形容詞	TLJJNN
来型変格	VKI	動詞性述語接尾辞	
カ行変格	VKH	母音動詞	TLVA
カ行促音便活用	VKN	子音動詞	
サ行変格	VSH	カ行 5 段	TLVKA
ザ行変格	VZH	サ行 5 段	TLVKA
ワ行文語音便	VWB	タ行 5 段	TLVTA
形容詞	J		
イ形容詞イ段	JJ	ワ行 5 段	TLVWA
イ形容詞イ段特殊	JIS	来型変格	TLVKI
イ形容詞アウオ段	JAUO	カ行変格	TLVKH
ナ形容詞	JNA	カ行促音便活用	TLVKN
ナ形容詞特殊	JNS	サ行変格	TLVSH
ナノ形容詞	JNN	ザ行変格	TLVZH
タル形容詞	JTR	ワ行文語音便	TLVWB
助動詞	MD	得る型(うる型)	TLVUR
イ形容詞イ段	MDJI	ます型	TLVUMS
イ形容詞アウオ段	MDJAUO	接続詞	CC
ナ形容詞	MDJNA	指示詞	DM
ナノ形容詞	MDJNN	名詞形態指示詞	DMNN
助動詞ぬ型	MDNU	副詞形態指示詞	DMRB
助動詞く型	MDKU	連体詞形態指示詞	DMRT
助動詞だろう型	MDDR	副詞	RB
判定詞型	MDDT	連体詞	RT
助詞	S	特殊	L
格助詞	SCA	括弧始	LKS
接続助詞	SCC	括弧終	LKE
終助詞	SFL	句点	LSD
副助詞	SSB	読点	LRD
判定詞	DT	空白	LBK
感動詞	EX	記号	LSY
		形容詞活用系列	
		ダ列	_DA
		デアル列	_DR
		語幹	st
		未然形	mi
		基本形	bs
		意志形	wl
		命令形	od
		基本条件形	cd
		基本連用形	ry
		基本連体形	rt
		基本推量形	gs
		文語未然形	mib
		文語連体形	rtb
		文語基本形	bsb
		特殊連体形	rts
		タ形	ta
		タ系条件形	cdta
		タ系連用テ形	rytate
		タ系連用タリ形	rytatr

活用しない語と形態素：品詞表のコード (例：サ変名詞 NS)

活用する語と形態素：品詞表のコード+活用表のコード (例：ナ形容詞ダ列連用形 JNA_DArY)

複合語が活用形によりコードに存在しない場合：品詞を示す品詞表のコード+活用形を示すコード (例：行うように VJNA_DArY 全体の品詞は動詞で、最後の複合要素である助動詞の活用形がナ形容詞ダ列の活用)。