自然言語処理 120－6
(1997. 7. 24)

日英機械翻訳のための単語辞書

横尾昭男*1 宮崎正弘*2 池原 悟*3 白井 諭*1 阿部さつき*4

*1 NTT コミュニケーション科学研究所
〒239 神奈川県横須賀市光の丘1-1
{ayokoo, shirai}@cslab.kecl.ntt.co.jp

*2 新潟大学 工学部 情報工学科
〒950-21 新潟市五十嵐２の町8050
miyazaki@info.eng.niigata-u.ac.jp

*3 鳥取大学 工学部 知能情報工学科
〒680 鳥取市湖山町南4-101
ikehara@ike.tottori-u.ac.jp

*4 NTTアドバンステクノロジ株式会社
〒244 横浜市戸塚区川上町90-6 東戸塚ウエストビル
satsuki@totsuka.ntt-at.co.jp

あらまし

日英機械翻訳における高品質な意味解析を実現するため，語彙に関する知識を収録する単語意味辞書を構築するための単語の収録基準や収集方法を提案し，それに基づいて構築した単語辞書について報告した。日本語の単語は，様々な字種で表現され，活用や表記上のゆらぎがあり，それらに対応する基準を定めた。この基準に基づき，一般語１２万語，固有名詞２０万語，専門用語５万語，その他３万語の合計４０万語を収録した単語辞書を構築した。収録した各単語には，表記情報や品詞などの文法情報の他，約３０００属性からなる単語意味属性を付与した。この単語辞書は，日英機械翻訳システム ALT-J/E の解析処理で使用されている。

キーワード 機械翻訳，日英翻訳，意味解析，単語意味辞書，語彙

Semantic Word Dictionaries for Japanese-to-English Machine Translation

Akio YOKOO *1 Masahiro MIYAZAKI *2 Satoru IKEHARA *3 Satoshi SHIRAI *1 Satsuki ABE *4

*1 NTT Communication Science Labolatories
1-1 Hikari-no-oka, Yokosuka 239, JAPAN
{ayokoo, shirai}@cslab.kecl.ntt.co.jp

*2 Faculty of Engineering, Niigata University
8050 Ikarashi 2-nocho, Niigata 950-21, JAPAN
miyazaki@info.eng.niigata-u.ac.jp

*3 Faculty of Engineering, Tottori University
4-101 Koyamacho-minami, Tottori 680, JAPAN
ikehara@ike.tottori-u.ac.jp

*4 NTT Advanced Technology Corporation
90-6 Kawakamicho Totsuka-ku, Yokohama 244, JAPAN
satsuki@totsuka.ntt-at.co.jp

Abstract

In order to realize high-quality semantic analysis in Japanese-to-English machine translation, we proposed the criteria to select and collect words to build the semantic word dictionaries. We collected a total of 400,000 words, including 200,000 proper nouns and 50,000 technical terms, based on these criteria. We gave each word morphological information, semantic information such as part-of-speech, and semantic categories from about 3,000 attributes. These dictionaries are used in the semantic analysis process in the machine translation system, ALT-J/E.

key words machine translation, Japanese-to-English translation, semantic analysis, semantic word dictionary, vocabulary

1 はじめに

機械翻訳システムにおいて，高品質な訳文を得るには，言語表現中の単語がどのような約束（語義）で使用されているかを正確に決定できる質の良い意味解析が要求される。そのためには，現実の言語表現上での用法に関する知識を計算機処理に可能な形に整理する必要がある。この考えに基づき，筆者らは，まず，対象を概念化する際の視点を分類・整理することにより名詞の意味体系化を行うものとして名詞の意味属性体系を提案した[1,2]。そして，この意味属性体系をベースとして，日本語の形態素解析・意味解析に必要となる単語辞書や構文辞書[3]などを構築してきた。

本稿では，筆者らが研究開発を進めている日英機械翻訳システム ALT-J/E で実際に使用している「単語辞書」の構築に関し，単語辞書に収録すべき見出し語の収録条件，および，単語の収集方法について報告する。

2 見出し語収録条件

日本語の語は漢字，ひらがな，カタカナ，英数字など種々の字種で表記される。とくに，漢字には通常複数の読みがあり，熟語や固有名詞では特殊な読みを持つ場合も多く，同形語も少なくない。また，漢字は造語力が強く，複合語を自由に作り出せる。さらに，日本語においては，明確な正書法が確立されていないため，表記上のゆれがある。このような特徴をもつ日本語の語をどのような基準で単語辞書の見出し語として収録したかについて述べる。

2.1 一般語・固有名詞・専門用語などの収録基準

単語辞書に収録する語彙は，日英翻訳実験の対象として選んだ新聞記事などのような現代国語の記述文で使用される語を対象とし，以下のような考え方に基づき一般語１２万語，固有名詞２０万語，専門用語（電気電子・情報関連用語）５万語，その他（時事用語）３万語の合計４０万語を収録した。

(1) 一般語

一般語として収録した語の品詞はおおよそ以下の通りである。

名　詞：一般名詞，用言性名詞（サ変動詞型名詞，形容動詞型名詞），転生名詞（連用形名詞，～さ，～み，…），副詞型名詞（時詞，数詞など），連体詞型名詞，代名詞，形式名詞
動　詞：本動詞／補助動詞，自動詞／他動詞
形容詞：本形容詞／補助形容詞
形容動詞：ダ型／タルト型，形容動詞派生形態（～な，～に，～たる，～と）
副　詞：副詞派生形態（～と，～に，～だ，～する，･･･）
連体詞：指示連体詞（例：この），限定連体詞（例：ある），形容詞的連体詞（例：大きな），疑問詞的連体詞（例：どの）
接続詞：文接続詞，句接続詞
感動詞：例：ああ，はい
接　辞：接頭辞，接尾辞（表１に例示）
助動詞：受身，使役，希望，打消，過去，断定，推量，伝聞，様相の助動詞
助　詞：格助詞，副助詞，接続助詞，終助詞，準体助詞
記　号：文末記号，文節末記号，その他の記号

表１　接辞の分類

接頭辞	前置助数詞型	約，第，延，およそ
	否定型	無，不，非，未
	敬意添加型	御，ご，令，相
	純体言型	東，核，県，女
	連体詞型	各，全，同，本，元
	形容詞型	大，新，高，軽，快
	副詞型	再，最，既，仮，直
	動詞型	超，反，脱，対，過
接尾辞	助数詞型	回，毎，年，Kg，％
	助数詞承接型	目，台，強，弱
	数詞承接型	以上，以下，未満
	固有名詞承接型	県，駅，山，様，さん
	純体言型	さ，み，たち，者，機
	サ変動詞型	化，視
	形容動詞型	的，げ，そう
	連体詞型	性，風，用，型，式
	副詞型	前，後，間，中，上
	動詞型	がる，めく，つく，過ぎる
	形容詞型	らしい，（っ）ぽい，やすい

(2) 固有名詞

新聞記事などに現れる地名，人名，組織名，その他の固有名詞（地名，人名，組織名以外のとき，コト，モノの名など）を収録する。

(3) 専門用語

新聞記事（電気電子・情報関連）の専門用語（複合名詞などのような名詞が大部分を占める）を収録する。

(4) その他

新聞記事などに頻出する時事用語（複合名詞などのような名詞が大部分を占める）を収録する。

2.2 単語単位の扱い

日本語では，膠着言語の特徴の一つとして造語力が強く，漢字などを組み合わせることによって複合語が限りなく作り出される。とくに，専門用語や固有名詞には，複合語の形態をしたものが多い。このため，このような複合語のすべてをあらかじめ単語辞書に収録しておくことはできない。そこで単語辞書には原則として語基や接辞などの短単位語を収録し，派生語を含め複合語などの長単位語は，解析処理により短単位語の組合せに分割することとした[4]。ただし，以下のような語については，例外的に長単位で単語辞書に収録した。

(1) 短単位語の組合せに分割できない長単位語や部分（短単位語）から全体（長単位語）の意味や読みなどを合成できない慣用表現，熟語，複合語など。

(2) 3つ以上の語基が対等の関係で結合した並列語（例：松竹梅），並列語などの圧縮表現である縮退語（例：冷暖房＜＝冷房・暖房＞）。

(3) 国語辞典に子見出し語や派生語などとして収録されている一般用語。

(4) 格助詞相当の連語（格助詞相当語。例：によって，…），様相，アスペクト，テンスなどの法情報を表す連語（例：はずがない）。

(5) 以下のような複合固有名詞（全体が固有名詞となる複合語）。

- 日本の行政区画名（例：～都，～府，～県，～支庁，～郡，～市，～区，～町，～村。短単位の「～」の部分も収録）
- 自然地形名（例：～山，～川，～湖，～島，～海，～湾）
- 施設名の一部（例：～塔，～寺，～鉱山，～温泉，～道，～通り，～橋）
- 有名人名（「姓＋名」の形式のものなど）
- 組織名の一部（例：～省，～庁，～大会，～審議会，～条約）
  「～銀行，～商事」などの形式の企業名は長単位では収録せず，「～」の部分（短単位の固有名詞＜企業名＞）のみ収録
- その他の固有名詞の一部（例：～事件，～祭，～法，～地震，～賞）

(6) 専門用語の複合語。

2.3 活用語の扱い

べた書きされる日本語文の形態素解析を効率的に行うため，活用語は，通常以下の2つの形態で単語辞書に収録した。

(1) 規則的な活用を行うもの

不変化部分と変化部分に分離し，別見出し語として単語辞書に収録する。五段活用型動詞や形容詞では語幹が不変化部分，活用語尾が変化部分となる。ここで，形態素解析における1文字単語候補を減少させるため，ひらがな1文字語幹の五段活用型動詞（例：あく，あう）は，すべての活用形を単語辞書に収録した。一段活用型動詞では動詞連用中止形が不変化部分となり，「る」，「れ」，「ろ」，「よ」など少数の語尾が変化部分となる。一段動詞の未然形，連用形は不変化部分と同形のため，これらの活用形を不変化部分から，処理により生成する必要がある。

(2) 不規則な活用を行うもの（助動詞，カ変動詞，サ変動詞など）

すべての活用形を単語辞書に収録する。なお，「行く → 行っ」のように，例外的な活用形となるもの（カ行五段活用動詞はイ音便となる）については「行っ」のような形態（活用形）を単語辞書に収録した。また，形容詞のウ音便において，語幹が変化するもの（例：高＜タカ＞い → 高＜タコ＞う）については例外処理を必要とする。

2.4 表記のゆれの扱い

日本語においては，明確な正書法が確立されていないため，同じ語を漢字，ひらがな，カタカナ，およびその混ぜ書きといった様々な形式で表記したり，数詞を漢数字，算用数字，およびその混ぜ書きといった様々の形式で表記したり，送りがなのゆれがあるなど，表記上のゆれがある。

このようなゆれのある語を処理で対処しようとすると，すべての可能な形態を生成して辞書検索するため，単語辞書とマッチしない不必要な辞書検索が増加してしまう。一方，すべてのゆれを単語辞書に収録しようとすると，辞書の収録語数が膨大となってしまう。そこで，表記のゆれのある語のうちその代表形（標準表記）を設定し，辞書に記述することを基本とし，さらに以下のような辞書登録基準を設け，処理と組み合わせて表記上のゆれが吸収できるようにした。

(1) 送りがなのゆれ

すべての可能な形態を単語辞書に収録する。

例1）行う，行なう

例2）申込，申込み，申し込み

(2) 同一字種内の表記上のゆれ

すべての可能な形態を単語辞書に収録する。

例3）二葉，双葉

例4）コンベア，コンベヤ

例5）データー，データ

(3) 異なった字種間の表記上のゆれ

表記されることの多い形態を単語辞書に収録する。

例6）見る，みる

例7）鮎，アユ

例8）把握，は握

例9）パーセント，％

(4) 漢字の異体字と標準字体間のゆれ，カタカナ外来語の表記のゆれ

標準字体のみを単語辞書に収録しておく。異体字については変換テーブルで標準字体に変換すれば単語辞書の検索ができる。

例10）附属 → 付属

例11）ヴァイオリン → バイオリン

(5) 繰り返し記号に関するゆれ

原則として繰り返し記号を用いない形態で単語辞書に収録する。繰り返し記号のある語は，処理で繰り返し記号を対応する文字に置換したうえで単語辞書を検索すればよい。

例12）たゝみ → たたみ

例13）かゞみ → かがみ

例14）いちゝゝ → いちいち

例15）一歩々々 → 一歩一歩

ただし，「々」が連続しない場合（例：「日々」，「三々五々」，「佐々木」）は，例外的に「々」を用いた形態も単語辞書に収録する。従って，検索する際は，「々」を対応する漢字で置換せずに単語辞書を調べ，辞書検索に失敗した時，「々」を対応する漢字で置換して単語辞書の検索をやり直せばよい。これは，固有名詞などは通常「々」を用いて表記することが多いこと，「々」を漢字で置換することにより文字情報の一部が縮退すること（例：副詞「一々」は「一一」となり数詞と同形語となってしまう）などのためである。

2.5 同形語の扱い

品詞や読みの異なる同形語は，原則として別見出し語として単語辞書に収録した。なお，読みの同じ同形の固有名詞（例：清水<姓と地名>）は語数が多いことから，収録情報を圧縮して一つの見出し語として単語辞書に収録し，収録語数の削減による単語辞書のコンパクト化を図った。

2.6 用言性名詞や副詞派生形などの扱い

(1) 名詞化するサ変動詞語幹，形容動詞語幹

用言性名詞（サ変動詞型名詞，形容動詞型名詞）として単語辞書に収録する。用言化する場合は処理によりサ変動詞や形容動詞を生成すればよい。

(2) 連用形名詞

- 五段活用型動詞から名詞化したものは，動詞から連用形名詞を処理で生成しようとすると不変化部分（五段活用型動詞語幹）に活用語尾（連用中止形）を付加する必要があるため連用形名詞を単語辞書に収録する。
- 一段活用型動詞から名詞化したものは，単語辞書に収録せずに動詞不変化部分（一段活用型動詞連用中止形と同形，連用名詞化するか否かの情報をもつ）から処理により生成できるようにする。なお，動詞が名詞化した場合，元の動詞の意味から直接推定できない意味で使われるようになった語については，連用形名詞（例：流れ<川>）も単語辞書に収録する。

(3) 副詞派生形

副詞に「と」，「に」などが後接して副詞となるもの（例：はっきりと，すぐに）は単語辞書に収録しない。副詞語基に「と」，「に」などが後接すると副詞となるか否かを調べ，処理によりこのような副詞派生形を生成すればよい。

2.7 数表現の扱い

数表現は無限に生成されるため，その表現をすべて単語辞書に収録しておけない。そこで，以下のような数表現の基本要素（短単位）のみ単語辞書に収録する。したがって，数表現は解析処理で基本要素に分割すれば処理できる。

(1) 基本的な数詞「0，1，…9，〇，一，…，九，十，百，千，万，億，兆，京」，および不定数を表す「何，数，幾」

(2) 前置助数詞（例：第，約，延べ，計）

(3) 後置助数詞（例：本，個，枚，円，センチ，ｋｇ，％）

(4) 助数詞承接型接辞（例：強，弱，台，当り，以上，以下，未満）

(5) 数詞承接型接辞（例：以上，以下，未満）

(6) 数詞関連記号（例：位取りを表すカンマ「，」，小数点を表す「・」と「.」），および正負の区別を示す符号（例：「＋」，「－」）

2.8 辞書に登録されない語の機械処理の方法

すでに述べたように，ある一方の語から他方の語が処理によって生成できるような場合は，辞書容量の節約のため，他方の語はできるだけ辞書に登録せず，プログラム的な処理によって生成することとした。そこで，日本語解析を行う際，辞書検索の結果からこのような派生語をどのようにして生成して原文と照合するかについて，日英機械翻訳システムALT-J/Eの処理の例を付表に示す。

3 見出し語の収集方法

前節では，機械翻訳用の単語辞書として，どのような見出し語を含んでいるべきかについて述べた。各単語に付与されるべき情報は，出版されている人間用の辞書から収集できる部分もあるが，常識を期待できない計算機処理では，はるかに多くまた精密な情報が必要である。また，情報間の相互矛盾を防ぐには，機械翻訳システムに実装した上での様々な実験的検証が必要であった。本節では，実際にどのように辞書情報を収集してきたかについて述べる。

3.1 単語辞書の見出し語の収集

単語辞書を編集するため，単語辞書の見出し語となる一般語，固有名詞，専門用語などをどのように収集するかがまず問題になる。そこで，まず，単語辞書の見出し語を収集するうえでの問題点とその解決策について述べる。

(1) 一般語の収集

人間用の国語辞典には，様々な品詞の日本語の一般語が収録されている。しかし，人間用の国語辞典の見出し語セットは，日英機械翻訳用の辞書である単語辞書の見出し語セットとしては，必ずしも十分ではない。人間のもつ高度な知的能力（言語理解能力，推論・類推能力など）と膨大な知識を前提として編集されている人間用の国語辞典と異なり，このような高度の知的能力や知識をあてにできない計算機が使用する辞書では，見出し語セットの網羅性，完全性が要求される。計算機による文解析において，単語辞書にない語（未知語）を含む文は，人間のように未知語を正しく認識しその意味を推定することができず，解析に失敗することが多いからである。また，計算機による文解析では，計算機処理に適した形態で語を単語辞書に収録する必要がある。とくに，漢字かな混じりでべた書きされる日本語文では，文解析の入口の処理である形態素解析が重要な位置を占め，単語辞書は形態素解析で利用しやすいように構築されることが望まれる。

以上の点を考慮して，国語辞典の見出し語，子見出し，派生語をベースに以下のようにして不足する語彙を補充し，単語辞書の見出し語とした。

1) 国立国語研究所の新聞語彙調査（「現代新聞の漢字」）などに出現する基本語（短単位語）の中から，国語辞典には収録されていないが，単語辞書に収録したほうがよい語を抽出して見出し語とした。とくに，複合語の解析が行いやすいように，「的」，「さ」など本来，接辞とみなされるものの他に，「県」，「山」など自立的に使われるが，複合語の中に頻出し接辞的に使われる一字漢字（接辞性字音語基）も接辞として見出し語に加えた。

2) 短単位語の組合せに分割できない長単位語や部分から全体の意味や読みを合成できない長単位語を収集し，見出し語とした。

3) 格助詞相当語や法情報を表す連語など文解析で重要な役割を果たす付属語的な連語を収集し見出し語とした。

4) 活用語処理が行いやすいように，規則的活用を行う語を変化部分と不変化部分に分離し，不変化部分と全ての変化部分の形態，不規則的活用を行う語の全ての活用形などを見出し語とした。

5) 人間用のシソーラス（「分類語彙表」，「角川類語新辞典」）などに出現する基本語の中から，国語辞典には収録されていないが，単語辞書に収録したほうがよい語を抽出して見出し語とした。

6) 新聞記事（「日本経済新聞」）１年分の語彙統計を実施し，国語辞典には収録されていないが，単語辞書に収録したほうがよい基本語を抽出して見出し語とした。

(2) 固有名詞の収集

固有名詞については主として以下のように収集し，単語辞書の見出し語とした。

地名

- 「日本行政区画番号帳」から日本の行政区画名
- 「日本地名索引＜上，下＞」から行政区画名以外の日本の地名
- 「時刻表」などから日本のＪＲや私鉄の駅名
- 「地図帳」，「地名辞典」などから国際地域名，国名，外国の地名，日本の地方名や歴史的地名
- 「理科年表」などから天体名

人名

- 「日本の名字＜表記編，表音編＞」などから日本人の姓

・「日本人の名のデータ（電話帳統計）」などから日本人の名
・「人名辞典」などから歴史上の有名人名
・「各種人名録」などから現代の有名人名
・「角川類語新辞典」などから神仏名

組織名

・「会社四季報」などから日本の企業名
・「現代用語の基礎知識」，「新・記者ハンドブック」などから日本の官庁名，団体／党派名，国際機関名，条約名
・「大学・高専コード表」から日本の大学，短大，高専名

その他

・「歴史年表」などから年号，時代名，事件名
・「時刻表」などから一部の乗り物名（日本の列車，船舶，航空機，宇宙船等の愛称）
・「現代用語の基礎知識」，「新・記者ハンドブック」，「新聞記事１年分の語彙統計」，「教科書」などから行事名，言語名，宗教名，流派名，出版物名，法律名，制度名，民族・人種名，プロジェクト名，作品名，商品名，動物名，植物名，宝物名，現象名，その他

(3) 専門用語等の収集

電気電子・情報関連の専門用語を日英対訳の科学技術用語辞典（「科学技術用語２５万語辞典」）などから収集し，見出し語とした。

新聞記事に頻出する時事用語については，時事用語集（「現代用語の基礎知識」，「imidas」，「記者ハンドブック」，……）などから収集するとともに，(1)で述べた新聞記事１年分の語彙統計より単語辞書に収録したほうがよい時事用語も加えて見出し語とした。

3.2 辞書間の見出し語の整合性確保

日英翻訳システム ALT-J/E では，日英翻訳のため「単語辞書」，「構文辞書[3]」，「日英対照辞書」など複数の辞書が利用される。これらの辞書間で見出し語の整合をとるため，表記のゆれのある日本語の見出し語に対して，その代表形を標準表記として設定した。日本語の見出し語の表記のゆれは，単語辞書に記述された標準表記により解消され，標準表記は他の辞書を検索するときのキーとなる。

3.3 単語辞書における語の記述情報

単語辞書には，その語の品詞などの文法情報だけでなく，その語の意味属性や共起情報などのようなその語の意味や用法に関する情報も記述され，計算機による文の解析や生成には，このような情報が縦横に使われる。さらに，単語辞書の収録語数の増大に伴って単語辞書内の同形語や類義語が増大し，それらを判別する情報も必要となる。計算機による文の解析や生成の品質は，単語辞書などの各種の辞書に記述される知識の質と量に左右され，計算機で容易に扱えるような曖昧さのない明確な形式で記述されなければならない。

上記の点を考慮して，単語辞書に収録する情報の体系を，「表記情報」，「音韻情報」，「形態情報」，「文法情報」，「意味情報」などから構成することとした[5]。表２にその一部を示す。

表２　単語辞書に収録する情報（抜粋）

表記情報	見出し語の表記	・見出し語の字面
	標準表記	・見出し語の表記上の代表形
	読み	・見出し語のひらがな表記
音韻情報	連濁化情報	・見出し語の連濁化の可否
形態情報	文字情報	・漢字語／ひらがな語／カタカナ語／英数字語／混せ書き語の区別
文法情報	品詞・活用	・見出し語の品詞／活用型／活用行／活用形を示すコード
	品詞補足情報	・自／他動詞，本／補助動詞，引用の「と」，連用形名詞化の表示 ・形容動型名詞の「～な／～に／～たる／～と」形の可否 ・複合語内形容動詞化／複合語内連体詞化 ・副詞の「～と／～に／～で」形の可否 ・擬音語／擬態語 ・格助詞相当語／接続助詞相当語
	前方（後方）接続カテゴリ	・見出し語の前方（後方）に文法的に接続可能な品詞のグループを表すコード
	文型情報	・用言の必須格パターン（助詞の組み合わせ）
	ｱｽﾍﾟｸﾄ・ﾃﾝｽ属性	・継続／瞬間／状態などの区別
	様相属性	・意志／無意志動詞の区別
	接続属性	・接続詞・付属語的な語の接続属性
意味情報	意味カテゴリ	・見出し語の意味属性
位相情報	語種情報	・一般語／固有語／専門語の区別
	使用分野	・専門分野種別
重要度・統計的情報	重要度	・基礎語，最重要語／重要語
	統計的情報	・頻度統計結果による見出し語の出現頻度
同形語選択情報	優先属性	・同形語内での選択優先度
辞書検索制御情報	最長単語情報	・見出し語を包含する長い語がないことを表示
	同形最終語情報	・同形語の最終レコードであることを表示
	見出し語内単語連鎖情報	・見出し語に包含される単語列があることを表示

4 おわりに

本稿では，日英機械翻訳システム ALT-J/E で使用している単語辞書をどのようにして構築してきたかについて述べた。まず，様々な字種で表記され，活用や表記上のゆれのある単語を収録するための基準を定めた。そして，一般語，固有名詞，専門用語などを系統的・網羅的に収集し，各単語に形態・文法・意味などの情報を付与することにより単語辞書を構築してきた。

この単語辞書は，まだ構築の途中にある。見出し語数は，現在，約４０万語であるが，語彙数としてはまだ不足している。また，より深い意味解析や単語の解釈の曖昧性の解消のためには，辞書情報と解析処理の両方向から高精度化を行う必要がある。今後も，収録単語数，収録情報の種類の両面で拡張を図っていく予定である。

参考文献

[1] 池原, 宮崎, 横尾: 日英機械翻訳のための意味解析用の知識とその分解能, 情報処理学会論文誌, Vol.34, No.8, pp.1692-1704 (1993)

[2] 宮崎, 池原, 横尾, 白井: 日英機械翻訳のための意味属性体系, 電子情報通信学会, 言語理解とコミュニケーション研究会, NLC-97 (1997)

[3] 白井, 横尾, 中岩, 池原, 宮崎: 日英機械翻訳のための構文辞書, 電子情報通信学会, 言語理解とコミュニケーション研究会, NLC-97 (1997)

[4] 宮崎, 池原, 横尾: 複合語の構造化に基づく対訳辞書の単語結合型辞書引き, 情報処理学会論文誌, Vol.34, No.4, pp.743-754 (1993)

[5] 宮崎正弘: 辞書の記述と利用 —機械辞書の観点から—, 日本語学, Vol.14, No.4, pp.52-61(1995)

単語の収集にあたっては，本文中に記載した各資料を参照した。

付表　辞書にない単語の処理による派生について（１／２）

分類	単語派生の種類	単語を派生する方法	例と備考
記号	「記号」	記号は原則としてすべて処理で生成する。	例：＃，＆ 備考：％(パーセント)・＄(ドル)などのように辞書登録されている記号も存在するが，それらの品詞は接辞となっている。
名詞	数式 (「一般名詞」)	数式は全体で一つの単語として，文字列から処理で生成する。	例：１＋１＝２ 備考：全体で「一般名詞」
	数詞 (「数詞」)	数詞は全体で一つの単語として，文字列から処理で生成する。	例：２千３百 備考：全体で「数詞」
	連用形名詞の生成	連用形名詞化可能な一段動詞語幹に対し連用形名詞を生成する。	例：届け(「他動詞」) → 届け(「転生名詞」) 備考：なお，五段動詞の連用形名詞は辞書登録されている。
	連体詞型名詞の生成	複合語内で形容動詞型の係り受けが成立した場合，キー単語の品詞を連体詞型名詞に補正する。キー単語が元から連体詞型名詞や形容動詞型名詞である場合を除く。	例：近代(「時詞」)/美術(「一般名詞」) → 近代(「一般名詞・連体詞型」)/美術(「一般名詞」) 備考：『近代』と『美術』の間に形容動詞型の係り受けが成立するので，『近代』の品詞を連体詞型名詞に補正する。
	固有名詞の生成	複合語内係り受け検定が成立した場合などに，品詞の多義を解消する。	例：真弓(「固有名詞」)/明信(「固有名詞・名」) → 真弓(「固有名詞」)/明信(「固有名詞・名」) 備考：『真弓』の品詞は「姓」，「名」，「地名」の多義があるが，『明信』(「名」)との係り受けが成立することにより，品詞を「姓」に決定する。
	形式名詞『の』の生成	準体助詞の『の』の品詞を形式名詞に変更する。	

付表　辞書にない単語の処理による派生について（2／2）

分類	単語派生の種類	単語を派生する方法	例と備考
動詞	一段動詞の未然形・連用形の生成	一段動詞語幹に対し，未然形および連用形を生成する。	例：調べ(「他動詞」) → 調べ(「他動詞・未然形」)， 調べ(「他動詞・連用形」)
	副詞＋『する』のサ変動詞化	サ変動詞化可能な副詞の直後にサ変動詞語尾がある場合，全体を一語化し，品詞をサ変動詞にする。	例：はっきり(「副詞」)/する(「他動詞」) → はっきりする(「自動詞」)
	動詞の語幹語尾の一語化	動詞の語幹と語尾を一語化する。また，サ変名詞＋サ変動詞語尾も一語化する	例：動(「自動詞」/く(「自動詞」) → 動く(「自動詞」) 調査(「用言性名詞」)/する(「他動詞」) → 調査する(「他動詞」)
形容詞と形容動詞	副詞＋形容動詞語尾の形容動詞化	形容動詞化可能な副詞の直後に形容動詞語尾がある場合，全体を一語化し，品詞を形容動詞にする。	例：たいそう(「副詞」)/な(「形容動詞」) → たいそうな(「形容動詞」)
	連体詞の形容詞化	連体詞相当語の連体詞（形容詞が連体詞化したもの）に対し，品詞を形容詞にし，標準表記も修正する。	例：大きな(「連体詞」) → 大きい(「形容詞」) 備考：形容詞形が辞書登録されているかどうかには関係なく派生させる。
	形容詞・形容動詞の語幹語尾の一語化	形容詞・形容動詞の語幹と語尾を一語化する。また，形容動詞型名詞＋形容動詞語尾も一語化する。	例：赤(「形容詞」)/い(「形容詞」) → 赤い(「形容詞」) 偉大(「形容動詞」)/だ(「形容動詞」) → 偉大だ(「形容動詞」) 完璧(「用言性名詞・形容動詞型」)/だ(「補助動詞」) → 完璧だ(「形容動詞」)
副用語 副詞と連体詞	自立語＋格助詞の連体詞化と副詞化	自立語(名詞・副詞・連体詞)と直後の助詞の組合わせが以下の条件を満たす場合，一語化して品詞を副詞または連体詞にする。 副詞/副詞型名詞 ＋ と → 副詞 副詞/副詞型名詞 ＋ に → 副詞 副詞/副詞型名詞 ＋ で → 副詞 連体詞型名詞 ＋ の → 連体詞	例：がたがた(「副詞」)/と(「格助詞」) → がたがたと(「副詞」) 一斉(「副詞」)/に(「格助詞」) → 一斉に(「副詞」) 至る処(「一般名詞・副詞型」)/で(「格助詞」) → 至る処で(「副詞」) 大型(「一般名詞・連体詞型」)/の(「格助詞」) → 大型の(「連体詞」)
	動詞連用形＋『て／で』の副詞化	動詞連用形＋接続助詞『て／で』かつ動詞が動詞転成副詞の場合，一語化して品詞を副詞にする。	例：急い(「自動詞」)/で(「接続助詞」) → 急いで(「副詞」)
独立語	独立語（接続詞・感動詞）では，処理で作成する単語はない。		
接辞	一段動詞の未然形・連用形の生成	一段活用型接尾辞に対し，未然形および連用形を生成する。	例：染み(「接辞・他動詞型」) → 染み(「接辞・自動詞型・未然形」)， 染み(「接辞・自動詞型・連用形」)
	体言型接尾辞の品詞多義の解消	複合語内係り受け検定が成立した場合などに接尾辞品詞の多義を解消する。	例：西明石(「固有名詞」)/駅(「接尾辞」) → 西明石(「固有名詞」)/駅(「接尾辞」) 備考：『駅』の品詞は「固有名詞承接型」，「純体現型」の多義が『西明石』(「地名」)との係り受けが成立することにより，品詞を「固有名詞承接型」に決定する。
付属語	付属語（助詞，助動詞）では，処理で作成する単語はない。		