ＰＩＶＯＴ　Ｊ－Ｅ：日本語複合語解析

2B-3

宮部　隆夫*　坂井　信輔*　谷　健一*　川端　麻友美**

*日本電気　株式会社　**日本電気技術情報システム開発　株式会社

1　はじめに

本稿では、多言語機械翻訳システムＰＩＶＯＴの日本語形態素解析、特に複合語解析を紹介する。複合語とは、複数の基本的な単語の組合せから構成される単語であり、容易に新しい語が加わる。従って、複合語全てを網羅することは不可能であり、複合語を構成している単語の意味とその構造から、複合語全体の意味を推定することが必要となる。しかし複合語を構成する基本的な単語は多義であることが多く、しかも複合語は、活用などの形態的特徴が少ない語幹の形や名詞の組合せから構成され、構造推定は難しい。複合語には係り受けを利用した解析手法[1]が知られているが、複合語の意味を認定するためには係り受け構造だけでなく、構成語の語義推定や意味関係の指定が必要となる。以下、形態素解析全体の中で複合語解析の位置を示した後、上記問題に対応する複合語解析について説明する。

2　形態素解析の概要

形態素解析の目的は、各語彙の持つ情報に基づいて、文節を認定し、その文節内の構造を確定することにある。その主要部は、複数の形態素を組合せて単語を合成するフェーズとその単語を組合せて文節を合成するフェーズの２フェーズよりなる。前処理としての辞書引き・語切り(segmentation)の後、単語・文節構造を確定し、その後部分的な構文構造を推定する。

語切り(Segmentation)、辞書引き　可能な語切りの選択肢の中から、語の長さ、文字種、隣接しやすさ等の情報を使って、最も確からしい切り方を選択する。[2]

単語構造確定　単語構造を確定するために、接辞と活用（語尾）により品詞を確定する。隣接条件を中心にして、語彙の選択を行い、選ばれた形態素の組合せから単語を合成する。

文節構造確定　単語構造の確定後、単語同士の結合により文節構造を確定する。用言と助動詞により用言句を合成し、接辞や名詞を結合して複合語を合成し、さらに、それら自立語（句）と助詞との結合により文節構造を確定する。用言句合成では、テンス、アスペクト、モダリティ、意志性等の意味情報により、句全体の意味計算と語義選択を同時に行う。[3] 他方、複合語では語の用法情報により、構成語の語義を選択し複合語構造を確定する。

構文構造推定　文節確定後、助詞の種類や一部副詞を用いて文中の係り受けを部分的に推定する。非交差条件を満足する範囲で、並立助詞の種類や文パタンを基にして並立句構造を推定し、副助詞・呼応の副詞を利用して呼応表現中の副詞と用言句を対応付け、さらには、接続助詞の種類により用言句の係り受けを推定する。[4]

3　複合語解析

複合語の構造認定とは語を構成する多義的な語彙の組合せから、適切な語義の組合せを選択しその意味関係を推定することにある。従来、構造認定の鍵として、接辞の語彙情報が重視されてきたが、その情報をより広範囲な一般語彙に拡張し、語用の情報として整理する。その結果を表１に示すように、その働きの差に基づいて、①単語区間指定　②係り受け指定・語義選択　③意味構造の指定の３種類に分類した。

単語区間指定　単語区間推定の代表例として、固有名詞と数詞表現がある。表１の＜単語区間指定＞の例を参照すると、「さん」はその前に固有名のみを、「ｋｇ」は数詞を必要とする。区間内の語についてはその品詞内容を固有名あるいは数詞と指定する。この区間推定の機能は、「県－市－区」や「日－時－分」といった包含関係を持つ語についてはより強力となる。「川崎市宮前区」という表現では、市＞区の結び付きより、複合語全体が地名となり、「宮前」、「川崎」はその固有名称となる。「川崎市」宮といった誤った解析は有り得ない。また、「１時１０分」と「１時間１０分」についても、「時」と「時間」とにより、それぞれが時刻と時間として意味を一意に確定し、「分」の多義は解消される。

係り受け・語義選択　複合語内の係り受けは、個々の語彙情報によって拘束される場合がある。また、多義語については、その語義の差が係り相手の内容と対応して決まることが多く、係り相手の条件により語義選択をする。表１の＜係り受け・語義選択＞に示したように、係り受けに必須な条件、相手にあってはならない禁止条件の他に、その語義の選択の適切さを示す優先条件があり、その条件の充足度合が語義選択の尤度を表す。係り受けに関する受けの情報として語用品詞を設け、物名型、固有名型、形容動詞型等の品詞に準じた値を持たせる。先の係り受け相手の条件としては、意味と共にこの語用品詞が有効である。「係り受け相手条件」の例の「達」は前接語が普通名詞か固有名詞かによって意味が異なる。固有名詞の場合は、「山田」さんとその他の人、普通名詞の時はその複数形となる。

係り先に関しては、表１の例の「超（高速）」「（作成）者」のように、原則として隣接形態素に係るもの（形態素型）が多いが、「当」「的」のように、単語（複合語）に係るもの（単語型）もある。この単語に係る場合は、形態素から複合語を構成して、その語用品詞を確定した後、その語との係り受けを定める。

PIVOT J-E: Japanese Compound Word Analysis

Takao MIYABE　Shinsuke SAKAI　Ken-ichi TANI
NEC Corporation　〃　〃

Mayumi KAWABATA
NEC Scientific Information System Development,Ltd.

表1　語用法情報

項目	内容　用例
< 単語区間指定 >	区間の指定 ：（前方） 品詞の指定 ： 固有名詞、数詞
単独区間指定 （区間と品詞）	（EX.） 山田さん　１０ｋｇ
包含型の 区間指定 （区間と品詞） （包含関係）	区間の指定 ：（前方） 品詞の指定 ： 数詞、固有名詞 包含の共有軸 ：地名、時刻、等 包含の大小関係： 県 > 市 > 区
	（EX.）川崎市宮前区 １時１０分　１時間１０分
<係り受け・ 語義選択 >	語用品詞： 形容動詞型、固有名型、等 （係り受けの、受けのカテゴリで品詞に準じて設定）
語用品詞 （受けのカテゴリ）	（EX.） 高温（形動）　ＮＥＣ（固名）
係り受けの 相手選択条件 係受け決定条件 語義選択条件	係り受け相手語への条件 条件の種類： 必須：　必ず必要 禁止：　禁止 優先選択：該当するほど、 係り受けとして適切 （語義選択条件） 条件の内容： 語用品詞、意味、形態、等
	（EX.）超　超[高温／人] 達　[子供／山田さん]達
係り（受け）先 形：形態素型 語：複合語型	係り／受け相手の種類（２種類） 形（形態素型）：　隣接形態素 語（複合語型）：（複合）語
	（EX.）（形）　超（高速）計算機 データ（作成）者 （語）　当（高速計算機） （データ駆動）的 （）内の語が係り受けの相手
< 意味・構造指定 >	構造の型(係受け×意味構造)10種 係り受け(方向)：係り、受け2種 意味的な構造：(5種) 意味的なヘッド 意味的なディペンダント 関係 情報吸収 情報伝達
構造指定 10(=2*5)種	
	（EX.）作成者　防火施設　高圧 マスク付き演算　１対１ 新橋・横浜間 ２点　美しさ
意味指定	意味の指定 関係の指定：意味関係 情報の変更：品詞、意味 情報の付加：意味 意味；否定、アスペクト等
	（EX.）　不確定　未発表

意味指定　意味推定には、構造を指定する場合と意味関係、情報を指定する場合がある。構造指定については、係り受けと意味関係の２種類がある。係り受けでは係と受けの２種類があり、意味関係では意味的なヘッド、ディペンダント、意味関係自身、及び、情報吸収、情報伝達の５種類があり、組合せで１０種類の構造が存在する。表１の＜意味・構造指定＞中の例を参照すると、通常は「作成者」のように、後の語が係り受けの受けとなるが、「防火施設」の「防」のように左側の語が受けになる例もある。「マスク付き演算」は、「マスク」と「演算」との関係を表す場合であり、「さ」や「不」のように結合相手の単語に吸収されて、相手の品詞や意味を変化する場合もある。

意味指定では、表中の「不」のように語用品詞変更したり、否定の意味を付加するものと、「未」のように副詞的な意味関係を表すものがある。

以上の情報を使った解析例を、表２に示す。

表2　複合語解析例

原文	川崎市宮前区の不特定の .. 高温の .. １対１の ..　１０円高で
語切り	川崎/市/宮/前/区/の/不/特定/の .. 高/温/の ..１/対/１/の .. １０/円/高/で
語区間	(「川崎」市「宮前」区) の　..　「固有地名」
選択	[高]温の ..　１０円[高]で (高)-<属性>-(温)の　(１０円)-<値>-(高)
構造	(１-[対]-１)の　(水素２)-[対]-(酸素１)
意味	(不特定) ..[否定]

（ ）：語域、　< >：関係名　：係受け、
/ ：分かち書き　[] ：演算対象

4　おわりに

本稿では、語彙情報に基づいて文節を認定する形態素解析を紹介した。特に複合語について、単語区間、係り受け、語義および構造推定のために語用法の情報の情報を定義し、その内容について説明した。この語彙的な語用法の情報を使うことにより、複合語内の構造、意味推定が可能となった。

参考文献

[1] 宮崎　：「係り受け解析を用いた複合語の自動分割法」情処論文誌　Vol.25. No.6 ('84)
[2] 坂井他：「PIVOT J-E: 日本語形態素分解」　６３年度信学会全大
[3] 谷他　：「　〃　日本語多義助動詞の意味推定」〃
[4] 亀井他：「Lexical Discourse Grammar の提案」　信学技報　Vol.86 No.189 NLV86-7 ('86)