ＰＩＶＯＴ：日本語接辞解析

6E-3

坂井信輔　　宮部隆夫　　村木一至

日本電気（株）Ｃ＆Ｃ情報研究所

１．はじめに

日本語には、助詞を伴わない漢字語の連続による複合語が多く見られるという特徴がある。その中でも、接頭辞、接尾辞による造語能力は、他の言語と比べて強い。しかしながら、接辞の表記は一般に多義的であり、接辞を語彙として扱う際には、多義解消処理が重要となる。本稿では、日英機械翻訳システムＰＩＶＯＴにおける日本語の接辞の処理に関して、中間構造、結合の性質、多義解消に用いる情報について述べ、解析アルゴリズムを説明する。

２．接辞処理の問題

単独で自立語として用いられることがないが、他の自立語と結合して複合語を作る能力のある形態素を接辞であるとする。すると、機械翻訳というタスクを考えると、日本語では、接辞として扱われるが、英語などのターゲット言語においては、独立の単語となるような語彙が多く存在し、名詞、動詞などの自立語と比較して、日本語の接辞は、ターゲット言語の様々な品詞・表現に対応することが分かる（図．１）。また、日本語の１文字～２文字漢字には１つの表記が様々な語用をもつという多義性の問題があり、この複数の候補の中から１つを選択する必要がでてくる（図．２）。

排他【性】	exclusiveness
取引【高】	the amount of the transaction
カナダ【製】の机	a desk made in Canada

図．１　接辞を含む表現の対応

(a)	売上【高】	the amount of sales
(b)	浦和【高】	the Urawa high school
(c)	ドル【高】	a strong dollar
(d)	【高】純度（の）	(of) high purity

図．２　接辞の表記の多義性

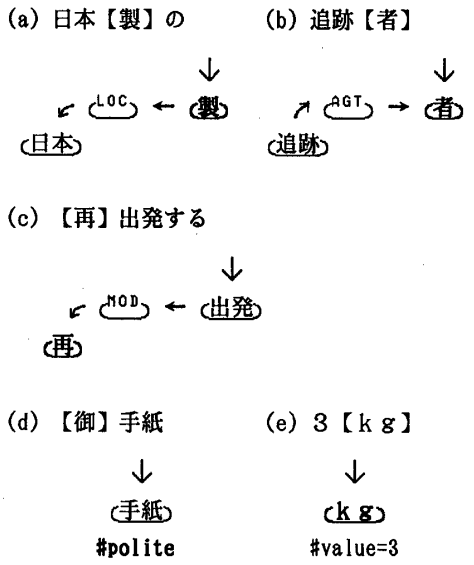
３．接辞と中間表現

われわれが扱う接辞は、結合の方向によって 接頭辞・接中辞・接尾辞に分かれる。解析の結果得られる中間表現[1]における接辞の役割には、次に挙げられるいくつかのものがある：

(1) 結合相手と格関係（あるいは意味関係）を結ぶ（図３．(a) (b) (c)）

(2) 結合相手に属性を付与する（図３．(d)）

(3) 結合相手を属性として取り込む（図３．(e)）



図．３　中間表現における接辞の役割

４．接辞の結合の性質

複合語中の語と語、あるいは語と接辞の結合は、一種のかかり受け関係であり、複合語中のある接辞は、その中のある特定の語と結合する。結合の相手は、必ずしも隣接の語とは限らず、その中の部分的な複合語である場合がある。例えば、図．４(2)の「大」は、隣接語と結合するが、(3)の「的」は、「総合／工学」という複合表現全体と結合する。そこで、接辞の結合の強さとして基本的に３つのレベルを設定し、接辞の辞書情報とした。また、この他に、「本」や「当」などの指示的な接頭辞、「兼」、「対」などの接中辞については、特別な扱いをするので、これらも結合の性質に関する独立のカテゴリーとして辞書に記述している（図．４）。

ＰＩＶＯＴ：Analysis of Japanese affixes

Shinsuke SAKAI, Ken-ichi TANI, and Takao MIYABE
C&C Information Technology Research Laboratories, NEC Corp.

(1) 【無】／給油／式　　(2) 【大】／容量／記憶

(3) 総合／工学／【的】　　(4) 【本】／記憶／装置

(5)宿舎／【兼】／調査／会場

(1) 他の接辞より結合度の強いもの。
(2) 単語と単語の結合よりも先に結合する。
(3) 単語と単語の結合よりも後に結合する。
(4) 指示的な接頭辞
(5) 接中辞

図4．接辞の結合の性質の分類

5．多義解消のための辞書情報

接辞の表記には、図．2に示したように、多義性が多く見られ、この中からの選択を精密に行なう必要がある。例えば「高」という表記が何をあらわすかを決定する場合、左側の語が《行為》の意味をもつ場合はその行為の結果できたものの量をあらわし、《地名》の意味をもつ場合は、その地名を名前とする高校をあらわすことが推定できる。また、右側の語が《属性》の意味をもつ場合は、その属性値が「高い」ことをあらわすと推定できる。このような隣接語に対する選択情報を同表記のおのおのの接辞（および単語）の辞書情報としてもつことにしている。また、接辞には、複合語内での共起により、それらの接辞および、結合相手の多義解消が行えるものがある（図．５）。これらには、順序をもって共起する接辞のグループ名と、その中での出現順を、辞書情報として与えている。この他に、各接辞に対して、結合相手が必ず満たさなければならない条件があれば、それを辞書情報として記述している。

(1) ３時１０【分】　　時刻
(2) ３時間１０【分】　　時間

図5．複合語内での接辞の共起による多義解消

6．解析アルゴリズム

接辞の解析の手続きを以下に示す。この手続きへの入力は、形態素分割処理[2]の結果であり、各区間に複数の単語あるいは接辞の候補をもつ分割区間の系列となっている。

(1) 否定型接頭辞の同定、結合処理
「無」、「未」などの否定型接頭辞は、結合度が強く、結合の結果できる（部分的な）複合語のもつ情報が、その他の語や接辞の多義解消に必要なことが多い。よってこれらとその結合相手の語に対して、まず隣接語に対する選択情報を用いて多義解消を行い、結合処理を行なう。

(2) 共起する接辞の多義解消マーキング
共起する接辞のグループに属する接辞の辞書項目が存在するならば、それと複合語内で適切な順序で共起する同グループの他の接辞を探し、共起するものが存在すれば、選択のためのポイントを与える。

(3) 隣接関係による多義解消マーキング
接辞の辞書項目に記述されている選択条件、必要条件を左右の隣接語が満たしているかをチェックして、選択のためのポイント付けを行なう。

(4) 接中辞の同定、結合相手マーキング
接中辞の多義解消を行なう。もし、接中辞が選択されたならば、その左側隣接語の性質を用いて右側の語の結合相手を決定する。

(5) 指示的接頭辞の同定、結合相手マーキングを行なう。

(6) 隣接関係によるマーキングを用いた多義解消
複合語内の各区間において、これ以前の処理において行なわれたポイント付けの結果から、もっとも得点の多いものを選択する。

(7) 結合処理
強結合型接辞、一般複合語、弱結合型接辞、共起型接辞、指示型接辞、接中辞の順に結合処理を行なう。

7．おわりに

ＰＩＶＯＴにおける日本語接辞解析について述べた。今後、ここで明らかにした接辞の隣接語に関する制約をもちいて、複合語の分割の精度を向上させて行きたいと考える。

最後に、有益な御教示をいただいたＣ＆Ｃ情報研・谷健一氏に感謝します。

参考文献

[1] 市山他：PIVOT：中間表現　情処全大38 '89年3月
[2] 坂井他：PIVOT J-E:日本語形態素分解 信学全大 '88年秋季,D-138