

第4章 形態素解析

柏野和佳子 131

1. 形態素解析の必要性	131
2. ことばの単位	133
さまざまな単位…134／短めの単位の特徴…135／長めの単位の特徴…136／『現代日本語書き言葉均衡コーパス』で採用した単位…137	
3. 形態素解析処理の手法	137
規則に基づく手法…137／統計・確率に基づく手法…140	
4. 形態素解析ツール	141
JUMAN…145／茶筌…147／MeCab…156／「茶筌 + UniDic」と「MeCab + UniDic」用の「茶まめ」…158	
5. 形態素解析システム用の辞書	161
品詞分類…161／JUMANの辞書…166／茶筌の辞書…170／MeCabの辞書…172／UniDic(茶筌, Mecabの辞書)…174	
6. 形態素解析の活用のために	182

第5章 機械翻訳ソフト

杉浦滋子 187

1. はじめに	189
調査の概要…187／機械翻訳ソフトの性質…188／句読法と機械翻訳の精度…189／文字の表記…200／固有名詞…202／言語間の形態面での対応のずれ…203／言語間で形態面に対応する要素に見られる意味的なずれ…209	
2. 機械翻訳の実際	210
英語…210／韓国語…219／中国語…224	
3. 結び	230

第1章

荻野 綱男

ソフトウェア概説

1. はじめに

コンピュータを使うことは事実上ソフトウェア（ソフト）を使うことである。なぜならば、ソフトを入れ替えて、それによってさまざまな働きをさせることができがコンピュータの本質だからである。

パソコンの場合も同様であり、どんなパソコンを使うかよりも、どんなソフトを使うかのほうがはるかに重要な問題になる。

ソフトには、いろいろな種類があるが、ここではアプリケーションソフトに限定して話を進めることにする。言いかえるとOSに関わる話はしないということである。OSとアプリケーションソフトの関係については第1巻第1章に記述してあるが、パソコンと一緒に動くのがOSであり、OS上で具体的なさまざまな処理を行うのがアプリケーションソフトである。

日本語学を中心とする言語研究者がパソコンを使う場合、どんな使い方をするかというと、大きく3つの使い方がある。1つは論文を書くときに使うということであり、ワープロソフトが典型である。2番目には、日本語に関するデータを処理するときに使うということであり、表計算ソフトや統計ソフトなどが典型である。そして3番目には各種調査のためにインターネットを使うということであり、ブラウザソフトが典型である。

ワープロソフトが扱う対象は「文書」であり、表計算ソフトや統計ソフトが扱う対象は「データ」である。その意味で両者は性格が違うのだが、パソコンの中では文書もデータもファイルの中に格納されているので、ファイル

を扱うソフトは、2つの使い方の両方に関係してくる。ソフトの中にはユーティリティソフトと呼ばれる一群のソフトがある。これらはユーティリティ、つまり「道具」であり、主としてファイル関連の処理を行うようになっている。OSそのものがそれ単独でファイルを扱えるように設計されているのだが、ユーティリティソフトを使うことでさらに便利にファイルを扱うことができるようになっている。ある意味ではユーティリティソフトはOSの機能を補完するものであり、パソコンを使いやすくするものともいえよう。このあたりの知識があると、パソコンを買ってきていたとき、OSそのままで処理するよりも便利になるよう思う。

本章では、WindowsのOS上のソフトのうち、日本語学関係者が使いそうなソフトをいくつか選んで解説していく。ただし、それぞれの使い方の具体例は省略し、使うか使わないかを決めるために役立つように、それぞれの特徴などを述べるにとどめる。大まかには、各種ソフトを文書作成、データ処理、ユーティリティ、ウェブ関連の4つに分けながら概観していくことにする。また、アプリケーションソフトでありながら、OSと密接につながっていて、ある意味ではOSの一部とも見られるカナ漢字変換システムについてもここで扱う。

2. ソフトウェアのタイプと入手法

2.1 ソフトウェアの入手と使用

ソフトを入手法の観点で分類すると、ただで入手できるもの（フリーウェア）と有償で購入するものがある。後者は、ソフト会社が開発・作成し、家電ショップなどで売っている「市販ソフト」と、プログラマ個人が開発・作成し、比較的安価にインターネット上で売っている「シェアウェア」がある。フリーウェアの入手はインターネット経由である場合が多いので、入手法からいえば、フリーウェアとシェアウェアが近く、市販ソフトはかなり違った流通形態を取っている。

フリーウェアとシェアウェアは、インターネットのサイトからダウンロー

ドして使うようになっている。代表的なサイトとして、以下のようなものがある。後者は、名前からもうかがえるように、Windows限定のダウンロードサイトである。

- Vector (<http://www.vector.co.jp/>)
- 窓の杜 (<http://www.forest.impress.co.jp/>)

フリーウェアは無料で使えるソフトである。とはいえ、よくできたものも多いので、いろいろ試して自分にあったものを見つけ、使いこなすとパソコンが生まれ変わったようになる。使い方がわからない場合は、付属の文書中に作者の連絡先が書いてあるので、そこに質問すればよい。たいてい、素早い返事が返ってきて驚かされる。また、中には、ネット上に会議室が用意されていて、そこで質問するというスタイルを取っている場合もある。

自分の使っているパソコンが複数台ある場合、「フリーウェアは、一度入手したら、それぞれのパソコンにコピーして使ってかまわない。」

ただし、フリーウェアは無料であっても、著作権があるので、フリーウェアを勝手にコピーして友人にプレゼントしたりすることは問題になりうる。そういう場合は、「ここからダウンロードできる」という所在情報を友人に教えてあげればいい。

シェアウェアは、市販ソフトよりもずっと安価に使うことができるソフトである。しかし、機能は市販品に勝るとも劣らない。普通は、試用期間（例えば1ヶ月）が設定されていて、その間は自由に使えるが、その後は機能の一部が使えなくなる場合が多い。ソフトを使ってみた上で、本当に便利でこれからも使いそうだという場合には、しかるべき金額を支払うようにするといいだろう。シェアウェアは、まさにそういう試用ができることが最大の売りである。

シェアウェアが複数台のパソコンで使えるかはソフトごとに異なるので、注意事項をよく読んでルールに従う必要がある。以前は、同時に起動しない限り、複数台のパソコンにインストールすることを認めるシェアウェアが多くかった。したがって、1本購入して、勤務先のパソコンと自宅のパソコンの両方にインストールするような人もいた。