

# 日本語形態素解析

Melanie Siegel, Oliver Scherf

## 1. はじめに

本研究の出発点である **Verbmobil** プロジェクトの目的は、ドイツ語・英語・日本語間の機械翻訳システムの構築であった。本システムに関する詳細な解説は **Wahlster (2000)** を参照していただき、本稿ではこのシステムの基本的な流れを簡単に紹介し、本研究の **Verbmobil** システム内での位置、目的、そして日本語形態素の解析・日本語生成のプロセスを解説する。

**Verbmobil** 翻訳システムはインプット言語（英・独・日本語の話言葉）を取りこみ、文字化（日本語の場合はローマ字化）し、翻訳に必要な情報を抽出する。その結果をアウトプット言語へと変換し、翻訳文を機械的に作成、音声データ化して出力する。我々の開発した日本語形態素解析ツールは、ローマ字化された入力情報を分析する際に主に使用される。このツールの目的を我々は以下のように設定した。

活用のある語（動詞・形容詞）の語幹、語尾を見極めることで、様々な語彙情報（文構造、他動性、敬語表現、意味に関する情報）を入手する。例えば動詞“参る”の過去形“mairimashita”は、語幹：mair(u) + 語尾：mashita に分割することができる。語幹からは semantic:’go’, valence: SUBJ, honorific: SUBJ を、そして語尾からは tense:PAST などの情報を得ることができる。

**Verbmobil** システムで使用されている辞書に登録されていない動詞・形容詞が出現した場合の処理。未登録動詞・形容詞を語幹・語尾に分割し、活用形から原形

Melanie Siegel Oliver Scherf

を導出し、翻訳処理を可能にした。

日本語文書生成の際に必要な語の全ての活用形 (“hanas-i-te” “hanas-a-nai” “hanas-i-masu” “hanas-e-ba”...) を必要に応じて語幹 (“hanas(u)”) から機械的に作成する。このツールによって、辞書の大きさを大幅に制限し、システムの負担を軽減することができた。

## 2. 形態素解析・日本語文書生成プロセス

日本語形態素の分析はまず語尾の選定から始める。語尾(例えば“hanas-a-nai”の場合、語尾は“nai”)が確定した時点で、語尾と語尾の前半の言葉 (“a”) は組み合わせられて語幹と語彙情報が生成される。このデータ構造はシステム辞書に登録されていない動詞と形容詞が読みこまれた際、その語の語尾を確定し、語尾の活用情報から語幹を割り出し、語幹からその語の原形を導出する。

日本語文書を作成するには、まず与えられた意味情報から語幹と生成するべき言葉のタイプに適合した語尾を見つけ、それらを組み合わせる。

有限状態変換マシン<sup>1</sup>(Finite State Transducer)は形態素部門の現象を取り扱うのに適している。特に日本語の形態素は語幹の最後の文字によって選択肢を制限し、語尾を結合することによって確定される。有限状態変換マシンによる形態素情報での語彙の直訳も可能である。我々は語幹・語尾の語彙データベースから分析の際必要となる変換プログラム、そして日本語生成のために使用される変換マシンを生成するツールを開発した。

---

<sup>1</sup> 入力記号列を出力記号列へと変換する効率的・数学的モデル。

## 日本語形態素解析

### 3. まとめ

我々は日本語形態素を解析し、単語リストからシステムが使用する辞書を生成し、システム辞書に登録されていない動詞や形容詞を分析する形態素解析ツールを構築した。このツールを使用することで、Verbmobil 機械翻訳システムが翻訳の際必要とする情報を全て得ることができた。このツールは辞書の語彙項目を生成するだけでなく、他のシステムにも対応可能であり、既に Verbmobil プロジェクトにおいては単語リストと意味論データベースから辞書をセットアップするのにも使用されている。今後このツールは日本語のオンライン処理や、自然言語処理システムにおける日本語生成に利用され得る。語幹辞書には 2209 の語幹が収集され 46178 の語彙項目が生成される。例えば語幹”hanas(u)”に対して、語彙見出しは 90 (76 種類) 生成し、すべてが文法的に正しいものであった。Verbmobil システムにとって未知のデータ分析でも、94.9%意図されたものであった。

有限状態変換プログラムは日本語形態素を効率的に処理することができる。我々は今後日本語文書の形態素処理に関して研究を進めるつもりである。他の知識ソースの可能性を確認して行きたいと考える。

参考文献：

Wolfgang Walster(ed). Verbmobil: Foundations of Speech-to-Speech Translation, 2000.