

単語の形態的分析のための Prolog による 2つの辞書の構築について

山 田 隆

はじめに

拙稿『ロシア語教科書に使われる単語の調査』(1987)においては、教科書内に出現する単語形をコンピュータの助けを借りて抽出する試みを例示した。ここで得られた単語というものは、いわゆる辞書見出し形ではなく、さまざまな語形変化を含んだ単語であり、ロシア語のように単語の形態変化をもつ言語においては、いくつかの利点にもかかわらず、結果的には不充分かつ、不完全な処理であった。これは、

- ・ 単語の最終的形態が得られないまま処理が終了していること、
- ・ 単語のもつ文法的、辞書的意味が不明のままであること

をさす。

本稿ではこれらの語尾変化形を、各々の品詞がもっている文法的指標を基準に分類して、単一の語形に収束する手順の作成を試みたい。すなわち、

- ① 語形の入力
- ② 与えられた語形を語末から2分化し、前半部〔語幹〕と後半部〔語尾〕の文字列について辞書との照合を行なう

③ 両者の照合が成功したときの副作用として単語のリストが生成される、という作業を想定している。この結果得られる情報は、

- ・ 単語の原形
- ・ 入力語形のもつ文法的意味のリスト

の2種である。

2つの辞書の構成について

本稿で使用, 生成される照合リストは3種類である。

① 語尾変化照合辞書 1)

② 語幹辞書 2)

③ 抽出リスト 3)

1) 語尾形を照合して, その文法的素性を抽出するための辞書。

2) 入力語形の前半部〔語幹〕の照合, ①で同定された変化類型との照合を行なうための辞書。

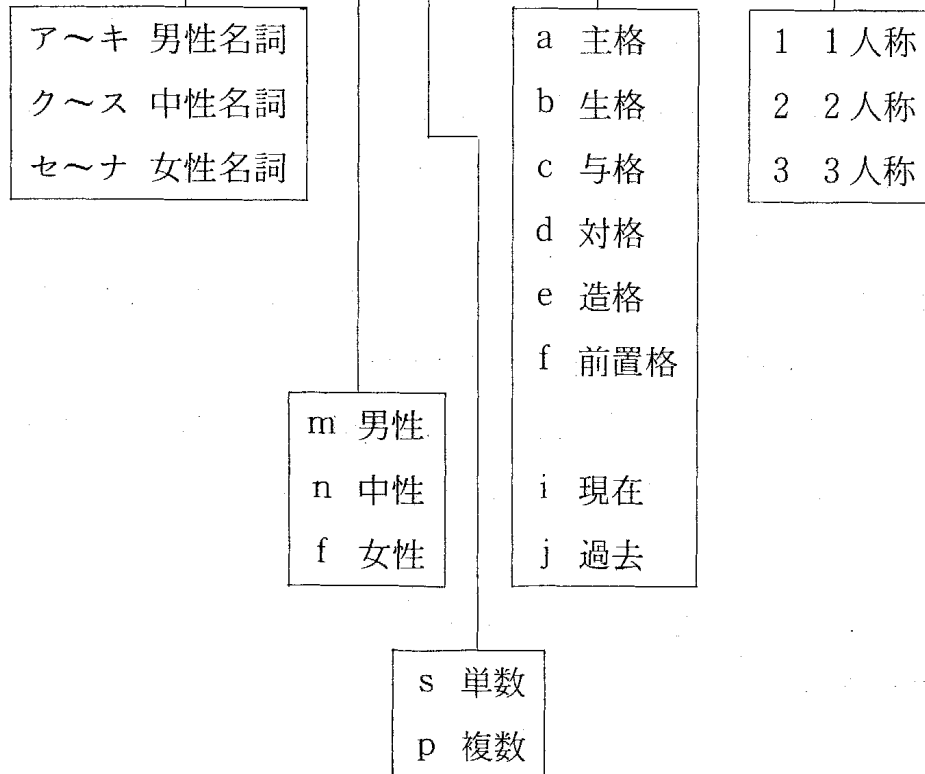
3) 語幹, 語尾の両成分が各々の辞書と一致した場合の副作用として生成される文法的素性のリスト。

各リストの構造を以下のように設定したい。

① 語尾変化照合辞書 (引数6)

d_s (語尾形, 語尾変化類型, 性, 数, 格 or 時制, 人称).

(記号一覧)



動詞過去	+	+			+
形動詞	+	+	+		
副詞					
接続詞					
前置詞					
助詞					
間投詞					

- ② 語尾処理が語幹処理に優先するので、文法的意味の情報源は語尾変化リスト $d_s(X)$ に求められる。
- ③ 処理に当たっては、各々の品詞に関する属性項目のみを参照し、該当しない文法項目については、空リスト $[\]$ として無視する。
- ④ 動詞属性値については、現在、過去、形動詞を総合すると最大5種類になるが、形動詞の分析を2段階（①形容詞分析、②動詞分析）に分離して処理すると、[時制]の項目を格と合併することができる。
- ⑤ 副詞以下の5品詞については、語尾変化の起こらない単語であるから、語幹辞書への参照だけで処理が終了する。

語幹辞書について

- ① 語幹辞書には、語彙論的情報の担い手としての機能を与える。拙稿の課題は単語の文法的意味の分析（しかも、名詞の場合）に限定されているので、語彙論的情報については次稿以降の処理段階にならなければ必要とされない。
- ② 2種の語幹を設定した理由は、出沒母音を含む音交替の処理軽減のためである。名詞の場合に可能な音交替は次のようなケースに考えられる。
- 語幹末から2文字目について、

- ① o か e を挿入（複数生格） $окно - окон, земля - земель$

- | | | |
|---|----------------|--------------------------|
| ② | o か e を削除 (斜格) | огонь—огня, ветер—ветра |
| ③ | ë を削除 | котёл—котла |
| ④ | e と ъ の交替 | лев—льва, свадьба—свадеб |
| ⑤ | ë と ъ の交替 | лёд—льда |
| ⑥ | e と й の交替 | китаец—китайца |
| ⑦ | ë と й の交替 | заём—займа; паёк—пайка |
| ⑧ | я と й の交替 | заяц—зайца |

理論上, これらの交替を語幹辞書参照の度に実行することが可能であるが, 実用の範囲を越えるので, あえて辞書項目のなかに並列記述している。

dic(lev, [lev, s, a], [1/v, [], []], m, ア, [1, 0, 0, 0]).

また, 同様の理由によって動詞語幹についても, 不定形語幹, 現在語幹の2種を設定しておくのが妥当と考える。

dic(bratノ, [bra, [], []], [ber, [], []], v, _, [0, 1, [], []]).

しかしながら, 大部分の動詞にとっては, 両者は共通である。このようなとき, 語幹第2は空リストになる。

dic(rabotatノ, [rabota, [], []], [[], [], []], v, _, [0, 0, [], []]).

- ③ 活動体・不活動体以外のことで見出し語形と等しい語形が斜格に存在するときは, 語幹を並記し, 必要な情報を列記する。

dic(raz, [raz, s, a], [raz, p, b], m, ア, [0, 1, 0, 0]).

- ④ 3個以上の語幹が可能な場合には, 語幹辞書項目を同一見出しで並記する。

dic(цыплёнок, [цыплёнок, s, a], [цыплёнк, [], []], m, ?, [1, 0, 0, 0]).

dic(цыплёнок, [цыплят, p, b], [[], [], []], m, _, [1, 0, 0, 0]).

実 行 例

г.

WORD: сeurnal


```
cf (Inf, Arti, Num_2, Case_2, Act),  
fail.
```

```
b :- !.
```

```
% 語尾を検索 (第 3, 3, 4 段階まとめて)
```

```
c (Put, Last) :-  
    word (Word),  
    substring (Word, 0, Last, Inter),  
    d_s (Put, Type, Num, Case),  
    string_term (Inter, Out),  
    dic (Inf, [Out|_], [_, _, _], Arti, Type, [Act|_]),  
    assert (rec (Inf, Arti, Num, Case)),  
    cf (Inf, Arti, Num, Case, Act),  
    fail.
```

```
c (Put, Last) :-  
    word (Word),  
    substring (Word, 0, Last, Inter),  
    d_s (Put, Type, Num, Case),  
    string_term (Inter, Out),  
    dic (Inf, [_, _, _], [Out|_], Arti, Type, [Act|_]),  
    assert (rec (Inf, Arti, Num, Case)),  
    cf (Inf, Arti, Num, Case, Act),  
    fail.
```


単語の形態論的分析のための Prolog による (山田 隆)
2つの乱書の構築について

参 考 文 献

- 『自然言語理解』, 田中穂積, 辻井潤一共編, オーム社, 1988
『機械翻訳』, 野村浩郷, 田中穂積編, 共立出版, 1988
『Arity Prolog v 5.1 リファレンスマニュアル』, 1988
Программирование на языке Пролог, У. Клосин, К. Меллиш, Мир, М., 1987