

1N-4

LangLAB上でのインドネシア語文法試作

Hartono, 田中穂積  
(東京工業大学 工学部)

1 はじめに

インドネシア語は言語の形態的分類上からはこう着語に属していると言われ、約三六種類の接辞を所有している。これらの接辞による素語(原形の語)からいろいろな意味、品詞の分出語(新しい構成された語)を構成することができる。本稿では、インドネシア語の語形変化の規則を説明し、分出語(語形変化形)から接辞と素語を抽出する方法について述べる。またこれらの形態素解析の成果を確認するためにLangLAB上で開発したインドネシア語の文法とその実験結果についても報告する。

2 インドネシア語における語の形態素、語形変化とその分出語の品詞

一般に、インドネシア語における語の形態素は大きく以下の三つに分けることができる。

- a. 接頭辞や挿入辞や接尾辞が付加している語
  - 例: menggetar (接頭辞 me + 素語 getar){振動する}
  - getaran (素語 getar + 接尾辞 an){振動}
  - gemetar (挿入辞 em + 素語 getar){震える}
- b. 単語の繰り返しでできた語(量語)
  - 1. 素語だけの繰り返し
    - 例: gunung-gunung (素語 gunung){山々}
  - 2. 素語の繰り返しに接辞を付加する
    - 例: keputih-putihan (接頭辞 ke, 素語 putih, 接尾辞 an){白っぽい}
  - 3. 素語の子音又は母音に変化する
    - 例: gerak-gerak (素語 gerak){仕種/動き}
- c. 合成語
  - 例: makan pagi (makan={食べる}, pagi={朝}){朝食}

辞形	素語の頭音	
	欠落しない	欠落する
me-	l,m,n,ng,ny,r,w,y	-
mem-	b,f,v	p
men-	d,j,sy,c,z	t
meny-	-	s
meng-	g,h,kh,母音	k

表-1 接頭語 Me- の辞形

Me+lihat →Melihat{見る}	Me+bayar →Membayar{払う}
Me+dapat →Mendapat{得る}	Me+tanya →Menanya{問う}
Me+gali →Menggali{掘る}	Me+kirim →Mengirim{送る}

表-2 Me-分出語の例

語形変化について、接頭辞Me-の辞形を見てみよう。接頭辞Me-はMe-分出語を構成する際に、素語の頭音の種類によって、表-1のように辞形が変化したり、素語の頭音が欠落したりする(表-2はその例である)。接頭辞Me-に付加される素語の品詞は、自動詞、他動詞、名詞又は形容詞である。新しく構成された分出語は動詞又は形容詞になり、以下のように分類される。

- ① 素語が名詞の時、他動詞又は形容詞を構成する
  - 例: busur {弓} → membusur {弓のような/曲がった}
  - paku {釘} → memaku {釘を打つ}
- ② 素語が形容詞の時、自動詞の分出語を構成する
  - 例: keras {堅い} → mengeras {堅くなる}
- ③ 素語が他動詞ならば、分出語は同じ他動詞になる
  - 例: ambil {取る} → mengambil {を取る}
- ④ 素語が抽象名詞又は自動詞の時、分出語は他動詞となる
  - 例: jamin {保証} → menjamin {を保証する}
  - susun {積み重ねる} → menyusun {を編集する}

3 分出語の処理

英語の語形変化は名詞の複数型、動詞の三人称単数型、過去型、現在分詞、過去分詞、形容詞の比較級最上級であるが、どちらでも語尾変化だけで、また普通品詞も変わらない。これに対して、インドネシア語の場合は、今まで述べたように、素語に接辞を付加することによって、いろいろな分出語を構成することができる。一つの素語に異なる接辞を付加すると、その接辞は、異なる品詞や異なる意味の分出語を構成する役割を持っているだけではなく、文の構造にも影響を与えている。例えば、menanyaという単語は、ditanyaと変化することによって受身の動詞として使用される。

John menanya Mary {John asks Mary}

という文を受身に直すと、

Mary ditanya John (又は Mary ditanya oleh John)

{Mary is asked by John}

となる。英語の構文を見ると、受身の文を構成するためには、be 動詞 is が付け加えられている。これに対してインドネシア語の場合は接頭辞Meを接頭辞Diに変えるだけで受身の文が構成される。つまり、文がどのような文型をしているかを知るためには、単語にどのような接辞が付加しているかを知る必要がある。このような情報を得るには、形態素解析を行う。この形態素解析は、まず分出語を分解し、接辞と素語を抽出する。付加された接辞と素語の品詞から、分出語の品詞を決定し、構文解析の際に必要な情報をも抽出する。

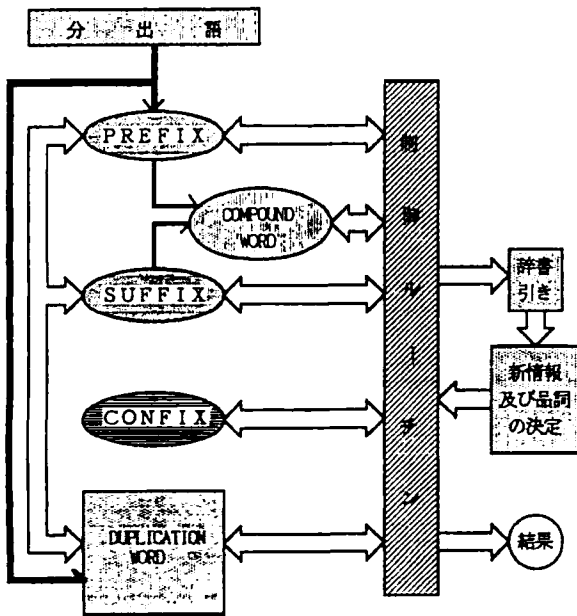


図-1 形態素解析システムの構成

インドネシア語の接辞は接頭辞、挿入辞と接尾辞の三種類からなっている。このうち、接頭辞と接尾辞については解析を行うが、限られた素語にしか付加しない接辞は解析を行わないことにする。また挿入辞の場合も限られた単語にしか付加しないことと、構成された分出語もほとんど一つの独立した単語として使われていることから、解析を行わない。その処理システムとして、図-1のように構成される。

制御ルーチンは接辞を抽出する際にすべての手続きを管理する。制御ルーチン内のプログラムによって、分出語が分解され、辞書引きをしながら、語根と接辞を抽出する。接辞を処理する順番として、まず接頭辞から調べ、なければ、接尾辞を見る。両方とも失敗に終わったら、バックトラックによって、複合辞を調べる。このように、どちらかが成功になったら、ただちにその得られた語根と接辞に基づいて、分出語の品詞と必要とされる情報を決定される。畳語と熟語についても同じ処理方法を利用している。

ここで、例として、menanya (接頭語 me + 素語 tanya) {問う} という分出語からその接頭辞と素語を抽出するアルゴリズムを見てみよう。まず、menanya を [m,e,n,a,n,y,a] に展開する。次に、2 に述べた接頭辞 Me- の辞形の変化の規則を見てみると、「語根の頭音が "t" なら、Me- が付加すると、Me- は Men- になり、頭音が欠落」となっている。この一連の手続きは Prolog のアサーションで次のように表現される。

prefix([m,e,n,a]Rest],[t,a]Rest),me).

第三引数は、接頭辞を示す。このアサーションにより、

?- prefix([m,e,n,a,n,y,a],X,Prefix). の質問に対して、

X = [t,a,n,y,a] と Prefix = me

を得る。つまり、menanya という分出語の原形 tanya と接辞 me を抽出することができたのである。

#### 4 インドネシア語の文法作成

インドネシア語の文法は、名詞句の単語レベルでの構造を除けば、一般的に英語の文法体系に似ている。文の種類には命令文、平叙文と疑問文がある。英語の場合は文頭の構文からその文の種類を分類できるが、インドネシア語の場合は、命令文を除いて、日本語のように、文の最後まで見ないと、その文の種類が判定できない。またインドネシア語は、英語などと違って、動詞の時制変化がない。その代わりに、時制を表現するための単語がある。例えば、過去完了を表現するには、"telah" を動詞の前に加えなければよい。【食べる】 makan を【もう食べた】にするには、telah makan となりまた未来型の場合は"akan"を付けて、akan makanとなる。われわれは、これを形勢副詞と呼んでいる。また文の中に、【昨日】とか【二日前】といった時制を表す副詞があれば、形勢副詞を省略してもよい。

今回、約100個のインドネシア語の文法規則をLangLAB上で開発した。これらの規則は、接辞の処理を十分考慮した上で記述したものである。特に動詞句の場合は、動詞の接辞を重視して、規則を作成した。図-2は、出力結果の一例である。

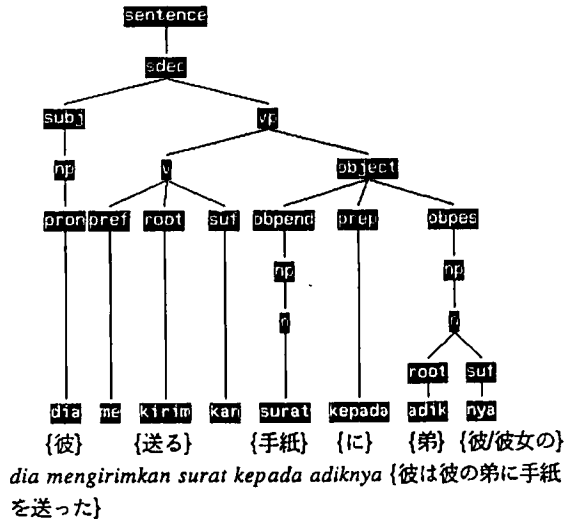


図-2 構文解析の結果の例

#### 5 おわりに

本研究は、インドネシア語の語形変化の規則を開発し、語形変化形から原形を復元するプログラムを作成した。これにより、以下のような解析が実現可能になった。

- a. 分出語から接辞と素語の抽出
- b. 接辞と語根の品詞から、分出語の品詞の決定

また、形態素解析の成果を確認するために、約 100個の規則からなるインドネシア語の文法を開発した。開発した文法規則は、小規模ではあるが、実験から、小学校の高学年程度の一般的なインドネシア語の文にも対応出来ることが確認された。

#### 参考文献

- ① Hartono : 『Prologによるインドネシア語の形態素解析と構文解析に関する基礎的研究』、東京工業大学情報工学科卒業論文(1986)
- ② S.Wojowasito : 『PENGANTAR SINTAKSIS INDONESIA』、Penerbit SINTHA DHARMA, Bandung,1976