論文 89-26-4-7

한국어의 형태소해석

(Morphological Analysis of the Korean Language)

李 秀 炫,* 小 澤 愼 治,* 李 柱 根**

(Soo Hyon Lee, S. Ozawa and Joo Keun Lee)

要 約

한국어의 구문 및 의미해석등에 필요한 정보를 도출하기 위한 입력문의 형태소해석에 대하여 기술한다.

명사구에서 명사와 조사를 분리하고, 복합명사의 분리점을 선택하는 조건을 규정하며, 변형된 복합명사를 처리하는 규칙을 표시한다. 그리고 규칙동사에서 어간과 어미를 분리하고, 변칙용언과 음운축약 등을 효과적으로 처리하기 위한 논리표현 형식을 제안한다. 이 논리표현은 해석규칙과 속성값으로 구성한다.

명사사전의 중복을 배제하기 위하여 "명사형 하다" 동사의 분리처리와 Q parameter 도입에 의한 "이다"의 처리방법을 보이고, 또한 부정문의 처리형식도 유도하여 복합술부의 형태소와 기본형식을 제시한다.

Abstract

A morphological analysis is described to extract the informations which are required in syntactic and semantic analysis of the Korean language.

The noun and particle are separated in a noun phrase, the selecting conditions are specified to analyze the compound noun and a restoring rule is represented to process the irregular compound noun. The stem and ending are separated in normal verbals and a logical representive form is proposed to the anomalously inflected word and contracted vowels. The logical representation is composed of the attribute value and analyzing rule.

The redundancy of noun is reduced in the dictionary as the verb of a "Nounformed HA-" is processed by "noun" and "HA-", separately and a predicative "IDA" is analyzed by Q parameter. The processing form of negation is also derived and the morpheme and basic structure of compound predicative parts are presented.

*正會員, (日本) KEIO大學 理工學部
(Faculty of Science and Technology, Keio Univ., Japan)

**正會員, 仁荷大學校 電子工學科
(Dept. of Elec. Eng., Inha Univ.)

接受日字 : 1988年 11月 7日

Ⅰ. 서 론

형태소해석은 입력문자열에서 구문의 최소단위인 형태소로 분리하여 구문 및 의미해석에 필요한 모든 정보를 사전에서 추출해 내는 과정이다. 형태소해석의 내용은 언어구조의 특징과 서법에 따라 크게 다르다. 한국어의 형태소해석은 영어, 일본어의 그것과

는 많은 차이가 있다.[1~4] 한국어의 형태소해석은 부분적 발표를 볼 수 있다.[5,6]

본 연구는 한국어에서 구 단위로 입력되는 문자열에서 조사 및 활용형을 분리하고, 변형용언의 구조변환, 음운축약 및 명사형 동사를 분리하는 등 제 현상을 거론한다.

해석은 연속적으로 입력된 자모열에서 단음절을 구분하고, 또 그들 몇개의 단음절 집합으로 구성되는 문자열의 단위에서 형태소를 분리하여 사전을 수색하고, 한 음절씩 증가시키면서 사전을 대조한다. 이때 어절간의 접속관계와 변형구조, 음운현상 등을 조사하여 문의 형태소를 단위로 하는 모든 해(solution)를 사전에서 추출해서 전후관계와 최장일치의 조건에 따라 형태소를 도출한다.

특히 음운축약 및 변칙용언등은 한국어의 고유현상으로서 해석과 생성에서 양면성을 가지고 있다. 변칙용언은 속성값과 결합규칙의 논리표현에 의하여 해석하고, 같은 관점에서 변형복합명사의 처리를 적용한다. 또 "명사형 하다" 동사의 분리방법을 제안하여 사전의 중복을 배제하고, Q parameter의 도입에 의한 "이다" 해석방법도 제시했다. 나아가서 간단한 복합술부의 구조형식과 해석방법도 제시했다.

자연어의 처리에서 첫 단계인 형태소처리가 명확하지 않으면, 그것이 점차 전성되어 구문해석에서 애매성(ambiguity)을 발생시키는 요인이 된다.[7~9] 특히 한국어와 같이 형태소가 고정되어 있지 않은 언어에서는 형태소해석이 매우 중요하다.

II. 체언의 형태소해석

1. 입력형식과 정형화

입력문의 서법은 절 단위에 대한 자모의 연속된 음절열이며, 출력은 그것을 재구성하여 출력하는 방식이다. 그 때문에 자모의 열로서 이루어지는 단음절을 자동적으로 구분하고, 또 구분된 몇개의 단음절로서 구성되는 형태소의 후보를 그림1과 같이 분리한다.

```
입력(CVCVCCVCVCVC---)      ←-
        ⇨(CVCVCCVCVCVC---)
            ↑   ↑  ↑  ↑
        ⇨(CVCVC / CVCVCVC / ---)
↑ : 단음절의 구분점
←- : 수색방향
/ : 형태소의 분리점
```

그림 1. 형태소의 분리
Fig. 1. Separating of morpheme.

문의 정형화는 입력문의 기호, 숫자, 알파벳 표기의 약어 등을 번역하지 않는다.

기호 : (; , . , :, …?, #, …)
숫자 : (1, 2, 3, …)
Alphabet 약어(AI, MT, CPU, …)
고유어 : (Seoul, Tokyo, …)

2. 명사구의 분리

명사구는 명사에 조사가 결합하는 형으로서 체언의 기본이 된다. 명사에 결합되는 격조사는 문법기능을 가지며 용언의 어미와는 구별된다.

입력된 문자열의 후미로부터 역으로 한 단음절씩 추출하여 조사테이블을 참조하여 대응되는 조사를 선택한다. 다음에 한 단음절씩 증가시키면서 점검하여 조사가 발견되지 않을때 조사의 검색을 정지하고, 나머지를 명사사전에서 수색한다. 만약 실패하면 앞에서와 같이 명사의 검색을 반복하여 판정하고, 사전으로부터 각각의 모든 정보를 추출하여 형태소의 해석을 완료한다.

만약, 명사의 사전에 그 단어가 등록되어 있지 않을때는 error 메세지를 출력한다.

예) "학교에서는"⇨학교 / 에서 / 는
"우리와의"⇨우리 / 와 / 의
"어린이에게도"⇨어린이 / 에게 / 도

여기서 "어린이"의 "이"는 조사로도 분리되고, "어린이"의 명사로 되어 애매성이 발생할 수 있다. 이때 명사를 우선 판정함으로써 애매성을 배제할 수 있다.

2. 복합명사

일반적으로 한국어는 한개의 어절에 수개의 명사를 연속해서 나열하는 경우가 많다. 이때 같은 명사가 반복 나열되어 사전의 중복을 초래한다. 따라서 사전의 용량을 고려해 넣고, 적절히 기본단위를 규정하는 것이 바람직하다. 간단한 예를 다음에 표시한다.

A) 단순한 복합명사

예) (1).	(2). "신발자국"
학교 / 교육	신 / 발 / 자 / 국
기술 / 교육	신 / 발 / 자국
의무 / 교육	신 / 발자국
대중 / 교육	신발 / 자국
교육 / 기관	
교육 / 정책	
교육 / 과정	
교육 / 시설	

(2)는 어느 것으로 분리해도 가능하지만, 어느 것

을 형태소로 인정하는 것이 효과적 시스템이 될 것인가의 판단을 위해서 선택조건을 둔다.

B) 변형 복합명사

한국에서 두개의 명사가 복합될 때, 기본음절이 음운변화를 하는 복합명사가 있다. 간단한 예를 다음에 표시한다.

예) 소나무⇨솔 / 나무
고깃배⇨고기 / 배
쇠고기⇨소 / 고기
나뭇잎⇨나무 / 잎
버드나무⇨버들 / 나무

이들 구조로 부터 변형 음소를 속성값으로 하는 결합규칙의 논리식을 정의하여 다음의 처리형식을 제시한다.

〔규칙〕

CHA_NN (rule attribute)　　(1)
CHA_NN : 변형 복합명사

예)

[소나무 / 버드나무] → [소(+ㄹ) / 나무 / 버드(+ㄹ) / 나무] → [솔 / 나무 / 버들 / 나무]

[고깃배 / 나뭇잎 / 쇠고기] → [고깃(－ㅅ) / 배 / 나뭇(－ㅅ) / 잎 / 쇠(－이) / 고기] → [고기 / 배 / 나무 / 잎 / 소 / 고기]

↑ 변형복합명사　↑ rule-attribute　↑ 원형

이때 rule은 첨가, 삭제(+, －), attribute는 (ㄹ, ㅅ, …)이며, 이 rule과 속성은 해석규칙이 되는 동시에 사전이 된다. 또 위 (1)식은 한국어의 특수현상인 음운축약어, 변형용언, 축약과 변형의 혼합형 등의 제현상을 일의적으로 처리할 수 있게 된다.

Ⅲ. 용언의 형태소해석

용언은 어간에 어미가 결합하여 활용하는 성질이 있다. 모든 정보가 용언에 집중되어 있으며, 여러가지의 형태소로 분리 가능하다. 용언이 활용되는 기본형을 (2)식에 표시한다.

VP＝STM＋Fi　　(2)
VP : 용언구
STM : 어간
Fi : 어미
(1≤i≤n)

그러나 해석의 대상이 되는 한국어의 용언은 반드시 규칙적인 것만이 아니고, 대단히 복잡한 변칙용언이 상당히 많다. 두 음절간의 음운축약, 변칙과 음운축약의 혼합형 등 한국어에만 존재하는 제현상, 또는 다른 품사로부터 전성되어 입력되는 것도 많다. 따라서 용언의 형태소를 해석하기 위하여 규칙적인 것으로부터 변칙 및 명사가 전성되어 동사 및 형용사적으로 사용되는것 등의 구조를 효과적으로 처리하기 위해서는 형식화하는 것이 중요하다.

1. 용언구조의 형식화

용언을 형성하는 최소단위는 "STM＋Fi"이다. 이 최소단위를 기본으로 하여, 명사, 부사, 보조어간 등이 결합되어 모든 한국어의 기본용언이 형성되는 구조형식은 표 1과 같다.

표　1. 용언의 구조형식
Table 1. Strutural form of inflected word.

	형　　식	빈 도 %
1	STM＋Fi | (J)	68. 32
2	STM＋AUXSTM＋Fi | (J)	11. 7
3	N＋STM＋Fi | (J)	16. 24
4	N＋STM＋AUXSTM＋Fi | (J)	2. 72
5	AD＋STM＋Fi | (J)	0. 93
6	AD＋STM＋AUXSTM＋Fi | (J)	0. 08

N : 명사,　　AD : 부사
STM : 어간　　AUXSTM : 보조어간
Fi : 어미　　\| (J) : 생략가능한 조사

위 6형식을 통합하면 다음과 같은 한개의 식으로 표시된다.

VP＝(N) \| (AD) \| ＋STM＋(AUXSTM)＋
Fi \| (J)
( )내는 생략 가능한 품사
\| 는 선택적임을 표시한다.　　(3)

표 1의 형식에서 기본이 되는 것은 1), 2)의 경우이며, 어간(STM)도 규칙적인 것과 변칙적인 것이 있으므로 어간의 형태를 분석하는 과정이 필요하다. 또 3)～6)의 경우는 명사(N) 혹은 부사(AD)의 결합에 의하여 용언이 되는 형식이다. 이들 형식속에 축약형도 내포되는 경우가 있고, 위 6형식에서 기본적으로는 Fi가 종결어미 이지만, 활용도 되고, 때로는 조사J가 첨가되는 경우가 있으므로 (J)는 선택적으로 표시했다.

2. 보조어간

보조어간은 어간과 어미의 사이에 삽입되어 시제, 존칭 등의 정보를 보강하는 역활만 하며, 이것은 어간의 형태를 취하고 있기 때문에 보조어간의 뒤에도 어미가 활용된다. 이때 활용되는 어미는 어간뒤에 활용되는 어미와 똑같은 역활을 한다. 그러므로 보조어간을 분리해내도 어미의 형태는 변하지 않는 특징이 있다. 따라서 보조어간의 정보만을 별도로 추출해 낼 수 있다. 한국어의 보조어간은 일본어의 조동사의 위치에 있으나 문법적으로는 매우 다르고,활용도 하지 않으며, 그 뒤에 어미가 결합되어 활용된다. 보조어간의 순서와 해석에 필요한 정보는 표2에 표시한다.

표 2. 보조어간의 분류
Table 2. Classification of auxiliary stem.

태(Voice)	사동(EMP)	기, 리, 이, 히, 우, 추
	피동(PAS)	기, 리, 이, 히
존칭(R)	경칭(R1)	(으)시
	겸양(R2)	오옵, 사옵, 삽, 잡
Tense(T)	현재(T1)	ㄴ, 는
	미래(T2)	겠, ㄹ, 을, 를
	과거(T3)	ㅆ, 던, 았, 었
	대과거(T4)	ㅆ었, 었었, 았었

3. 음운축약의 처리

한국어는 인접하는 두개 단음절의 특정모음이 결합해서 한개의 음소로 축약되는 성질이 있다. 이것은 주로 용언에서 발생하며 음운조화에 의하여 글자에 반영되어 변형을 일으킨다. 이때 변형된 입력은 원형이 아니기 때문에 사전에 등록되지 않음으로 원형으로 환원하기 위한 5형식을 표3에 표시한다.

표 3. 음운축약의 규칙
Table 3. Rule of contracted word.

축 약 어	규 칙
과	ㅗ+아
궈	ㅜ+어
괘	ㅚ+어
걔	ㅑ+여
켜	ㅣ+어

이것을 논리대수적으로 표시하면 인접하는 두 음절 "C1V1+C1V2C2"의 모음(V1, V2)가 논리적으로 결합하여 V(=V1+V2)로 축약된다.

(C1V1+C1V2C2+C1V2 ...)
→ (C1(V1+V2)C2 / C1V2 ...)
→ (C1VC2 / C1V2)
C1 : 자음
V1, V2 : 모음
V = (V1+V2) : 복합모음

축약어는 어간과 보조어간을 음운적으로 결합시켜 문자에 반영되는 특이한 현상으로 용언의 구조 해석에서 그 영향이 매우 크다. 이들에 대한 사전에는 원형만 등록되고, 문의 표층에는 변형 형태만 나타난다.

예) a)형 : 1) 왔다→오+았다→오 / 았다
2) 졌다→지+었다→지 / 았다
b)형 : 3) 했다→하+였다→하 / 았다
4) 줬다→주+었다→주 / 았다
5) 됐다→되+었다→되 / 았다
↑ 변형 ↑ 어간/어미

위의 규칙에 대한 논리표현을 다음 (4)식과 같이 표시한다.

〔규칙〕

CNSTT(Rule(Attribute_1, Attribute_2))
CNSTT : 축약형 (4)

이것은 축약어의 사전형태가 되는 동시에 처리규칙이다.

4. 변형용언의 형태소

A) 변형용언

규칙용언의 해석은 명사구의 해석과 거의 같기 때문에 설명을 생략하고, 구문처리로 넘긴다. 변형용언은 언어학에서는 11종으로 분류하지만, 본 연구는 그 성질을 3개의 군으로 분류해서 속성값과 조건, 처리규칙을 설정하여, 변형된 어간에서 원형을 도출한다. 표4에 변형용언의 변환규칙을 표시하였다.

위의 규칙과 의미표현을 위한 지식은 (5)식으로 표시하며, 이것은 사전의 표현이 되는 동시에 처리규칙이 된다.

〔규칙〕

IRR_STM(rule attribute) (5)
IRR_STM : 변형어간

표 4. 변형 용언의 변환규칙
Table 4. Transformative rule of anomalously inflected word.

rule	attribute	condition
KX (+)	ㄹ, ㅅ, 으 우, ㅎ	탈락된 어간의 최종 음소를 속성값으로 하여 변형된 어간에 더한다.
KY (−)	ㄷ, 르 러, ㅂ	변형된 음소를 탈락시키고 첨가될 음소를 속성값으로 하여 더한다.
KZ (0)	여	"하"로 끝나는 어간 다음에 후속되는 "여−"로 시작하는 어미를 "어−"로 바꾼다.

B) 축약과 변형용언의 혼합형

변형용언외에 모음축약과 변형칙용언이 복합되는 형태가 있다. 이것은 앞에서 표시한 축약규칙과 변형용언을 동시에 처리한다. 이 경우의 예를 보이면

예) 도와서→도오(−ㅂ) / 아서→돕 / 아서
누워서→누우(−ㅂ) / 어서→눕 / 어서
↑ ↑ ↑
변형 rule-attribute 원형

와 같다. 위의 축약 및 변형용언의 혼합형태는 음운적으로 복합현상을 수반하고 있으며, 처리형식을 (6)식에 표시한다.

〔규칙〕

IRR_CNSTT (rule attribute) (6)
IRR_CNSTT : 변형용언과 축약 혼합어 CNSTT

일반적으로 종래의 언어학에서는 변형용언과 음운축약어를 분리해서 생각했다. 따라서 혼합형 처리에서 애매해지기 때문에 본 논문에서는 혼합형식을 도출함으로서 해석을 용이하게 하였다.

이상의 변칙용언, 축약어등은 기계처리를 위한 해석과 생성에서 양면성을 가지며, 변칙용언의 처리 subroutine 의 예는 다음에 보인다.

```
(DEFUN ANOMAL (LAMBDA (ARG1 ARG2)
 ((AND (GET ARG1'KX) (ASSOC ARG2 ENDTAB))
  (SETQ LG (SUB1 (LENGTH ARG1)))
  (SETQ ARG* (PACK (LIST (SUBSTRING ARG1 1 LG) (GET ARG1'KX))))
 ((AND (GET ARG1'KY) (ASSOC ARG2ENDTAB))
  (SETQ ARG* (PACK (LIST ARG1(GET ARG1'KY)))))))
;
(DEFUN KZ (LAMBDA (RT JS KK)
 ((AND (ASSOC RT KVTBL) (EQUAL'HA (SUBSTRING RT (SUB1 KK))))
       ((EQUAL'YEO (SUBSTRING JS 13))
        (SETQ JOSA (PACK (CDR (UNPACK JS))))
        ((PREDCATE RT JOSA) T))
       ((AND (EQUAL'E (SUBSTRING JS 11)) (NOT (EQUAL'EUIJS)))
        (SETQ JOSA (PACK (LIST'EO (SUBSTRING JS 2)))
        (SETQ JOSA (PACK (LIST'EO (SUBSTRING JS 2))))
        ((PREDCATE RT JOSA) T) ]
```

5. 부정문의 형태소

부정문은 긍정문에 대립되는 문장으로서 어떤 사물이나 명제에 대해서 부정하는 뜻을 가지며, 술어동사에 부정요소가 매립된다. 한국어에서 부정문은 동사를 중심으로 전, 후에 다양하게 표현된다. 부정구조의 패턴은 표 5 에 표시한다.

표 5. 부정구조의 패턴
Table 5. Pattern of negation.

형식	No.	규 칙	연결어미
I	1	Neg_A+V	Fi
	2	Neg_B+V	
II	3	V지+[Neg_A]	Fi
	4	V지+[Neg_D]	
	5	Vㄹ+[줄 Neg_D]	
III	6	V지+[Neg_A]+V	Fi
	7	V서는+[Neg_A]+V	
	8	V지+[Neg_B]+V	
	9	V지+[Neg_C]+V	

Neg_A : 안, 않, Neg_b : 못, Neg_C : 말, Neg_D : 모르
V : 동사의 어간

표 5 에 보인 바와 같이 I 형식은 부정요소가 서술동사V의 앞에 오고, II 형식은 서술동사의 뒤에 온다. 또 III 형식은 서술동사의 중간에 매립된다. 이와 같이 한국어의 부정문은 영어, 일본어 등의 부정문과는 많은 구조적 차이가 있다. 따라서 번역시스템에서는 대상어에 따라서 구조변환을 해야하기 때문에 형태소 레벨에서 구조형식을 만드는 것이 필요하다.

Ⅳ. 보조용언의 형태소

한국어의 서술부는 한개의 어간과 어미로 구성되는 경우도 있지만 일부의 보조용언 및 이들에 상응하는 형태는 본 용언의 의미를 보강하는 역활을 한다. 따라서 본 논문에서는 한국어의 처리에 적합하도록 본 용언과 보조용언사이에서 연관되는 어미의 형태를 분석하였다. 본용언과 보조용언으로 이루어지는 서술부의 구조는 본 용언의 어미의 종류와 보조용언의 기능에 따라서 3가지 유형을 도출한다.

1. 주종관계의 구조

이 관계는 본 용언과 보조용언이 결합하여 한개의 서술부가 되고, 보조용언에 의하여 본 용언의 의미를 보강하는 형식이다. 이때 전후의 연결관계는 본

용언의 어미에 의하여 결정되며, 의미의 보강은 보조용언의 어간이 맡는다. 이와 같은 역활을 하는 어미는 부사형어미(F6)와 연용어미(F7)가 있다.

F6	아(어), 게, 지, 고
F7	러, 려, 면, 기도, 어도, ...

또한 이들의 어미 뒤에는 제한된 보조용언이 후속되어 의미를 보강하며, 이들을 기능별로 분류하면 표 6과 같다.

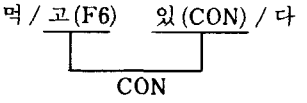
표 6. 보조용언의 기능분류
Table 6. Functional classification of auxiliary verbal.

개념	기호	종 류
시발	STA	-러 들-
계속	CON	-고 있-, -어 있-, -(어/아) 가-,
종료	COM	-어 버리-, -(어/아) 내-, -(어/아) 놓-,
중지	INT	-다 놓-, -다가 두-, -다가 말-,
반복	REP	-어 쌓-, -어 대-,
의향	IND	-려 하-, -려고 하-, -고자 하-,
소망	HOP	-기 바라-, -기 원하-, -면 하-,
시행	OPE	-어 보-, -아 보-,
봉사	SER	-어 주-, -어 드리, , -어 바치-,
당위	REA	-어야 하-, -어야 되-, -어야만 되-,
용인	ADM	-어도 되-,
인용	QUO	-다고 하-,
희망	DES	-고 싶-,
피동	EMP	-게 되-, -아 되-, -어 지-,
사동	PAS	-게 하-, -게 만들-,

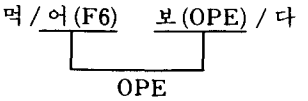
서술부의 주종관계의 구조형식의 예는 (7)과 같다.

STM · F6 | F7+Aux_V · Fi (7)

예) (1) "먹고 있다"



(2) "먹어 보다"



2. 불완전명사가 결합된 보조용언의 구조

본 용언과 보조용언 사이에 불완전명사가 포함되어 이 불완전명사와 보조용언이 결합하여 본 용언을 보강한다. 이때 본 용언의 어미로서 불완전명사와 연결관계를 맺는 어미는 그 종류가 한정되어 있으며, 이것은 F8로 표시한다.

F8	ㄹ, 을, 를, ㄴ, 은, 는

그리고 전체의 불완전명사를 국문학에서는 57개로 분류하고 있지만, 그 중에서 본 용언과 보조용언간에 결합되는 불완전명사(AUX_N)는 13종류이다.

AUX_N	수, 줄, 리, 턱, 양, 체, 번, 상 만, 듯, 사, 듯, 척

불완전명사와 결합되는 보조용언을 기능별로 분류하면 표 7과 같다.

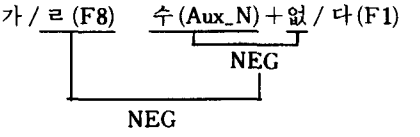
표 7. 불완전명사와 결합된 보조용언의 기능분류
Table 7. Functional classification of auxiliary verbal combined with imperfect noun.

개념	기호	종 류
부정	NEG	-ㄹ수 없-, -ㄹ리 없-,
인지	REC	-ㄹ줄 알-, -ㄹ지 알-, -ㄹ줄을 알-,
가능	POS	-ㄹ수 있-, -ㄹ수가 있-,
가식	FAU	-는 체하-, -는 척하-, -는 양하-,
불성	MIS	-ㄹ번하-,
추정	INF	-ㄹ듯하-, -는 듯하-, -ㄹ 상싶-,
가당	VAL	-ㄹ만하-, -ㄹ 사하-,

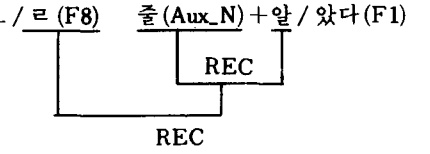
이에 대한 구조형식은 다음과 같이 표시한다.

STM · F8+Aux_N+STM · Fi (8)

예) (1) "갈 수 있다"



(2) "올 줄 알았다"



3. 병렬적 구성

앞 뒤의 두 용언이 주종관계에 있는 것이 아니고, 동등한 역활을 하는 경우이다. 즉 두 용언의 동작이 병립하고, 앞의 용언의 어미가 병렬적 기능을 가지며, 이러한 어미를 F9로 표시한다.

F9	(으)며, (으)면서, 다가, ...

이것은 다음과 같은 형태가 된다.

STM · F9+STM · Fi　　(9)

예) (1) "노래하며 춤춘다"

노래 / 하 / 며(F9)　춤추 / 다(F1)

parallel-relation

(2) "문을 닫고 들어왔다"

닫 / 고(F9)　들어오 / 았다(F1)

parallel-relation

이상에 대한 subroutine들은 다음과 같다.

```
(DEFUN AUX_PROC(LAMBDA(ARG ARG1 ARG2)
  (FINDSUF ARG1)
  ((AND   (MEMBER APC_FG'(F6 F7))
          (SEARCH_AUX ARG2))
         ((NEG ARG1 AUXSTM)T)
         ((APT ARG1 AUXSTM)T)
         ((MODE ARG1 AUXSTM)T))
  ((AND   (EQUAL APC_FG F8)
          (MEMBER ARG2 Aux_N_LST)
          (SEARCH_AUX(CADDR SENTENCE)))
          ((NEG ARG1 AUXSTM)T)
          ((APT ARG1 AUXSTM)T)
          ((MODE ARG1 AUXSTM)T))
  ((PARA_AUX_PROC ARG ARG1 ARG2)T)
    ]
    ;
    ;
```

V. 명사형의 동사화

명사형의 동사화에는 두가지의 관점이 있다. 이 두가지 관점은 매우 민감한 것으로서 하나는 "명사+하다" 결합이고, 다른 하나는 "명사+이다" 결합이다.

1. "명사(N)+하다"의 해석

이것은 명사에 어미가 결합하여 활용하는 경우이다. 명사와 어미가 결합해서 동사의 역활을 하는 어간이 매개적으로 사용된다. 이것은 "명사+하다" 동사로서 이것과는 또 다른 원래의 "하다" 동사가 있다. 이 양자는 의미적으로 다르다. "명사+하다"는 그 양이 대단히 많고, 또 그 명사는 명사사전과 동사사전에서 중복된다. 따라서 그림2와 같이 "명사+하다"로부터 명사를 분리하는 것이 바람직하다.

그러나 "명사+하다"로부터 "명사"와 "하다"를 분리하면, "명사+하다" 동사가 문장속에서 지배하는 구문 및 의미적 역활은 상실해 버린다. 즉 명사는 오직 원래의 기능만 가지고 분리된 "하다" 동사도 명사적 의미는 상실하여 원래의 "하다"로 돌아간다.

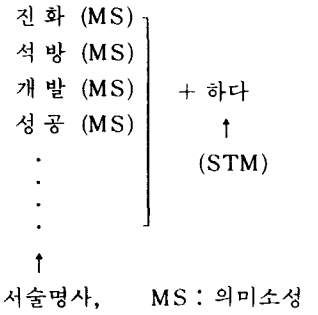


그림 2. "명사+하다" 동사
Fig. 2. Verb of "noun+HA-"

따라서 본 연구에서는 "명사+하다" 동사를 분리하는 방법을 다음과 같이 기술한다.

동사 : "하다"(ViQj)
"명사+하다"(Vi'Qj')
분리 : "명사"(MS)
　　"하다"(ViQj)

단, ViQj, Vi'Qj'는 각각 동사의 구문 및 의미의 격지배 정보이다.

이 문제의 해결을 위한 방법으로서 명사에 "하다"가 결합된 상태의 의미를 명사에 기록해 두고, 그 "명사"가 "하다"와 결합할 때는 그림3과 같이 명사의 의미소성을 "하다"의 의미소성에 대치하는 방법을 생각할 수 있다.

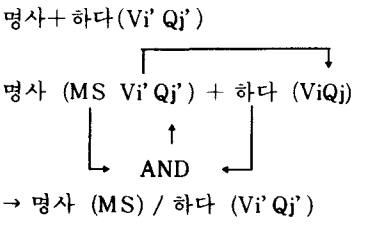


그림 3. "명사"와 "하다"의 분리방법
Fig. 3. Separating method of "noun" and "HA-".

이와 같이 하여 "명사"와 "하다"를 분리해서 기억된 사전으로부터 각각 추출한 "명사"와 "하다"를 결합할 수 있어서 명사사전의 중복을 배제할 수 있다. 이 "명사+하다" 동사는 한국어 용언의 30% 이상이 됨으로 이로 인한 사전의 효과가 매우 높아진다.

2. "명사+이다"의 해석

"명사+하다"와는 별개로 명사 N에 문법소 "이다"가 붙어서, 용언과 같은 활용을 한다.

이 "이다"에 대해서는 언어학에서도 많은 논란이 있었다.[10] 조사해본 결과, "이다"의 "이"는 조사도 아니고, 용언의 어간도 아니지만 활용하는 미묘한 성질이 있고, 대부분의 명사에 접속되어 활용된다.

이는 단순히 "명사"와 "어미"를 결정해 주는 요소로서 명사와 결합해도 "하다"와 같이 의미가 변하지 않는다. 따라서 본 논문에서는 Q parameter를 정의하여 용언형태와 유사한 형식을 취한다. 즉 "명사(N)+이(Q)"에 어미 Fi을 결합하는 구조형식을 다음과 같이 제시한다.

$$\underline{(N+Q)}+Fi \qquad (10)$$

Fi= 다, 냐 : 종결형
구, 냐
ㄴ, ㄹ : 전성형
ㅁ, 기
고, 면 : 연결형
거, 어서
.
.
.
.

여기서 Q parameter는 "하다" 동사와는 달리 어간이 아니고, 의미를 가지고 있지 않으며, 음운적 연결기능만 가지고 있기 때문에 명사 N과 결합하거나 분리해도 명사 N의 의미는 변하지 않는다. 따라서 Q parameter에 의하여 효과적으로 처리할 수 있다. 이상의 명사의 동사화 subroutine을 다음에 표시한다.

```
(DEFUN NOUNFORM_PROC (LAMBDA (ARG1 AGR2)
   ((AND (SEARCH ARG1 KNTBL)
         (NOT (FINDKPTC ARG2 T0TAB T1TAB T2TAB)))
        ((AND (EQUAL 'I (SUBSTRING ARG2 1 1))
              (ASSOC (SUBSTRING ARG2 2) ENDTAB))
                 (SETQ MORP_FG T)
                 (SETQ ROOT ARG1)
                 (SETQ HROOT (LIST ARG1))
                 (PUSH TAKNO TRANSTACK)
                 (VERBPRO ARG2) T)
             ((AND (EQUAL COMP 'MS)
                   (EQUAL 'HA (SUBSTRING ARG2 1 2)))
                 (SETQ MORP_FG T)
                 (SETQ VSSS T)
                 (SETQ MSLT (CADR KT))
                 (PUSH TAKNO TRANSTACK)
                 (SETQ HROOT (LIST ARG1))
                 (VERBPRO ARG2) T)
        ]
   ;
   ;
```

Ⅵ. 해석결과

h : (KJMT)
KOREAN⇨천연자원이 부족한 우리나라는 외국으로부터 수입한 원료를 가공하여 수출하는 무역입국을 지향할 수 밖에 없읍니다.$

```
===== morphological analysis =====
(CHEONYEON : N(MCC))
(JAWEON : N(MCC))
(I : P)
(BUJOGHA : V/ (AJ 24))
(N : SUF/ (N T3) R1 F9)
(URI : N(MDC MAC HUM))
(NARA : N(MLC))
(NEUN : P)
(OEGUG : N (MLC))
(EUROBUSTEO : P)
(SUIB : N(MS V24S2))
(H A : V/ ((V27W9 62)))
(N : SUF/(N T3)R1 F9)
(WEONRYO : N (MCC))
(REUR : P)
(GAGOO : N(MS V24S2))
(HA : V/((V27W9 62)))
(EO : SUF/T1 R1 F6)
(SUCHUR : N (MS V24S2))
(HA : V/AUX)
(NEUN : SUF/T1 R1 F9)
(MUYEOG : N (MCC MS))
(IBGUG : N (MLC))
(EUR : P)
(JIHYAOHA : V/ ((V4W9 245)))
(R : SUF/ (R T2) R1 F9)
(SU : AUXN)
(BAGG : N (MLC))
(E : P)
(EOBS : V/(AJ 4))
(EUBNIDA : SUF/T1(EUB R2)F1)
T :
```

Ⅶ. 결 론

본 연구의 성과는 다음 같이 정리된다.

(1) 복합명사에서 형태소의 분리점은 여러가지로 분리 가능함을 감지하고, 선택조건에 의하여 적정선을 규정해 두는 것이 유용하였다.

(2) 복합명사에서 변형이 발생함으로 이것에 대한 구조형식을 도출하고, 처리규칙을 유도하였다.

(3) 종래의 변칙용언은 11종으로 분류하지만, 본 연구에서는 그것을 3개군으로 재분류하여 형식화하고, 속성값의 결합규칙을 유도하여 변형용언을 일의적으로 처리하였다.

(4) 음운축약은 단독으로 존재하는 경우와 변칙용언과 혼합형도 있다. 이들 축약과 변형용언을 일의적으로 처리함으로서 변형의 미묘한 부분도 효율적으로 처리할 수 있음을 확인하였다.

(5) "명사형 하다" 동사의 분리방법을 제안하여 용언의 30%를 점하는 명사의 중복을 배제하였다.

(6) "이다"는 조사도 아니고, 용언도 아니지만 활용하고 있어 이에 대한 품사인정 문제로 언어학에서는 많은 논란이 있었다. 본 논문에서는 이것을 음운소의 한 형태로 보고, Q parameter를 규정하여 가상적 어간 형태로 용언과 똑같이 처리함으로서 용언의 어미를 공용할 수 있어 매우 효과적이었다. 또 "이다"는 활용하는 양이 방대함으로 그 영향이 또한 매우 크다.

參 考 文 獻

[1] M. Nagao et al., "Data structure of a large Japanese dictionary and morphological analysis," IPSJ, vol. 13, no. 6. June 1978.

[2] Sakamoto, "Basic implementation of morphological analysis in Japaneses," IPSJ, Natural language processing 38-3, 1983. 7.

[3] K.J. Yoshimura et al. "Morphological analysis of Japanese sentences using the least cost method," IPSJ, Natural language processing 60-1, 1987. 3.

[4] K. Narazone et al., "On analysis of aspect," IPSJ, Natural language processing, 61-2, 1987.5.

[5] 김상연, 이주근, "한국어 문장의 자동 띄어쓰기 방법", 전자공학회논문지, vol. 7, no. 1, 1984. 7.

[6] 윤명기, 이주근, "자연어의 형태소 분석", 전자공학회논문지, vol. 7, no. 1, 1984. 7.

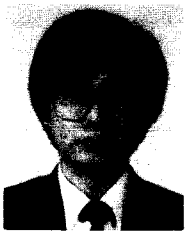
[7] S.H. Lee, K.R. Han, J.K. Lee, "A bidirectional MT system between Korean and Japanese based on a pattern net," SITA'87, Japan, Nov. 1987.

[8] J.K. Lee, J.H. Lee, "Determination of modificatory scope and inference for Korean," TENCON-87-IEEE Region 10 Conference, Aug. 25-28, 1987.

[9] J.K. Lee, K.R. Han, "Korean-Japanese machine translation system (III)," 85 SISEE, Nov. 1985 in Seoul.

[10] 최현배, 우리말본, 서울, 1961. *

著 者 紹 介



李 秀 炫(正會員)

1959年 10月 20日生. 1983年 2月 인하대학교 전자공학과 졸업. 1985年 8月 인하대학교 대학원 석사학위 취득. 현재 일본 KEIO 대학 이공학부 박사과정. 주관심분야는 자연언어 처리 및 기계번역 시스템 등임.

小 澤 愼 治 (非會員)

1967年 KEIO대학 전기공학과 졸업. 1972年 동 대학원 박사과정 수료 공학박사, 현재 KEIO대학교수. 1984年 Maryland대학 방문교수. 주관심분야는 화상, 음성신호처리 언어처리 등임.

●

李 柱 根 (正會員) 第25卷 第12號 參照

현재 인하대학교 전자공학과 교수