

# 보조어의 기능분석에 기초한 실마리어추출의 한가지 방법

리명일, 리예화

언어생활에서 보조어는 없어서는 안될 매우 중요한 언어의 한 부분이다. 보조어[1]는 독자적으로 쓰일수 없는 단어 즉 반드시 일정한 문법적형태를 갖춘 자립적단어와 결합하여 그 단어결합의 문법적의미표현기능을 수행하는 단어들이다.

선행연구[2]에서는 문장을 입력하여 형태부분석을 진행한 후 실마리어로 추측되는 단어들에 대한 판정에 기초하여 실마리어를 추출하였다. 이 방법에서는 문장전체에 대한 형태부해석을 진행하므로 문장수가 긴 문서에서는 실마리어추출속도가 뜨며 정확도가 낮은 결과가 나타난다.

론문에서는 우리 말 보조어의 기능들을 분석한데 기초하여 실마리어의 추출속도를 높이기 위한 한가지 방법을 제안하였다.

## 1. 보조어의 기능분석

보조어는 반드시 다른 단어의 도움에 의하여서만 그 쓰임이 가능한 단어로서 다음과 같은 기능을 수행한다.

① 보조어는 반드시 일정한 문법적형태를 갖춘 자립적단어와 결합하며 어휘적의미의 구체화와 함께 문법적의미를 표현하여주는 기능을 수행한다.

보조어모임을  $A$ , 자립적단어모임을  $B$ ,  $a \in A$ ,  $b \in B$  라고 하면 보조어  $a$ 와 자립적단어  $b$ 는  $b+a$ 의 형태로 나타난다.

실례로 《저기 보이는것이 학교이다.》에서 보조어로 쓰이는 불완전명사 《것》은 독자적으로 쓰이지 못하며 피동형단어결합인 《보이는》과 함께 쓰인다. 또한 불완전명사 《것》이 없으면 문법이 이루어지지 않는다.

② 보조어는 대부분이 고유어로 쓰인다.

실례로 《갈, 것, 나위, 녀, 노릇, 대로, 무렵, 즈음, ...》을 들수 있다.

③ 자립적단어로부터 추상화되어 점차 보조어로 전환됨으로써 자립적단어의 어휘적 의미를 강조하거나 보충하여주는 기능을 수행한다.

실례로 《에 관하여》, 《을 위하여》, 《에 비하여》, 《에 대하여》, 《로 인하여》 등을 들수 있다.

④ 보조어는 어음구성이 비교적 단순하고 하나의 말뿌리로 이루어져있다. 다시말하여 보조어들은 대체로 한소리마디로 된것이 기본을 이루며 2~3개의 소리마디로 된것은 극히 적다.

빈도수사전에 올라있는 체언형보조어 58개중에서 한소리마디는 47개, 두소리마디는 11개로서 한소리마디로 이루어진 보조어가 81%이다.

⑤ 보조어는 자립적단어와 결합할 때 자립적단어가 일정한 문법적형태를 갖출것을 필요로 한다.

실례로 《자립적단어(체언 또는 용언)》 + 불완전명사》의 구조를 이룬 경우에는 체언

과 용언은 자립적단어가 규정형을 취하고 불완전명사는 격형태나 맺음형을 취한다.

그러면 과학론문들에서 많이 쓰이는 보조어 《에 대하여, 을 위하여, 에 비하여, 에 기초하여, 로 인하여》 등에 대하여 보자.

실례로 문장 《그를 대할 때마다 마음은 즐거웠다.》에서 《대하다》는 분명히 동사로서 자립적단어이며 《인공지능학의 발전추세에 대하여》에서 《에 대하여》는 자립적단어 《대하다》로부터 나온 보조어 《에 대하여》로서 과학론문들에서 많이 리용된다.

이에 대한 표현구는 《에 의하여(의한)》, 《로 인하여(인한)》, 《에 관하여(관한)》, 《로 말미암아》, 《에 따라》, 《을 고찰하기 위하여(하기 위한)》, 《에 기초하여(에 기초한)》, 《에 준하여》, 《으로 인하여》, 《과 더불어》 등에서도 볼수 있다.

우와 같이 일정한 격형태와 결합된 단어들은 문법적형태를 갖추고있어 뻔해야 뻔수 없는 련관성을 가지고있으며 뒤의 단어들의 대다수가 자기의 본래 어휘적의미로부터 추상화되어 자립적의미가 아니라 문법적의미를 표현하는 기능을 수행한다.

이러한 단어부류는  $[b+to]+[a+to]$ 의 형태를 취한다. 여기서  $[to]$ 는 우리 말의 토를 의미한다.

## 2. 문서들에서 실마리어추출방법

우에서 보조어의 기능을 분석한데 기초하여 과학론문과 기술서적들에서 실마리어를 추출하기 위한 방법을 보자.

### ① 보조어모임(A)의 구성

문서들을 분석한데 의하면 실마리어와 함께 가장 많이 쓰이는 토는 속격토 《의》이며 보조어는 《에 의하여(의한)》, 《로 인하여(인한)》, 《에 관하여(관한)》, 《로 말미암아》, 《에 따라》, 《을 고찰하기 위하여(하기 위한)》, 《에 기초하여(에 기초한)》 등이다. 이러한 보조어들을 1부류 보조어라고 하며 이것들과 결합된 단어들은 실마리어로 될 확률이 가장 크다.

또한 보조어에 관하여 《No, a, count, sil》의 구조를 갖춘 실마리어목록표를 작성한다. 여기서 a는 보조어, count는 빈도수, sil은 실마리어를 의미한다. 이 표의 기능은 실마리어 추출에 참가하는 속격토 《의》와 보조어의 빈도수를 계산하며 실마리어추출과정에 반복되는 실마리어들의 중복을 피하는데 리용된다.

### ② 실마리어추출

N개의 초록문장을 가진 문서라고 가정하자.

먼저 문서의 제목을 입력한다.

문서의 제목에서 보조어모임의 원소인 속격토 《의》와 보조어를 탐색한다.

속격토가 존재하는 경우 NP + 《의》 + 《NP》결합에서 《NP》를 이루는 단어들에 대하여 실마리어사전에 존재하는가를 판단한다.

만일 존재하면 문서의 실마리어로 등록한다. 이때 실마리어목록표를 빈도수에 따라 갱신한다.

다음 빈도수가 가장 높은 순서로 보조어모임의 원소와 결합한 단어들에 대하여 실마리어사전에 존재하는가를 판단한다.

만일 존재하면 문서의 실마리어로 등록한다. 이때 실마리어목록표를 빈도수에 따라 갱신한다.

다음  $N$ 개의 초록(Abstraction)문장들에 대하여 우와 같은 과정을 반복한다.

다음 얻어진 실마리어모임( $S$ )의 원소들이 표준실마리어사전에 등록되어있는가를 검사하고 등록되어있으면 문서의 실마리어로 결정한다.

### 3. 실마리어추출에 대한 분석평가

여기서는 정보과학소논문(국문)에 대하여 실마리어를 추출하고 적중률과 완전률,  $F$ 값을 리용하여 체계의 성능을 평가하였다.

적중률, 완전률,  $F$ 값은 다음과 같이 계산하였다.

$$\text{적중률: } P = \frac{|S_h \cap S_s|}{|S_s|}$$

$$\text{완전률: } R = \frac{|S_h \cap S_s|}{|S_h|}$$

$$F\text{값: } F = \frac{(\beta^2 + 1) \times P \times R}{\beta^2 \times P + R}$$

여기서  $S_h$ 는 정답모임,  $S_s$ 는 체계가 출력한 대답모임,  $S_h \cap S_s$ 는 체계가 출력하는 대답모임중에서 정답과 일치하는 실마리어모임,  $||$ 는 모임의 농도,  $\beta$ 는 적중률과 완전률의 중요도를 조절하는 상수이다.  $\beta$ 를 3으로 설정하였다.

실마리어추출결과는 표와 같다.

표. 실마리어추출결과

척도 방 법	$R$	$P$	$F$	시험문장수
선행한 방법	0.331 4	0.496 1	0.342 8	2 800
제안한 방법	0.610 9	0.481 2	0.594 9	2 800

표에서 보여주는바와 같이 논문에서 제안한 방법은 높은 적중률과 함께 완전률을 1.73배 높였다.

### 맺 는 말

조선어에서 보조어는 반드시 일정한 문법적형태를 갖춘 자립적단어와 결합한다.

논문에서는 우리 말 보조어의 기능들을 분석한데 기초하여 실마리어의 추출속도를 높이기 위한 한가지 방법을 제안하고 실마리어추출을 위한 적중률과 완전률을 개선하였다.

### 참 고 문 헌

- [1] Y. Ding, M. Palmer; Natural Language Process, 3, 3, 24, 2014.
- [2] Mattia Tomasoni; Natural Language Process, 2, 4, 38, 2017.

## **A Method of Keyword Extraction Based on Function Analysis of Auxiliary Language**

*Ri Myong Il, Ri Ye Hak*

In this paper we suggested a method of keyword extraction based on function analysis of auxiliary language in Korean language.

Keywords: keyword extraction, linguistics, information extraction