

동사와 표현구를 리용한 조선어문장분류의 한가지 방법

리명일

경애하는 최고령도자 김정은동지께서는 다음과 같이 말씀하시였다.

《과학연구부문에서는 나라의 경제발전과 인민생활향상에서 전망적으로 풀어야 할 문제들과 현실에서 제기되는 과학기술적문제들을 풀고 첨단을 돌파하여 지식경제건설의 지름길을 열어놓아야 합니다.》

론문에서는 언어생활에서 쓰이는 문장들을 최대로 일반화하고 문장의 구조분석과 의미분석, 자연언어처리에 리용하기 위하여 조선어문장들을 분류하는 한가지 방법을 제기하였다.

선행연구[1]에서는 조선어기초문장을 토대로 하여 14가지 문장유형을 분류하였다.

대표적인 문장유형은 다음과 같다.

1유형: $N_{\text{주}} + V_{\text{술}}$ // 밀림이 설레인다.

3유형: $N_{\text{주}} + V_{\text{이술}}$ // 옥이는 모범학생이다.

14유형: $N_{\text{주}} + N_{\text{대}} + N_{\text{수}} + V_{\text{술}}$ // 철이는 논문을 컴퓨터로 편집하였다.

여기서 N 은 명사, V 는 동사, 《주》는 주어, 《술》은 술어, $V_{\text{이술}}$ 에서 《이》는 바꿈토, $N_{\text{수}}$ 에서 《수》는 수단을 나타낸다.

또한 선행연구[2]에서는 문장유형의 수에는 제한을 주지 않는 토를 리용한 문장유형 결정방법을 제기하고 실현하였다.

우에서 본 문장유형은 다음과 같이 결정된다.

① 문장유형 1의 《 $N_{\text{주}} + V_{\text{술}}$ 》은 다음과 같이 변경된다.

$$N_{\text{주}} \Rightarrow NP + T_{zg}$$

$$V_{\text{술}} \Rightarrow V_z$$

$$N_{\text{주}} + V_{\text{술}} \Rightarrow NP + T_{zg} + V_z. \text{ 즉 문장유형패턴은 } \{ _T_{zg} _V_z _ \} \text{이다.}$$

여기서 T_{zg} 는 주격토, V_z 는 자동사이다.

② 문장유형 3의 《 $N_{\text{주}} + V_{\text{이술}}$ 》은 다음과 같이 변경된다.

$$N_{\text{주}} \Rightarrow NP + T_{zg}$$

$$V_{\text{이술}} \Rightarrow NP + T_{zg} + T_b + T_m$$

$N_{\text{주}} + V_{\text{이술}} \Rightarrow NP + T_{zg} + NP + T_b + T_m$. 즉 문장유형패턴은 $\{ _T_{zg} _T_b _T_m \}$ 이다. 여기서 T_b 는 바꿈토, T_m 은 맺음토이다.

③ 문장유형 14의 《 $N_{\text{주}} + N_{\text{대}} + N_{\text{수}} + V_{\text{술}}$ 》은 다음과 같이 변경된다.

$$N_{\text{대}} \Rightarrow NP + T_d$$

$$N_{\text{수}} \Rightarrow NP + T_{zo}$$

$$V_{\text{술}} \Rightarrow VP$$

《 $N_{\text{주}} + N_{\text{대}} + N_{\text{주}} + V_{\text{술}}$ 》 \Rightarrow 《 $NP + T_{\text{zg}} + NP + T_d + NP + T_{\text{zo}} + VP$ 》 즉 문장류형패턴은 $\{-T_{\text{zg}}-T_d-T_{\text{zo}}-VP\}$ 이다. 여기서 T_d 는 대격토, T_{zo} 는 조격토이다.

우에서 $N_{\text{주}}$ 는 행동, 상태의 주체로 되는 대상을 나타내는 주격형명사구로서 명사구와 주격토의 결합으로, 《 $V_{\text{이술}}$ 》은 명사구와 바꿈토 《이》, 맺음토 《다》의 결합으로 이루어진다.

1. 동사와 표현구를 리용한 문장류형결정

조선어는 토교착어로서 문장의 류형과 구조, 의미는 토에 의하여 분석될수 있다.

한편 동사는 대상의 움직임을 나타내는 품사로서 문장의 마지막위치에 놓이면서 문장성분들과의 구조적관계, 의미적관계를 맺고있는 중요한 문장구성요소이다. 조선어동사는 인칭과 수에 따라 문법적형태를 바꾸지 않으며 《가다, 오다》등과 같이 미정형으로는 절대로 쓰이지 않는다는 특성을 가지고있다. 또한 《이란 ... 을 말한다.》, 《왜냐하면 ... 때문이다.》 등 표현구에 관해서도 문장류형을 결정할수 있다.

동사와 표현구의 이러한 특성으로부터 문장류형을 표현하는 기본요소로 동사와 표현구를 정하였다. 즉 동사모임을 V , 토모임을 T , 표현구모임을 P 라고 할 때 문장류형요소모임은 $Y=\{V, T, P\}$ 로 된다.

$V = \{\text{Rec } V_i \mid i = \overline{1, n_{\text{동}}}\} (n_{\text{동}}: \text{세부동사의 레코드끝번호, Rec } V_i: \text{세부동사의 레코드})$

$T = \{\text{Rec } T_i \mid i = \overline{1, n_{\text{토}}}\} (n_{\text{토}}: \text{세부토의 레코드끝번호, Rec } T_i: \text{세부토의 레코드})$

$P = \{\text{Rec } P_i \mid i = \overline{1, n_{\text{표}}}\} (n_{\text{표}}: \text{표현구의 레코드끝번호, Rec } P_i: \text{표현구의 레코드})$

① $\text{Rec } V_i$ 의 구조

$\text{Rec } V_i = (V_{\text{no}}, i, V, V_{\text{pt}}, V_{\text{tr}})$

여기서 V_{no} 는 동사의 유일번호, i 는 세부동사의 레코드번호, V 는 동사의 원형, V_{pt} 는 세부동사의 이름, V_{tr} 는 동사의 변형형태이다.

실례로 1류형문장 《밀림이 설레인다.》의 술어에는 일반적으로 자동사가 놓이는데 모든 자동사가 오는것이 아니라 주로 자연현상을 나타내는 자동사 V_3 (개이다, 흐리다, ...), 정신적 및 육체적작용을 나타내는 자동사 V_4 (자다, 놀다, 굶다,...), 사건의 발생을 나타내는 자동사 V_5 (생기다, 일어나다,...), 피동형의 동사 V_6 (열리다, 시작되다, 건설되다, ...) 등이 놓인다.

실례

$\text{Rec } V_i = (121, 3, \text{자동사}, [\text{설레이다}], [\text{설레인다, 설레입니다, 설레었다, 설레었습니다.}])$

② $\text{Rec } T_i$ 의 구조

$\text{Rec } T_i = (T_{\text{no}}, i, T_{\text{na}}, T)$

여기서 T_{no} 는 토의 유일번호, i 는 세부토의 레코드번호, T_{na} 는 세부토이름, T 는 토이다.

③ $\text{Rec } P_i$ 의 구조

$\text{Rec } P_i = (P_{\text{no}}, P, P_{\text{ex}})$

여기서 P_{no} 는 표현구의 유일번호, P 는 표현구, P_{ex} 는 표현구실례이다.

문장류형결정을 위한 일반적요구는 다음과 같다.

① $N+T_s+N$ 의 명사구에서 속격토(의)를 류형요소모임(Y)에서 제거한다. 그것은 속격토가 있는 경우와 없는 경우의 문장류형이 일치하기때문이다.

실례

나는 소설을 읽고 많은것을 느꼈다.

나는 천세봉의 소설을 읽고 많은것을 느꼈다.

② 동사가 술어부분에서 표현되는 비존경－현재, 비존경－과거, 비존경－미래, 존경－현재, 존경－과거, 존경－미래는

$$\text{Rec } V_i = (V_{no}, i, V, V_{pt}, V_{tr})$$

의 V_{tr} 에 반영한다.

③ 도움토 《는/은》은 문장에서 주격형태를 이루므로 주격토 《가/이》와 같은 부류에 속하는것으로 한다.

④ 《?, !》 등의 부호를 문장의 류형요소로 정하고 문장을 크게 알림문, 물음문, 명령문, 추김문으로 나누어 등록한다.

조선어문장류형작성절차는 다음과 같다.

- ① 문장입력
- ② 형태단어분리
- ③ 형태단어에서의 토분리 및 레코드번호결정
- ④ 표현구검색 및 표현구레코드번호결정
- ⑤ 동사검색 및 동사레코드번호결정
- ⑥ 문장류형요소 V, T, P 를 리용하여 문장류형결정

2. 실험결과 및 분석

론문에서는 조선말대사전에 등록되어있는 6 000개 문장들을 대상으로 문장류형을 작성하기 위한 방법을 제기하였다.

표에 제안한 문장류형체계와 이전 체계의 지표별성능을 보여주었다.

표. 지표별성능

지표	선행방법[3]	제안방법
문장류형개수/개	1 300	1 500
문장구조분석률/%	77	83
문장의미분석률/%	55	60

표에서 보는바와 같이 토, 표현구, 동사를 류형결정요소로 하는 제안방법은 토를 기본으로 하여 문장류형을 결정하던 선행방법[3]보다 문장개수는 200개 증가하였으며 구조분석률은 6%, 의미분석률은 5% 개선되었다.

맺 는 말

토, 표현구, 동사를 문장유형을 결정하는 요소로 하고 이것을 리용하여 조선어문장을 분류하기 위한 분류체계를 구성하였다.

참 고 문 헌

- [1] 방금숙; 외국어로서의 조선어실천문법연구, 김일성종합대학출판사, 102~120, 주체98(2009).
- [2] 배광희; 컴퓨터응용언어학, 김일성종합대학출판사, 32~43, 주체92(2003).
- [3] Q. N. Rockiman; Natural Language Process, 1, 2, 41, 2016.

주체108(2019)년 2월 5일 원고접수

A Method of Korean Sentence Classification Using Verb and Expression Phrase

Ri Myong Il

Korean sentences used in language life is expressed with various types by verb, expression phrase and particles.

In this paper we suggested a method for Korean sentence classification by using verb and expression phrase.

Key words: language process, database