

동음이의어처리를 위한 단어들사이의 관계목록작성의 한가지 방법

리 명 일

언어생활에서는 동음이의어가 많이 리용된다.

동음이의어란 서로 다른 뜻을 가지면서도 표현(소리)이 똑같은 개념들을 말한다.

실례로 《밤》은 《날》을 나타내는 의미와 《열매》를 나타내는 의미를 포함하고있다.

일반적으로 동음이의어는 이야기가 진행되는 환경이나 문맥에 따라 구별된다.

그러므로 컴퓨터에 의한 동음이의어처리에서는 환경과 문맥정보를 자동적으로 추출할수 있는 알고리즘을 설계하여야 한다.

선행연구[1, 2]에서는 《조선말대사전》에 존재하는 동음이의어에 대하여 올림말과 매주석의미에 대응하는 단어를 련관시키는 방법으로 문맥정보를 추출하였다. 이 방법에서는 사전에 기초한것으로 하여 단어의 개수는 제한되어있으며 시기에 따라 출현하는 새 단어들에 대한 정보를 얻을수 없었다.

본문에서는 동음이의어처리를 위한 단어들사이 관계목록을 작성하는 방법을 제안하였다.

1. 동음이의어가 명사인 경우의 단어관계목록작성

조선어에서는 명사와 동사에서 동음이의어들을 많이 찾아볼수 있다.

여기서는 먼저 동음이의어가 명사인 경우의 단어관계목록작성방법에 대하여 보기로 한다.

품사가 명사인 경우의 단어들은 주로 명사들과 결합하여 복합명사를 이룬다.

실례 그는 수학기호를 리용하여 문장을 서술하였다.

이 경우에는 형태부해석을 진행하여 복합명사를 분해하고 매 단어들사이의 련관관계 목록을 작성한다.

말줄기를 L , 토를 To 로 표시하고 말줄기를 이루는 앞불이를 F , 말뿌리를 R , 뒤불이를 B 라고 할 때 형태단어(W_f)를 다음과 같이 표시한다.

$$W_f = \langle L, To \rangle = \langle F, R, B, To \rangle$$

여기서 $L = \langle F, B, R \rangle$ 이다.

이때 단어들사이 관계목록을 작성하는 방법은 다음과 같다.

- ① 형태단어 W_f 의 선택
- ② 토사전을 리용한 토(To)의 분리 및 말줄기(L)추출
- ③ 말줄기 L 에 대한 복합명사분해
 $L = \langle F, B, R \rangle$ 를 얻는다.
- 단어사전, 뒷불이(앞불이, 뒤불이)사전을 리용한다.
- ④ 단어들의 의미추출

매 단어들에 대한 의미를 의미계층사전을 통하여 얻는다.

⑤ 다음과 같이 단어들사이 관계목록을 작성한다.

단어들사이 관계목록구조를 S 라고 하자.

$$S = \langle \text{No}, W_0, W_1, \dots, W_n, N_0, N_1, \dots, N_n, f, \text{other} \rangle$$

여기서 No 는 목록에서의 번호, W_0 은 동음이의어, W_i 는 복합명사분해결과 얻어진 단어 ($i=1, n$), N_0 은 W_0 에 대응한 의미계층사전에서의 번호, N_i 는 W_i 에 대응한 의미계층사전에서의 번호, f 는 동음이의어의 출현회수, other 는 분야정보 혹은 새 단어출현정보이다. 새 단어출현정보는 새 단어인 경우 1로, 새 단어가 아닌 경우 0으로 설정한다.

①에서 형태단어 W_i 는 문장코퍼스에서 문장을 선택하고 형태단어들을 추출하는것으로 한다.

2. 동음이의어의 품사가 동사인 경우의 단어관계목록작성

동사는 일반적으로 거의 모든 언어에서 수많은 단어결합의 의미론적 및 문법적핵심의 하나로 되고있는 단어의 한 부류이다.

조선어에서 동사는 어휘적다의성과 함께 풍부한 문법적형태를 가지고 문장안에서 다른 성분들과 적극적으로 어울려 쓰임으로써 문장론적구조의 중심을 이룬다.

문장안에서 동사와 맞물리는 명사(체언)는 동사의 어휘적의미 또는 문법적형태에 어울리게 자기의 문법적형태를 갖추고 문장의 어느 한 성분으로 될수 있다. 때문에 동음이의어의 품사가 동사인 경우의 단어관계목록작성에서는 주격, 대상격에 대응한 단어들사이의 관계를 반영하여야 한다.

실례로 동사 《타다》의 다의성을 문장속에서 고찰해보자.(표)

표. 동사 《타다》에 대한 다의성분석

의미번호	실례문장	동작주격	격	격이름
1	그는 말을 탄다.	그는	말을	대상격
2	창고가 불에 탄다.	창고가	불에	상대격
3	그는 기타를 탄다.	그는	기타를	대상격
4	그는 색감을 탄다.	그는	색감을	대상격

표에서 《격》과 《격이름》은 동작주격이 요구하는 격과 이름을 나타낸다.

위의 내용들을 반영하여 동음이의어의 품사가 동사인 경우 다음과 같은 방법으로 관계목록을 작성한다.

① 문장코퍼스(Cor_S)에서의 문장(S_i) 선택

$$S_i \in Cor_S$$

② S_i 에 대한 형태부해석

③ 동사(V_i)의 선택

하나의 문장(S_i)에는 2개이상의 동사가 출현할수 있다.

④ 단어들사이 관계목록을 작성한다.

$$S = \langle \text{No}, V_i, \text{Frame}, \text{other} \rangle$$

$$\text{Frame} = \langle \text{Sub}, \text{Obj}, \text{Rel} \rangle$$

여기서 S 는 단어들사이 관계목록구조, No 는 목록에서의 번호, V_i 는 동음이의어(동사), $Frame$ 은 격틀에서의 격, Sub 는 주격, Obj 는 대상격, Rel 은 상대격, $other$ 는 기타 정보이다.

맺는 말

사전에 기초한 동음이의어처리방법의 제한성을 극복하기 위하여 단어들사이 관계를 목록화하는 방법을 제안하고 실현하였다.

참고 문헌

- [1] W. N. Pickexian; Natural Language Process, 2, 3, 112, 2017.
- [2] Taul Kiwetch; Natural Language Process, 3, 3, 17, 2018.

주체110(2021)년 5월 5일 원고접수

A Method of Relation List Creation Between Words for Homonym Process

Ri Myong Il

In this paper we proposed a method of relation list creation between words for homonym process.

Keywords: grammar, nature language process, structure analysis