

국소 구문 관계 및 의미 공기 정보에 기반한 명사 의미 모호성 해소

김영길⁰ 홍문표 김창현 서영애 양성일 류철 황은하 최승권 박상규
한국전자통신연구원 휴먼정보처리연구부

(kimyk, hmp63108, chkim, yaseo, siyang, ryuch, yinxia, choisk, parksk}@etri.re.kr

Word Sense Disambiguation Based on Local Syntactic Relations and Sense Co-occurrence Information

Young-Kil Kim⁰ Mun-Pyo Hong Chang-Hyun Kim Young-Ae Seo Chul Ryu Yinxia Huang Sung-Kwon
Choi Sang-Kyu Park
Dept. of Human Information Processing Department, ETRI

요 약

본 논문에서는 단순히 주변에 위치하는 어휘들간의 문맥 공기 정보를 이용하는 방식과는 달리 국소 구문 관계 및 의미 공기 정보에 기반한 명사 의미 모호성 해소 방안을 제안한다. 기존의 WSD 방법은 구조 분석의 어려움으로 인하여 문장의 구문 관계를 충분히 고려하지 못하고 주변 어휘들과의 공기 관계로 그 의미를 파악하려 했다. 그러나 본 논문에서는 동사구의 논항 의미 관계 뿐만 아니라 명사구내에서의 의미 관계도 고려한 국소 구문 관계를 고려한 명사 의미 모호성 해소 방법을 제안한다. 이 때, 명사들의 의미는 자동번역 시스템의 목적에 맞게 공기(co-occurrence)하는 동사들에 따라 분류하였다. 그리고 한층 자동 번역 지식으로 사용되는 명사 의미 코드가 부착된 74,880 의미 격들의 의미 공기 정보를 이용하였으며 형태소 태깅된 말뭉치로부터 의미모호성이 발생하지 않게 의미 공기 정보 및 명사구 의미 공기 정보를 자동으로 추출하였다. 실험 결과, 의미 모호성이 발생하는 명사들에 대해서 83.9%의 의미 모호성 해소 정확률을 보였다.

1. 서론

제한된 영역의 문장이 아닌 일반 문장을 대상으로 한 어휘의 의미 모호성 문제는 자연언어 처리 기술의 가장 어려운 문제들 중의 하나이다. 특히 자동번역 시스템에 있어 어휘 모호성 문제는 대상 언어의 적절한 대역어 선택과 직접 연관되어 번역의 질을 크게 좌우하는 중요한 요소이다. 지금까지 형태소 분석 및 구문 분석에 있어서의 모호성 해결은 그동안 활발히 연구되어 좋은 성과가 있었지만 의미 모호성 해소 방법은 현재 활발하게 연구되고 있는 분야이다.

현재까지 연구에서는 의미 발음치, 전자사전의 뜻풀이 용례 등의 자원을 이용한 방법론이 제안된 바 있으며, 일부 어휘에 대한 어휘 모호성 해소 실험에 그치고 있으며, 실제 일반 문장들을 대상으로 한 동형이의어의 의미 모호성 해소는 시도되고 있지 못하다. 이중 의미 태그 코퍼스를 이용한 확률 통계적인 방법은 대량의 신뢰성이 보장된 이용 가능한 자원을 구축하기 위해서는 상당한 비용을 필요로 하지만 언어의 특성을 가장 잘 반영한다. 따라서 본 논문에서는 한층 자동 번역 지식으

로 사용되는 명사 의미 코드가 부착된 80,000 의미 격들의 의미 공기 정보와 형태소 태깅된 말뭉치로부터 자동 추출한 의미 공기 정보 및 명사구 의미 공기 정보를 이용하는 의미 모호성 해소 방안을 제안한다.

그리고 기존의 WSD 방법은 구조 분석의 어려움으로 인하여 문장의 구문 관계를 충분히 고려하지 못하고 주변 어휘들과의 공기 관계로 그 의미를 파악하려 했다. 그러나 구문 관계를 고려하지 않는 경우, 물론 각 동사에 지배를 받는 논항의 핵심어는 지배 동사와 각 논항들 사이의 공기 관계로 그 의미 파악이 가능하다. 그러나 논항이 단일 명사가 아니라 복합 명사를 포함하는 명사구로 이루어진 경우, 그 명사구의 핵심어의 자매 성분들은 명사구 내에서 국소적으로 그 의미 관계를 파악해야 한다. 동사구의 논항 의미 관계 뿐만 아니라 명사구내에서의 의미 관계도 고려한 국소 구문 관계를 고려한 명사 의미 모호성 해소 방법을 제안한다.

2. 관련연구

어휘 모호성 문제를 해결하기 위하여 현재 사전 뜻풀이의 의미 정보를 이용하는 방법[1,2], WordNet, Roget 시소러스 등의 시소러스를 활용한 방법[3,4], 문맥 공기 정보를 이용하는 방법[5,6,7], 어휘의미 분석 말뭉치를 이용하는 방법[8] 등 활발히 연구되고 있다. 의미 부착 말뭉치 구축은 의미가 사람의 주관에 따라 변동적이기 때문에 일관성 유지하기가 어려운 단점이 있지만 언어의 특성을 가장 잘 반영할 수 있기 때문에 의미 모호성 해소를 위한 중요한 데이터이다.

그리고 의미 부착 말뭉치 구축을 위한 작업은 연세대학교에서 연세 한국어 사전을 참조로 의미 주석을 시도하고 있으며, 세종 계획의 일환으로 국립국어연구원에서 의미 주석 말뭉치를 구축하고 있으며, 인천대학교에서는 명사 시소러스를 데이터로 하여 의미 분석 말뭉치를 구축한 바 있다.

3. 명사 의미 분류

일반적으로 명사를 의미적으로 분류할 때 가장 많이 사용되는 방식은 개념적 기준에 따라 온톨로지(Ontology)를 구축하는 방식이다. 명사 의미분류를 논의할 때 가장 많이 언급되는 워드넷(Wordnet)도 가장 대표적인 예라고 할 수 있다[9]. 그러나 명사를 개념적 기준에 따라 의미 분류를 할 때 생기는 가장 큰 문제점은 (강범모, 2001)[10]에서 지적된 바와 같이 개념적 기준에 대한 정의 자체가 어렵고 모호하다는 점이다. 예를 들어 ‘악기’이라는 개념을 하위 부류로 나눌 때 연주 방법에 따라 ‘현악기’, ‘관악기’, ‘타악기’ 등으로 나눌 수 있지만, 악기의 기원에 따라 ‘동양악기’, ‘서양악기’ 등으로 나눌 수도 있다.

온톨로지는 일반적으로 일차원적인 모습의 트리(tree)구조를 띄게 된다. 그러나 이 때 많은 경우에 개념과 개념들은 단순한 ‘is-a’ 관계만이 아니라 ‘subject-field’, ‘is-a-function-of’, ‘is-a-part-of’ 등과 같은 관계를 지니는 통일되지 못한 모습을 보인다. 이러한 문제를 해결하기 위해 명사의 의미기술을 일차원적인 트리구조를 통해서가 아니라 격자(lattice)등을 통해 다중레벨에서 기술하는 방식이 사용되기도 한다[11, 12].

그러나 일반 시소러스등의 의미분류 방식의 가장 큰 문제점은 의미 분류를 자동번역 시스템에 적용하기가 어렵다는 점이다. 자동번역 시스템에서 명사 의미분류의 목적은 일반적인 정보 검색 시스템과는 다르다고 할 수 있다. 자동번역 시스템에서 명사의 의미는 주로 구조 분석과 대역어 선택을 위해 사용된다. 따라서 의미분류에 대해 여러 가지 기준이 있을 수 있지만 이러한 통사적 성질과 대역어 공유라는 기준도 의미분류를 위한 하나의 언어

학적 기준이 될 수 있을 것이다.¹

이러한 논의를 기반으로 우리는 한중 자동번역 시스템을 위해 한국어의 명사를 최대 4개 레벨 약 200여개의 의미코드를 사용하여 분류하였다. 명사의 최상위 레벨은 구체명사, 추상명사, 활동명사로 나뉜다. 이러한 분류 하에 각각의 의미 코드들은 다음의 두 가지 기준을 통해 세 분류 되었다:

- i) 공기 (co-occurrence)하는 동사들에 따라 분류화한다. (예: [기계]!를 수리하다)
- ii) 하나의 명사가 여러가지 의미코드를 가질 수 있다. (예: 동아일보 → 조직, 건축물, 인쇄물)²

첫번째 기준은 일정 명사 부류가 특정 동사들의 특정 문법 기능어 자리 (주어, 목적어, 간접 목적어 등)에 공통적으로 쓰일 수 있다면, 이러한 명사들을 하나의 부류로 간주할 수 있다는 직관에 근거한다.

두번째 기준은 자료 구조를 단순화하기 위해 의미코드 체계를 트리 구조로 구성하는 것을 유지하기 위해 동음 이의어나 다의어의 경우는 물론 일반적으로 의미가 명확하다고 생각되는 명사들에 대해서도 다수의 의미코드를 갖는 것을 허용한 것을 나타낸다. 다음 표 1은 현 시스템에서 사용되는 의미코드의 일부와 예제를 보여준다:

표 1. 명사 의미 계층 구조

| 레벨1 | 레벨2 | 레벨3 | 레벨4 | 예 |
|------|------|-----|-----|---------------------|
| 구체명사 | 생물 | 동물 | 포유류 | 개, 고양이, 호랑이, 사자 ... |
| | 식물기 | | | 뿌리, 줄기, 잎, 열매 |
| | 동물기 | | | 머리, 눈, 허파, 위, 경락 |
| | ... | | | ... |
| 추상명사 | 의무 | | | 책임, 과제, 숙제, 역할 |
| | 업적 | | | 공훈, 세계신기록, 공로, 구매실적 |
| | ... | | | ... |
| 활동명사 | 정신활동 | 생각 | | 기억, 인식, 주문, 증명, 착각 |
| | 종교활동 | | | 기도, 예배, 설교, 찬불 |
| | ... | | | ... |

¹ 강범모 (2001)에서는 명사와 공기하는 슬어에 대해 ‘적정 슬어’라는 자질을 도입하여 정보를 기술하고 있다

² 이 기준은 명사 의미의 Metonymy 현상을 설명하기 위해 필요하다

4. 국소 구문 관계를 고려한 의미 모호성 해소

기존의 Word Sense Tagging은 구조 분석의 어려움으로 인하여 문장의 구문 관계를 충분히 고려하지 못하고 단순히 주변 어휘들과의 공기 관계로 그 의미를 파악하려 했다. 물론 각 동사에 지배를 받는 논항의 핵심어는 지배 동사와 각 논항들 사이의 공기 관계로 그 의미 파악이 가능하다. 그러나 논항이 단일 명사가 아니라 복합 명사를 포함하는 명사구로 이루어진 경우, 그 명사구의 핵심어의 자매 성분들은 명사구 내에서 국소적으로 그 의미 관계를 파악해야 한다.

용언구에 전자 통신 연구원의 신입 연구원들이 대한 연구원을 교육할 것이다

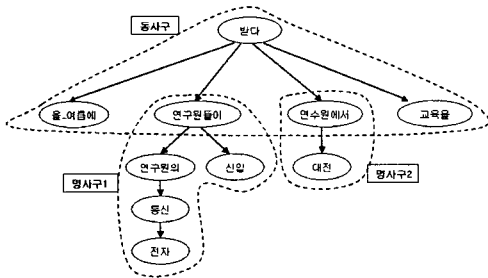


그림 1. 국소 구문 관계

예를 들어, “전자 통신 연구원에서 근무한다.” 라는 문장에서 “근무하다” 라고 하는 동사는 명사 어휘 “연구원” 과 “에서” 와는 밀접한 의미적 공기 관계를 가지지만, (통신, 근무하다), (전자, 근무하다) 는 명사구와 동사구의 국소 구문 관계를 넘어서는 공기 관계가 약한 정보이다. 대신 “연구원에서 근무하다” 라고 하는 동사구와 “전자 통신 연구원” 이라고 하는 명사구의 국소 구문 관계 내에서 각 단어들에 대한 공기 관계를 파악해야 한다. 따라서 WSD를 위한 정확한 의미 공기 관계 파악을 위해서는 한 문장에 있어 동사구와 명사구에 대한 의미 공기 관계의 단계별 적용이 필요하다.

1) 국소 구문 범위 결정

본 논문에서는 문장에서 한 동사가 지배하는 논항의 범위를 결정하기 위해서, 의존 구문 규칙 및 동사구 패턴의 한국어 격틀을 이용한 의존문법 파서를 사용하여 입력문의 단문 구조를 파악한다. 한국어 격틀의 사용은 이진 의미 패턴 적용과 격틀 적용의 두 단계로 나누어진다. 이진 의미 패턴은 명사어휘, 격조사, 동사 패턴을 격틀로부터 추출하여 의존관계를 결정할 때 적용하여 의존 트리의 수를 감소시키는 제약 조건으로 사용한다. 격틀의 적용은 격틀에서 기술하는 제약조건을 입력문과 비교하여 전체의 조건이 입력문과 일치하는 완전 매칭

과 일부 조건이 일치하는 부분 매칭에 따라 다른 가중치를 할당하여 단문 분석을 시도한다. 관형절의 경우 피수식어가 수식절의 격 정보를 채우게 되므로 피수식어를 포함한 수식절을 별도의 단문으로 취급하여 격틀을 적용해 본다.

구조 분석에 있어, 복합문의 수식 구조, 즉 동사구들간의 수식 관계는 현재 파악하기 어려운 실정이지만 의존 구문 규칙 및 한국어 격틀 정보를 유용하게 사용하면 정확한 단문 구조 인식은 가능하다.

2) 용언구 논항 핵심어의 의미 모호성 해소

용언구내에서의 의미 태깅이란 한 용언구의 구성 성분인 논항의 핵심어 어휘들의 리스트 $w_{1,n}$, 격 조사들의 리스트 $p_{1,n}$ 및 용언 v 가 주어졌을 때, 이 용언구에 대한 의미 확률을 최대로 하는 어휘들 ($w_{1,n}$)의 의미 리스트 $s_{1,n}$ 를 구하는 함수 WSD로 볼 수 있다. 이를 수식으로 나타내면 다음과 같다.

$$WSD_{vp}(w_{1,n}, p_{1,n}, v) = \operatorname{argmax}_{s_{1,n}} P(S_{1,n}=s_{1,n} \mid W_{1,n}=w_{1,n}, P_{1,n}=p_{1,n}, v) = \operatorname{argmax}_{s_{1,n}} P(s_{1,n} \mid w_{1,n}, p_{1,n}, v) = \operatorname{argmax}_{s_{1,n}} \frac{P(s_{1,n}, w_{1,n}, p_{1,n}, v)}{P(w_{1,n}, p_{1,n}, v)}$$

형태소 태깅에 있어 현재의 태그는 이전의 태그에만 의존한다는 HMM 모델의 Markov 가정과는 달리, 용언구의 각 논항은 인접 논항들보다는 문미의 용언과 공기하는 경향이 강하다. 따라서 본 논문에서 용언구내의 의미열에 대한 확률값은 각 논항의 핵심어 의미와 용언과의 공기값의 합을 반영하며, 식 (1)과 같이 정의한다.

$$WSD_{vp}(w_{1,n}, p_{1,n}, v) = \operatorname{argmax}_{s_{1,n}} P(s_{1,n}, w_{1,n}, p_{1,n}, v) = \operatorname{argmax}_{s_{1,n}} \sum_{i=1}^n P(s_i | w_i) * P(p_i, v | s_i, w_i) \approx \operatorname{argmax}_{s_{1,n}} \sum_{i=1}^n P(s_i | w_i) * P(p_i, v | s_i)$$

3) 명사구 의미 모호성 해소

명사구에 대한 정확한 의미 결정을 위해서는 명사구 내에서의 구문 관계를 고려해야 한다. 명사구에서는 용언구의 논항 핵심어가 명사구 구조의 뿌리 노드가 되는 구문 구조를 가지며 이때 단일 명사 또는 복합 명사가 관형격 조사나 콤마 등의 기호와 연결되어 하나의 노드를 형성한다. 예를 들어, 그림 3에서 “전자 통신 연구원의 신입 연구원” 이라는 명사구에서는 복합 명사인 “전자 통신 연구원” 과 관형격 조사 “의” 로 구성되는 관형격 논항과 명사구의 뿌리 노드인 “신입 연구원” 으로 구분된다. 따라서 명사구 NP의 뿌리 노드의 핵심어 “연구원” 과 논항의 핵심어 “연구원” 이 먼저 의

미 분별이 이루어져야 하고 이후 각 복합 명사 CN1, CN2 내에서의 명사 의미가 결정되어 진다.

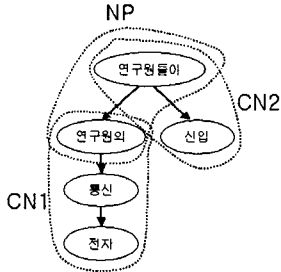


그림 2. 명사구의 2단계 구문 구조

관형격 조사와 그 명사의 핵심어들의 조합으로 M 개의 노드로 이루어진 명사구 $NP(w_{1p_1}, w_{2p_2}, \dots, w_{m-1p_{m-1}}, w_m)$ 에 있어서 명사구의 핵심어 w_m 의 의미는 동사구에서의 논항 의미 결정에 의해 s_m 으로 이미 결정되어 있다. 따라서 최적의 의미열 $s_{1,m-1}$ 을 탐색하기 위하여 통계적인 어휘 의미 확률 및 이진 의미 공기 정보를 이용하며, 식 (2)와 같이 각 의미들을 역방향(backward)으로 차례로 결정한다. 이때 식 (2)에서 의미와 어휘간의 공기 확률 정보는 데이터 부족 현상으로 인하여 반영하지 않았다.

$$\begin{aligned}
 WSD_{np}(w_{1,m-1}, p_{1,m-1}) &= \operatorname{argmax}_{s_{1,m-1}} P(s_{1,m-1}, w_{1,m-1}, p_{1,m-1}) \\
 &= \operatorname{argmax}_{s_{1,m-1}} \prod_{i=m-1}^1 \prod_{j=i+1}^m [\sum P(s_i | w_i) * P(s_j, p_j | s_i, w_i)] \\
 &= \operatorname{argmax}_{s_{1,m-1}} \prod_{i=m-1}^1 [\sum P(s_i | w_i) * P(s_j | s_i, w_i) * P(p_j | s_i, w_i, s_j)] \\
 &= \operatorname{argmax}_{s_{1,m-1}} \prod_{i=m-1}^1 [\sum P(s_i | w_i) * P(s_j | s_i) * P(p_j | s_i, s_j)] \\
 &- (2)
 \end{aligned}$$

명사구내의 각 논항의 핵심어 의미를 결정한 후, N개의 단어들로 이루어진 복합 명사 $w_1 w_2 \dots w_n$ 에 대한 의미를 결정한다. 이때, 복합 명사에서 핵심어 w_n 의 의미 또한 동사구 및 명사구의 논항 의미 결정에 의해 s_n 으로 이미 결정되어 있다. 따라서 통계적인 의미 공기 이진 관계를 고려하여 식 (2)에서 조사와의 공기 관계 $P(p_i | s_i, s_j)$ 를 제외하고는 같은 계산식으로 최적 의미열 $s_{1,n-1}$ 을 결정한다.

5. 실험 및 평가

한중 번역 시스템에서 한국 구문 분석 및 대역어 생성을 위해, 동사를 중심으로 하는 단문 중심의

동사구 패턴(고정 표현)에 대한 동사구 대역 패턴 구축하였다. 본 시스템에서 구축하는 동사구 대역 패턴은 패턴 내에 용언이 하나인 단일 용언형 대역 패턴을 말하며 현재 74,880 패턴이 구축되어 있으며 이로부터 의미 용언 공기 사전 SEM(의미, 조사, 용언)을 179,126쌍과 어휘 의미 빈도 정보를 159,296번 반영하였다. 동사구 패턴의 예를 살펴보면 아래와 같다.

(예1) 용례: 철수가 빵을 먹었다.

A=사람[철수]!가 B=식품[빵]!를 먹!다 > A
 吃:v B :: 먹다1-1

(예2) 용례: 나는 그녀에게 호감이 갔다

A=사람[나]!가 B=사람[그녀]!에게 호감!가
 가!다 > A □B □生:v 好感 :: 가다1-2

STEP 2000, ETRI 품사 태그드 코퍼스 1,500,000 어절을 대상으로 구문 분석의 모호성이 발생하지 않는 문미의 동사구에 대해 반자동으로 의미 부착 말뭉치를 구하였다. 의미 용언 공기 사전 SEM(명사 의미, 조사, 용언)을 29,841쌍과 명사구의 의미 공기 사전 SEM(명사 의미, 조사, 명사 의미) 68,847쌍, SEM(명사 의미, 명사 의미) 33,197쌍을 추출하였다. 그림 3에서는 품사 태그드 말뭉치로부터 의미 모호성이 발생하지 않는 문미 동사구에 대한 의미 부착 말뭉치 자동 추출의 일례를 보이고 있다.

그 말이 떨어지기 무섭게 그녀는 발작을 일으켰다.
 그/mm 말/nc+ 이/jc 떨어지/pv+ 기/etn 무섭/pt+ 게
 /ec 그녀/np+ 가/jc 발작/nc+ 을/jc 일으키/pv+ 었
 /ep+ 다/ef+ /s
 ->
 그녀/np+ 가/jc 발작/nc+ 을/jc 일으키/pv+ 었/ep+ 다
 /ef+ /s
 ->
 A=사람[그녀]!가 B=유생물상태[발작]!를 일으키다

그림 3. 의미 부착 말뭉치 자동 추출의 일례

STEP 2000, ETRI 원시 말뭉치에서 의미 부착 말뭉치 구축에 사용되지 않은 분야별 1,000문장에 대하여 의미 분석을 수행하였다. 분석 방식은 형태소 분석 및 구문 분석을 거쳤으며 일반 보통 명사를 대상으로 의미 모호성 해소를 시도하였다.

표 2. 명사 의미 모호성 해소 실험 결과

| | |
|-----------------|-------|
| 문장수 | 1,000 |
| 평균 어절수 | 13.2 |
| 의미 2개 이상인 명사 어휘 | 1,740 |
| 평균 의미수 | 2.7 |
| 에러수 | 270 |

| | |
|-----|--------|
| 정확률 | 83.9 % |
|-----|--------|

형태소 분석의 오류, 구조 분석의 오류, 용언구의 의미 공기 정보 부족, 명사구 의미 공기 정보 부족 등에 의해 의미 오분석된 경우가 있었으며 구조 분석기의 동사구의 범위를 인식 오류로 인하여 의미 공기 정보가 적용되지 않았다.

| |
|---|
| <p><형태소 분석 오류에 의한 오적용 일례> 배/N[식품:梨子, 교통:船, 동물기관:胃, 단위:倍, 행사활동:杯子]+로/P 물건/N[생활용품:日用品, 무생물성질:□量]+를/P 실/V+어/E 나르기/N[이동:搬運]+가/P 줄/A+있/PE +다/E+./S</p> <p><구조 분석 오류에 의한 오적용 일례> 이/ADN 곳/N[장소:地方]+는/P 물/N[액체:水, 광물:水, 자원:洪水, 약품:染料]+관리/N[지위:官□, 사람:官□, 관리:管理]+에/P 편리하/A+도록/E 수로/N[장소:水路]+가/P 잘/ADV 갖추어지/V+어/E 있/V+더/E+ ...</p> <p><의미 공기 정보 부족의 오적용 일례> 수도/N[장소:首都, 조직:水道, 건축물:水道, 도구:水道]+가/P 없/A+는/E 곳/DN+에는/P 수도/N[도구:水道, 건축물:水道, 조직:水道, 장소:首都]+를/P 놓/V+아/E 주/V+고/E+ ...</p> |
|---|

그림 4. 의미 모호성 해소 실패의 일례

그림 4는 의미 모호성 해소가 실패한 일례들을 보이고 있다. 분석 결과에서 첫번째 의미코드가 선택된 제 1후보이다. 형태소 분석 오류에서는 “실다”가 “실다”로 오분석되어 공기 정보가 적용되지 못하였으며, 구조 분석 오류에서는 “관리에”가 용언 “편리하”가 아닌 “갖추어지”에 잘 못 연결되어 의미 공기 정보가 오적용된 예를 보이고 있다. 그리고 용언구 및 명사구의 의미 공기 정보 부족으로 인해 의미 변별이 실패하는 경우도 있었다.

오류 유형을 분석해 본 결과, 용언구 논항에 대한 의미 모호성 해소는 비교적 잘 이루어진 반면 명사구는 데이터 부족으로 인하여 그 정확률이 다소 떨어졌다. 향후 의미 부착 말뭉치를 추가 확보함으로써 어휘 의미 확률 및 어휘 의미 공기 정보 등을 추가 반영함으로써 이러한 오류를 줄일 수 있을 것으로 판단된다.

6. 결론

기존의 의미 공기 정보를 이용하는 방법은 구문 분석의 어려움으로 인하여 특정한 범위에 나타나는 단어들에 대한 문맥 정보를 이용하였으며 이로 인해 문장의 구문 정보를 반영하지 못하였다. 본 논문에서는 의미 공기 정보를 이용하는 의미 모호성 방법 뿐만 아니라 구문 정보를 반영한 국소 문맥

기반 의미 모호성 방식을 제안하였다.

앞으로 의미 부착 말뭉치를 현재까지 확보된 의미 공기 사전을 이용하여 반자동으로 대량으로 구축할 예정이며, 어휘 의미 확률 값을 충분히 반영할 예정이다. 그리고 데이터 부족 현상을 해소하고 동시에 의미 변별력을 가지기 위한 의미 코드 정비 작업 또한 계속 진행할 예정이다.

7. 참고 문헌

- [1] 허정, 옥철영, “사전의 뜻풀이말에서 추출한 의미 정보에 기반한 동형의어 중의성 해결 시스템”, 정보과학회논문지 제 28권 제 9호, pp688-698, 2001
- [2] 이왕우, 이재홍, 이수동, 옥철영, 김철기, “Bayes 정리에 기반한 개선된 동형의어 분별 모델”, 제 13회 한글 및 한국어 정보처리 학술대회, 2001
- [3] 송영빈, 최기선, “동사의 애매성 해소를 위한 시소러스의 이용과 한계”, 제 12회 한글 및 한국어 정보처리 학술대회, 2000
- [4] 이창기, 이근배, “의미 애매성 해소를 이용한 WordNet 자동 매핑”, 제 12회 한글 및 한국어 정보처리 학술대회, 2000
- [5] David Yarowsky, “Unsupervised Word Sense Disambiguation Rivaling Supervised Method”, 33rd Annual Meeting of the ACL, pp189-196, 1995
- [6] Saim Shin, Youn-Seok Choi, Key-son Choi, “Word Sense Disambiguation Using Vectors of Co-occurrence Information”, Proceedings of the 6th Natural Language Processing Pacific Rim Symposium, Tokyo, Japan, 2001
- [7] Jong-Hoon Oh and Key-Sun Choi, “Word Sense Disambiguation using Static and Dynamic Sense Vectors”, Proceedings of the 19th International Conference on Computational Linguistics, Taipei, Taiwan, 2002.
- [8] 신지현, 최민우, 강범모, “어휘의미분석 말뭉치 구축의 절차와 문제”, 제 13회 한글 및 한국어 정보처리 학술대회, 2001
- [9] Miller G (1990) “Wordnet: An On-line lexical database”, International Journal of Lexicography, 1990
- [10] 강범모, 박동호, 이성현, 박진호, “한국어 명사 의미 부류 체계의 구축과 활용”, 한글 및 한국어 정보처리 학술대회, 2001
- [11] Pustejovsky, J., The Generative Lexicon, The MIT Press, Cambridge/London
- [12] 홍문표, 최승권, “다국어 자동번역시스템에서 부사격 조사의 올바른 대역어 선정을 위한 언어학적 모델링”, 제 13회 한글 및 한국어 정보처리 학술대회, 2001