

# 단어링크와 공기 단어를 이용한 의미중의성 해소

구영석<sup>0</sup>, \*나동렬  
연세대학교 컴퓨터과학·산업시스템공학과  
\*연세대학교 전산학과  
hey90tan@december.yonsei.ac.kr  
\*dyra@dragon.yonsei.ac.kr

## Word Sense Disambiguation Using Word Link and Word Cooccurrence

Yeong-Seok Ku<sup>0</sup>, \*Dong-Yul Ra  
Dept. of Computer Science, Yonsei University

### 요 약

본 논문은 문장 안에서 의미 중의성을 갖는 단어가 출현했을 때 그 단어가 어떤 의미로 사용되고 있는지 판별해 주는 방법을 제시하고자 한다. 이를 위해서 먼저 중의적 의미를 가지는 단어의 각 의미 (sense) 마다에 대하여 이 의미를 나타내는 주요단어 즉 종자단어와 연관성이 있는 단어들로 벡터를 구성하여 이 의미를 나타내게 한다. 종자단어와 말뭉치의 문장을 통하여 연결된 경로를 가진 단어는 이 종자단어에 해당하는 의미를 나타내는 데 기여하는 정보로 본 것이다. 경로는 동일 문장에서 나타나는 두 단어 사이는 링크가 있다고 보고 이러한 링크를 통하여 이루어 질 수 있는 연결 관계를 나타낸다. 이 기법의 장점은 데이터 부족으로 야기되는 문제를 경감시킬 수 있다는 점이다. 실험을 위해 Hantec 품사 부착된 말뭉치를 이용하여 의미정보벡터를 구축하였으며 ETRI 품사 부착된 말뭉치에서 중의적 단어가 포함된 문장을 추출하여 실시하였다. 실험 결과 기존의 방법보다 나은 성능을 보임이 밝혀졌다.

### 1. 서 론

단어의 의미 중의성 해소(word sense disambiguation)는 주어진 문맥안에서 해당 단어의 의미를 정확히 파악하는 작업이다. 단어의 중의적 의미를 파악해 낼 수 있다면 기계번역 시스템이나 정보 검색 시스템의 성능을 향상시켜 줄 수 있다. 중의적단어의 의미중의성 해소에 대한 기존의 연구방법들을 살펴보면 크게 지식기반의 방법과 데이터기반의 방법으로 나누어 볼 수 있다.

지식기반의 방법에는 의미계층구조를 이용하거

나 확률통계의 방법을 사용하여 의미정보를 확보한 다음 단어의 의미중의성을 해결한다. 의미계층구조를 이용하는 방법은 영어권에서는 Roget thesaurus나 Wordnet을 이용하여 의미중의성을 해결하는 반면 한국어에서는 각 연구실마다 자체적으로 시소러스를 구축하여 연구에 사용하고 있는 실정이다. 확률통계의 방법은 불규칙한 데이터의 출현을 반영하여 가장 빈도가 높은 것으로 의미부여를 하는 방법이다. 하지만 확률통계의 방법의 단점은 자료부족현상(sparse data problem)이 있을 때에는 기대이하의 결과가 나온다는 점이다.

데이터기반의 방법에는 사전을 이용하거나 대량의 말뭉치를 이용하는 방법이 있는데 이는 말뭉치를 이용하는 방법은 다시 품사 부착된 말뭉치를 이용하는 교사학습 방법(supervised learning algorithm)과 품사가 부착되지 않은 말뭉치를 이용하는 비교사 학습 방법(unsupervised learning algorithm)으로 분류할 수 있다.

[3]은 국소문맥을 사용하여 만들어진 Decision List를 통해 단어의 형태적 중의성을 제거하는 방법을 소개하였다. 이 방법은 최초의 종자 언어(seed collocation)로 1차 Decision List를 만들어 실험 말뭉치에 적용하고 태깅된 결과를 자가 학습하는 반복과정에 의해 Decision List의 수행능력을 향상시켰다. 이 방법은 단어의 형태적 중의성 제거에 일정 거리의 연어가 가장 큰 영향을 끼친다는 직관에 바탕을 두어서 비교사 학습 방식으로 수행되었다.

[2]는 방대한 untagged corpus를 사용하여 비교사 학습 방법을 가지고 단어의 의미중의성을 해소하였다. [2]는 두 가지 가정을 하였는데 첫 번째는 "one sense per collocation"로 단어들은 주어진 연어에 의해 한 가지의 의미로 결정될 수 있다는 것이며, 두 번째는 "one sense per discourse"로 각 담화 범위의 동일한 단어들은 모두 같은 의미로 쓰인다는 것이다.

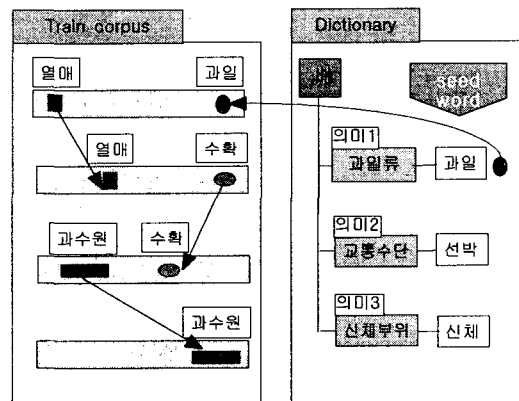
2장에서는 단어링크와 공기 정보를 사용하여 한국어 명사의 의미중의성 해소를 하는 방법에 대한 모델을 제시한다. 3장에서는 단어링크와 공기 단어를 이용하여 의미중의성을 해소하는 시스템의 구성을 설명한다. 4장에서는 실험 및 평가를 한다. 5장에서는 결론 사항을 언급한다.

## 2. 의미중의성 해소 모델

### 2.1 단어링크와 공기 단어를 이용한 모델

본 논문에서는 단어들이 문장에서 여러 번 출현할 때 서로 연관이 있다고 가정한다.

예를 들어 의미중의성이 있는 단어 '배'에 대해 사전을 기반으로 의미를 분류해 보면 '과일류', '교통수단', '신체부위'로 나누어 볼 수 있다. 또한 의미대표단어로서 '과일류'의 의미에 대해서는 '과일'을, '교통수단'의 의미에 대해서는 '선박'을, '신체부위'의 의미에 대해서는 '신체'를, 각각 종자단어(seed word)로 선정하여 준다.

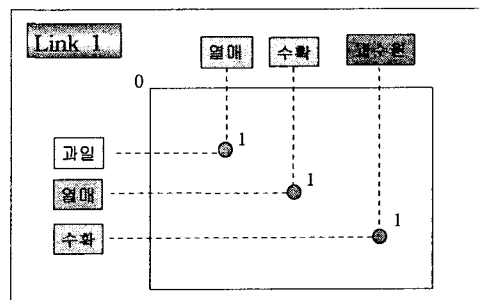


[그림 1] 단어링크와 공기 단어를 이용한 모델

'과일류'의 의미를 가지는 '과일'이라는 종자단어를 가지고 대량의 말뭉치에서 '과일'이 들어간 문장을 추출하고 공기 단어를 추출하게 된다. 본 논문에서 공기 단어는 문장 안에서 자주 출현하는 의미 있는 단어를 말한다. 예를 들어

"이 열매는 당분이 많고 시원한 과일이다"

라는 문장이 있다고 하면, 이 문장에서 공기 단어로 추출되는 것은 '과일', '열매', '당분', 등의 체언류와 '많', '시원하' 등의 용언류가 대상이 된다. '과일'에 대해서 공기 단어를 추출하고 여기에서 구한 체언류와 용언류에 대해 의미백터를 구하게 되고 그런 다음에는 체언들 중에서 다음 링크가 될 수 있는 링크대상단어를 선정하게 된다.



[그림 2] 동일 문장 안에서의 공기 단어 정보 추출

'과일'과 가장 연관이 있는 단어가 '열매'라고 가정해 보자. 이번에는 '열매'가 종자단어처럼 키워드가 되어서 다시 대량의 말뭉치로부터 '열매'가 들어있는 문장들을 추출하고 그 문장들 안에서 함께

자주 나오면서 의미가 있는 단어들이 공기 단어로 추출된다. 추출된 공기 단어들은 체언류와 용언류로 분류되어 각각 의미벡터로 저장된다. '열매'를 포함하는 문장들 중에

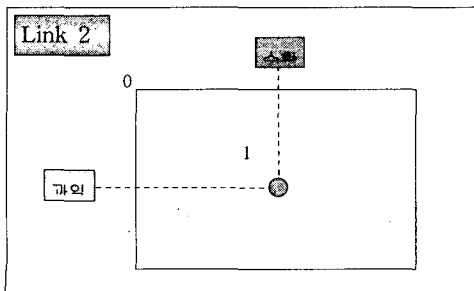
"담스러운 열매를 농부가 수확하고 있다"

라는 문장이 있다고 하자. 이 문장에서 공기 단어로 추출되는 것은 '열매', '농부', '수확' 등의 체언류와 '담스럽' 등의 용언류가 대상이 된다. 여기에서도 '열매'와 '농부', '수확'은 링크를 거치지 않은 문장 안에서 직접 함께 나타나는 공기 단어이다.  $\therefore$  음으로

"과수원에서 매년 수확하는 생산량이 늘고 있다"

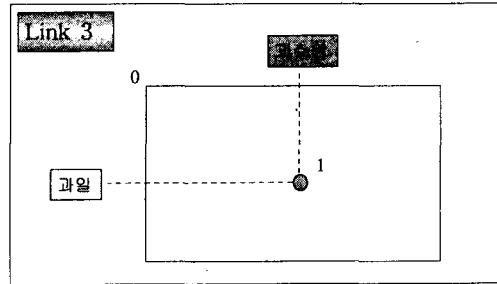
라는 문장이 있을 때 이 문장에서 공기 단어로 추출되는 것은 '과수원', '매년', '수확', '생산량' 등(3) 체언류와 '늘' 등의 용언류가 대상이 된다. 여기에서도 이런 단어들은 '수확'이라는 단어에 대해서 링크를 거치지 않은 문장 안에서 직접 함께 나타나는 공기 단어이다. 이렇게 동일한 문장에서 공기하는 단어쌍 사이에 관계를 1 개의 링크를 통한 관계로 보며 [그림2]처럼 나타내어진다.

여기에서 두 단어 '과일'과 '수확'이라는 두 단어 사이의 관계를 살펴 보자. [그림 2]에서와 같이 '과일'과 '열매'가 1개의 링크를 통해 연결되어 있고, '열매'와 '수확'이 1개의 링크를 통하여 연결되어 있다. 따라서 '과일'과 '수확'은 링크를 2번 통해서 연결되어 (공기하여) 있다고 말할 수 있다. 다시 말하면 '과일'이라는 단어와 '수확'이라는 단어는 서로 동일한 문장 안에서 한번도 나타난 적이 없지만 '열매'라는 링크 단어를 통해서 공기 단어한 것으로 본 것이다. [그림 3]은 이런 상황을 나타낸다.



[그림 3] 링크 2번을 통한 공기 단어 추출

이번에는 '과일'과 '과수원' 사이의 관계를 살펴 보자. 이 두 단어가 비록 동일 문장에는 나타나지 않았더라도 다른 단어 ('열매', '수확')을 통하여 연결되어 있음을 알 수 있다. 이 두 단어 사이에는 3 개의 문장에서의 공기 관계를 통하여 연결되어 있다고 볼 수 있으므로 [그림 4]와 같이 링크 3개를 통한 공기관계로 볼 수 있다.



[그림 4] 링크 3번을 통한 공기 단어 추출

링크를 통한 의미정보의 구축방법은 단어의 의미중의성 해소를 위한 연구에서 가장 큰 문제가 되는 자료 부족 문제를 해결해 주는 열쇠가 된다.

## 2.2 중의적 단어의 의미결정

본 논문에서 사용한 의미중의성 해소를 위한 최종적인 수식은 다음과 같다.

$$WSD(S_{selected}) = ARG \ MAX \ S_i \ Sim(C, S_k) \quad (1)$$

C는 테스트문장을 나타낸다. 주어진 테스트 문장을 다음과 같은 벡터로 나타낼 수 있다.

$$C = [x_1, x_2, x_3, \dots, x_i, \dots, x_N]$$

여기에서  $x_i$ 는 단어 번호  $i$  인 단어가 테스트문장에 나타났으면 1 아니면 0 이다.

$S_k$ 는 중의적 단어의 한 의미를 나타내는데 역시 다음과 같은 벡터로 나타낼 수 있다.

$$S_k = [v_1, v_2, v_3, \dots, v_i, \dots, v_N]$$

예를 들어 의미 중의성을 가지는 단어가 '배' 라고 하고 이에 대한 3가지 의미(sense)의 의미대표단어를 각각 '선박', '과일', '신체'라고 한다면  $S_1$ 는 '선

박',  $S_2$ 는 '파일',  $S_3$ 는 '신체'에 대한 의미를 나타낸다.

$Sim(C, S_k)$ 는 테스트 문장( $C$ )벡터와 의미정보벡터 ( $S_k$ ) 사이의 유사도를 나타낸다.  $Sim(C, S_k)$ 는 다시 체언간의 유사도  $Sim(C_N, S_{k_N})$ 와 용언간의 유사도  $Sim(C_V, S_{k_V})$ 로 나누어 볼 수 있다.

$$Sim(C, S_k) = Sim(C_N, S_{k_N}) + Sim(C_V, S_{k_V})$$

$C_N$ 은 테스트문장에서의 체언류를 추출하여 벡터값으로 가진다.  $C_V$ 는 테스트문장에서의 용언류를 추출하여 벡터값으로 가진다.  $S_{k_N}$ 는 중의적 단어의 사전적인 의미를 나타내는 정보벡터 중에서 체언류만 포함한 것이다.  $S_{k_V}$ 는 중의적 단어의 사전적인 의미를 나타내는 정보벡터 중에서 용언류만을 포함한 것이다. 각각 체언벡터끼리, 용언벡터끼리 식 (5)처럼 코사인 유사도 계산을 한 다음에 이를 더하게 된다.

$$Sim(C, S_k) = \frac{\sum_{i=1}^N x_i \cdot v_i}{\sqrt{\sum_{i=1}^N x_i^2} \cdot \sqrt{\sum_{i=1}^N v_i^2}}$$

$S_k$  벡터를 구하는 구체적인 방법에 대하여 알아보자.

$$S_k = [Sim(w_1, S_{k_{sw}}), \dots, Sim(w_n, S_{k_{sw}})]$$

$S_{k_{sw}}$ 는 중의적 단어의 한 의미에 대한 종자단어(의미대표단어)이다.  $w_i$ 는 이 시스템이 사전에 보유한 모든 단어 중 번호가  $i$ 인 단어를 나타낸다(4) 즉  $v_i$  값은

$$Sim(w_i, S_{k_{sw}}) = \sum_{j=1}^T weight_j \times Link_j$$

과 같이 구한다. 이것은 의미대표단어( $S_{k_{sw}}$ )와 단어( $w_i$ ) 사이의 말뭉치를 통한 연관도 즉 공기 관계 정도를 나타낸다.  $weight_j$ 는 링크를  $j$ 번 사용한 공기 관계에 대한 가중치로서  $j$ 가 클수록 줄어들게 된다. 여기에서  $T$ 는 고려할 최대 링크의 수,  $sw$ 는 체언 또는 용언을 나타낸다.

$$Link_j = \text{number of paths with } j \text{ links}$$

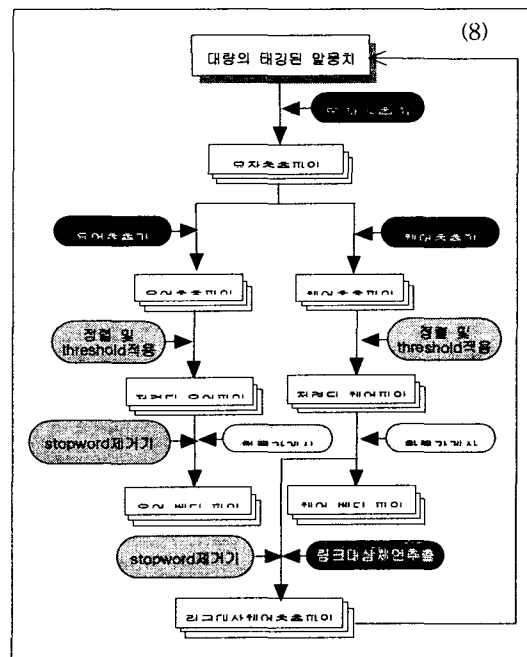
$Link_j$ 는 의미대표단어와 단어  $w_i$  사이에  $j$ 개의 링크를 통하여 연결된 경로의 총수이다. (5)

### 3. 의미중의성 시스템

#### 3.1 의미중의성 해소 시스템

중의적 단어의 의미중의성 해소 시스템(6) 구성도는 [그림 5]와 같다. 구성도를 보면 우선 대량의 품사 부착된 말뭉치에서 문장추출기를 통해 중의적 대상이 되는 체언단어를 키워드로 하여 문장들을 추출하게 된다. 이렇게 추출된 문장을 파일로 보관하게 되고 체언벡터파일을 구하는 단계와 용언벡터를 구하는 단계로 나누어지게 된다.

우선 체언벡터를 구하는 과정을 살펴보자. 문장추출파일에서 체언추출기를 통해서 체언(7)만을 추출하게 된다. 추출된 체언들은 각각 카운트를 하여 빈도를 계산하고 1번 이하로 나타난 체언들은 제거한 다음 빈도순으로 정렬을 하게 된다. 그 다음에는 확률값을 계산하여 체언벡터파일에 저장하게 된다. 또한 다음링크대상체언을 추출하기 위해서 링크대상체언추출기가 사용되는데 이때 불용어리스트가 필요하다. 체언에 대한 불용어리스트는 한국어와 영어로 이루어져 있으며 무의미한 체언들이 링크의 대상이 되는 것을 막는다.



[그림 5] 중의적 단어에 대한 의미중의성 해소 시스템

다음은 용언벡터추출하는 과정을 살펴보자. 우선 문장추출기에 의해 문장이 추출되면 이로부터 용언추출기를 사용하여 용언을 추출하게 된다. 용언이 추출되면 용언의 빈도를 계산하고 임계값이하를 버린다. 용언벡터가 되기 위해서는 역시 불용어리스트가 필요한데 용언에 대해 사용하는 것은 한국어에 있어서 고빈도단어 불용어리스트를 사용하여 불용어를 제거하게 된다. 예를 들면 ‘있’, ‘하’, 등의 용언은 한국어 문장에 있어서 가장 많이 나오는 용언들 중에 하나이다.

용언벡터가 완성되고 체언벡터가 완성되면 다음은 링크대상체언추출과일을 근거로 하여 다음링크를 출발하게 되고 이러한 링크대상체언추출과일에 있는 단어가 다시 키워드가 되어서 대량의 품사 부착된 말뭉치에서 문장추출을 하고 체언벡터와 용언벡터를 확보할 수 있게 된다.

### 3.2 불용어리스트 적용하기

불용어란 단어의 출현 빈도에 의한 중요도 평가를 함에 있어서 장애가 되는 것이 있는데 바로 문장 가운데에서 문장의 의미 전달에 큰 영향을 미치지 못하는 단어들을 말한다. 이러한 불용어들은 미리 제거되어야 성능 저하를 막을 수 있다. 일반적으로 전치사, 조사, 관사, 접속사 등과 같이 주제어(topic term)로서 의미가 없는 단어들을 불용어라고 할 수 있다. 명사 중에서도 아주 일반적이고 모든 문장에 의미없이 나타나는 단어들은 불용어로 제거되어야 한다.

본 논문에서는 연세대 한글정보처리 연구실에서 프로젝트로 구축한 불용어리스트를 이용하여 체언에 대해서는 영어의 불용어리스트와 한국어의 체언 불용어리스트를 사용하였고 용언에 대해서는 한국어의 고빈도 불용어리스트를 사용하여 불필요한 단어들을 제거하였다.

## 4. 실험 및 평가

### 4.1 실험 대상

실험에 사용된 말뭉치는 Hantec 품사 부착된 말뭉치를 가지고서 의미정보를 구축하였으며 ETRI 품사 부착된 말뭉치를 가지고 외부실험을 실시하였다.

[표 1] 실험 말뭉치

구분	말뭉치대상	어절 개수	형태소 개수
의미정보 구축	Hantec 품사 부착된 말뭉치	16,144,180개	30,824,320개
외부실험	ETRI 품사 부착된 말뭉치	561,129개	1,185,167개

[표 2]는 중의적 단어 3개를 선정하여 의미분류를 하였으며 종자단어후보를 각 의미 분류별로 제시한 다음 각각 종자단어 후보들이 실험 말뭉치에서 가장 출현빈도가 많은 것으로 최종 종자단어로 선정하도록 한다.

[표 2] 중의적단어와 의미분류 및 종자단어 후보군

중의적 단어	의미 분류	종자단어 후보군
배	과일류	과일 - 수박 - 나주배
	교통수단	어선 - 유람선 - 나룻배
	신체부위	신체 - 아랫배 - 내장
눈	기상현상	눈보라 - 북극 - 눈사람
	신체부위	시력 - 안과 - 눈병
차	교통수단	소방차 - 자가용 - 전용차
	음료	녹차 - 커피 - 엽차

[표 3]은 중의적단어 ‘배’에 대해서 과일류에는 ‘과일’이, 교통수단에는 ‘어선’이, 신체부위에는 ‘신체’가 종자단어로 선정되었다.

[표 3] 중의적단어 ‘배’와 종자단어 선정

중의적 단어	의미 분류	종자단어 후보	Training 말뭉치에서의			
			후보 단어출현 문장개수	공기 체언 개수	공기 용언 개수	링크 대상 체언개 수
배	과일류	과일	355	1679	231	1511
		수박	130	621	81	546
		나주배	1	5	0	4
	교통수단	어선	652	2014	326	1809
		유람선	94	460	66	408
		나룻배	9	43	13	38
	신체부위	신체	690	3655	347	3385
		아랫배	4	1	1	0
		내장	284	1976	145	1825

[표 4]는 중의적단어 ‘눈’에 대해서 기상현상에는 ‘북극’이, 신체부위에는 ‘시력’이 종자단어로 선정되었다.

종자단어 후보 중에서 ‘눈사람’의 경우를 보면 출현문장은 3개이지만 공기 체언 개수가 1이고 공

[표 4] 중의적단어 '눈'과 종자단어 선정

중의적 단어	의미분류	종자단어 후보	Training 말뭉치에서의			
			후보 단어출현 문장개수	공기 체언개수	공기 용언개수	링크대상 체언개수
눈	기상현상	눈보라	5	3	4	1
		북극	49	294	34	265
		눈사람	3	1	3	0
	신체부위	시력	78	316	45	280
		안과	65	230	53	198
	눈병	10	10	5	7	

기 용언개수가 3인 이  
유는 1  
번 이하  
로 나오  
는 것들  
은 제외  
되었기  
때문이다. 또한

링크대상 체언개수가 0인 이유는 불용어리스트에 의해 체언이 제거되었기 때문이다.

[표 5]는 중의적단어 '차'에 대해서 교통수단에는 '자가용'이, 음료에는 '커피'가 종자단어로 선정되었다

[표 5] 중의적단어 '차'와 종자단어 선정

중의적 단어	의미분류	종자단어 후보	Training 말뭉치에서의			
			후보 단어출현 문장개수	공기 체언개수	공기 용언개수	링크대상 체언개수
차	교통수단	소방차	194	543	140	462
		자가용	378	1641	199	1467
		전용차	71	463	39	408
	음료	녹차	18	118	19	98
		커피	329	1466	192	1306
	엽차	0	0	0	0	

#### 4.2 성능 평가

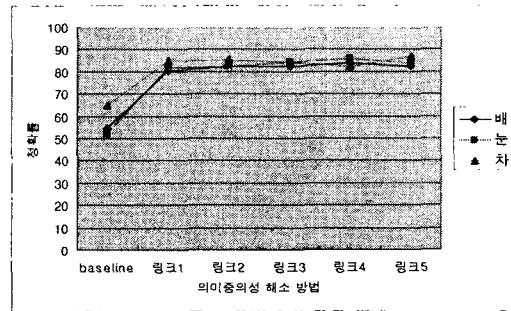
[표 6]은 ETRI 말뭉치에 대한 외부실험 테스트 결과이다. 중의적 단어 '배', '눈', '차'에 대해서 체언의미벡터와 용언의미벡터를 구한 다음 테스트 문장이 들어오면 테스트 문장에서도 용언과 체언으로 분류한 다음 각각을 유사도 계산하여서 유사도 값이 가장 큰 값을 나타내는 의미대표단어의 의미를 결정하였다.

[표 6] ETRI 말뭉치에 대한 테스트 결과

중의적 단어	test 문장 개수	baseline 정확률 (%)	정확률(%)				
			링크1	링크2	링크3	링크4	링크5
배	108	54.73	80.24	82.5	82.73	<b>83.78</b>	82.43
눈	629	52.27	82.06	82.35	84.62	<b>85.86</b>	84.25
차	457	65.16	85.29	85.63	84.6	82.57	<b>86.93</b>
평균	398	57.38	82.53	83.49	83.86	84.07	84.53

'baseline'은 중의적 단어에 대해서 가장 빈번하게

나타난 의미를 가지고 테스트 문장에 있는 중의적 단어의 의미로 결정했을 때의 정확률이다.



[그림 6] 중의적 단어에 대한 의미중의성 해소 결과

#### 5. 결론 및 요약

본 논문은 말뭉치의 공기 정보를 이용하여 단어 의미 중의성을 해소하는 기법을 제시하였다. 비록 두 단어가 같은 문장에서 공기하지 않았더라도 동일 문장에 나타난 두 단어 사이에 설정되는 단어 링크를 통하여 여러 링크를 거쳐 두 단어가 연결되는 경로가 있다면 공기 관계가 있는 것으로 간주되고 이는 단어의 의미를 나타내는 데 이용된다.

실험과정에서 링크를 한번도 가지 않은 경우에는 확보되는 체언벡터의 개수와 용언벡터의 개수가 작는데 비해서 링크대상 체언단어를 통하여 링크를 확장할수록 해당 의미의 하부구조로써 체언벡터의 개수와 용언벡터의 개수가 상당히 확보할 수 있음을 알 수 있었다. 이것은 자연어처리 연구에 있어서 가장 문제가 되는 자료부족문제를 해결하는 좋은 방법이 된다.

그러나 종자단어를 선정할 때에는 주의를 해야 하는데 종자단어를 가지고 처음 시작을 하기 때문에 해당 말뭉치에서 종자단어가 한번도 나오지 않는다면 문장을 추출할 수 없고 그래서 문장을 추출할 수 없기 때문에 체언이나 용언을 구할 수 없을 뿐만 아니라 링크대상 체언단어를 확보해서 확장을 할 수가 없다. 본 논문에서는 종자단어를 선정함에 있어서 후보단어를 선정하여 말뭉치에서 해당 후보단어가 들어있는 문장들을 추출하고 또한 1번 이하 나오는 것을 제외한 체언과 용언을 구하고 링크대상 체언을 구하여 비교한 다음 가장 우수한 단어를 종자단어로 선택하여 의미중의성 해소 시스템에 사용하였다.

추가적인 연구가 필요한 사항으로는 본 연구 결과를 개선하기 위한 한 방법으로 사전의 뜻풀이 말을 이용하는 것이다[5]. 즉 각 의미마다 여러 개

의 종자 단어가 가능하도록 하며 이들을 의미의 뜻풀이 말에 나타난 주요 단어들로 하는 것이다.

### 참고 문헌

- [1] Yarowsky, David. "Word-sense Disambiguation using Statistical Models of Roget's Categories Trained on Large Corpora." In Proceedings of COLING92, pp.454~460, 1992
- [2] Yarowsky, David. "Unsupervised word sense disambiguation rivaling supervised methods", in Proceedings of the 33rd Annual Meeting, pp.189-196, Cambridge, MA, June. Association for Computational Linguistics, 1995
- [3] 이충희, 송만석, "한국어 자연어 처리에서의 국소 문맥을 이용한 형태적 중의성 해소", 제12회 한글 및 한국어 정보처리 학술대회 발표집, pp.48~55, 2000
- [4] 이승우, 이근배, "국소문맥과 공기 정보를 이용한 비교사 학습 방식의 명사 의미 중의성 해소", 정보과학회 논문지 : 소프트웨어 및 응용 제27권 제7호, pp.769-783, 2000
- [5] 허정, 옥철영, "사전 뜻풀이말에서 추출한 의미 정보에 기반한 의미 중의성 해결", 제12회 한글 및 한국어 정보처리 학술대회 발표집, pp.269~276, 2000
- [6] 신사임, 최기선, "공기정보 벡터를 이용한 한국어 명사의 의미구분", 제13회 한글 및 한국어 정보처리 학술대회 발표집, pp.472~478, 2001
- [7] 서희철, 임해창, "유사어를 이용한 단어 의미 중의성 해결", 제11회 한글 및 한국어 정보처리 학술대회 발표집, pp.304~309, 1999
- [8] 임수종, 송만석, "가중치 정보를 이용한 한국어 동사의 의미 중의성 해소", 제10회 한글 및 한국어 정보처리 학술대회 발표집, pp.425~429, 1998
- [9] 박영자, 송만석, "사전에서 추출한 의미 속성에 기반한 명사 의미 클러스터링", 정보과학회논문지(B) 제25권 제3호 pp.585~595, 1998
- [10] 이호, 임해창, "분류 정보를 이용한 단어 의미 중의성 해결", 정보과학회논문지(B) 제24권 제7호, pp.779~789, 1997