

# 분포 정보를 이용한 의미 중의성을 지닌 한국어 동사의 의미 분별

조정미\*○ 김길창\*  
\* 한국과학기술원 전산학과

## Korean Verb Senses Disambiguation Using Distributional Information from Corpora

JeongMi Cho\*○ Gil Chang Kim\*  
\* Dept. of Computer Science, KAIST

### 요 약

본 논문에서는 말뭉치로부터의 분포 정보를 이용하여, 의미 중의성을 지닌 한국어 동사의 의미를 분별하고자 한다. 분포 정보란 말뭉치내에서 목적어-서술어 관계에 있는 명사와 동사의 분포를 의미한다. 이 분포 정보는 명사 분포와 동사 분포로 나누어 생각할 수 있는데, 본 논문에서는 이 두가지 분포 정보를 사용함으로써 명사 분포만을 이용하였을 때 나타나는 자료 부족 현상을 동사 분포 정보를 이용하여 보완하였다. 분포 정보간의 유사도 계산은 정보 이론에서 사용하는 상대 엔트로피를 이용한다. 품사가 태깅된 50만 단어의 한국어 말뭉치로부터 분포 정보들을 추출하여 한국어 동사 10개에 대해 실험하였다.

### 1. 서론

본 논문\*은 말뭉치로부터의 분포 정보를 이용하여 의미 중의성을 지닌 한국어 동사의 의미를 분별해내고자 한다.

우리는 일상 대화에서 의미 중의성을 지닌 많은 동사들을 사용하고 있다. 예를 들면, '점심을 들었다', '가방을 들었다', '예를 들었다' 세 문장에 나타난 동사는 모두 '들다'로 그 어휘는 같지만, 각각 의미하는 것은 세 문장에서 모두 다르다. 앞의 세 문장들에 사용된 '들다'의 의미는 순서적으로 '떡다의 높임', '놓인 물건을 잡아 위로 올리다', '어떤 사실이나 예를 내보이거나 지적하다' 이고, 동사 '들다'의 의미는 이외에도 '손에 가지다', '몸의 한부분을 쳐들어 올리다' 등의 의미가 있다. 그런데 이와 같이 의미 중의성을 지닌 동사들을 의미 애매성없이 사용할 수 있는 이유는 이런 동사들이 문장에서 사용될 때는 다른 단어들에 의해 그 의미를 제한받기 때문이다. 앞에서 예를 든 문장에서 보듯이, '점심', '가방', '예' 등이 바로 '들다'의 의미를 제한하는 단어들이다.

이와 같이 의미 중의성을 지닌 동사와 함께 나타나는 단어들은 그 동사의 의미를 분별하는데 매우 결정적인 역할을 한다. 본 논문에서는 대량의 말뭉치로부터 의미 중의성을 지닌 동사와 함께 나타나는 단어들의 분포를 이용하여 동사의 의미를 분별하고자 한다. 본 논문에서의 동사 의미 분별이란 의미 중의성을 지닌 동사들이 말뭉치내에서 여러 의미로 사용될 때, 이 동사들을 같은 의미로 쓰인 동사들끼리 분류하는 것을 의미한다.

### 2. 분포정보

말뭉치에 나타나는 단어들의 분포 정보는 통계적인 자연 언어 처리에 있어서 매우 유용한 정보이다. 말뭉치내의 단어 분포란 단순히 인접해서 나타나는 단어들에 대해서도 적용이 가능하며 특정 구문 관계로 나타나는 단어들에 대해서도 적용이 가능하다.

단어 분포를 이용한 지금까지의 연구에 대해서 간략하게 살펴보자. [Brown 1992]에서는 말뭉치내에서 좌우 50 단어 내에 인접해서 나타나는 단어들간의 Mutual Information을 이용하여 단어를 분류하였으며, [Yarowsky 1992]는 [Brown 1992]과 같이 말뭉치내에서 좌우 50 단어 내에 인접해서 나타나는 단어들을 이용하나, 단어 자체가 아니라 단어의 클래스를 이용하여 의미 중의성을 지닌 단어의 의미를 분별하였다. 그리고 [Schütze 1992]는 인접해서 나타나는 단어들만을 이용할 때 나타나는 자료 부족 현상을 보완하기 위해, 인접해서 나타나는 단어와 인접해서 나타나는 단어들의 분포 양상까지 이용하여 단어의 의미를 분별하는 실험을 하였다.

말뭉치에 나타나는 구문 관계를 이용한 연구들을 살펴보자. [Hindle 1990]은 말뭉치내에 나타나는 predicate-argument 구조에 따라 명사를 분류하였다. 먼저 구문 분석기를 이용하여 말뭉치내의 문장들에 대한 구문 구조를 얻고, 이로부터 predicate-argument 관계의 자료를 추출한 뒤, 이들간의 유사 정도에 따라 명사를 분류한다. [Pereira 1993] 역시 말뭉치로부터 서술어-직접목적어 관계의 정보를 추출하고, 이에 정보 이론에서 사용하는 상대 엔트로피를 적용하여 단어를 분류하였다. 즉

\*본 연구는 과학재단의 목적 기초 과제 "한국어 이해에 나타나는 중의성 문제 처리 모델에 관한 연구"의 부분 지원을 받은 것입니다.

비슷한 동사 분포를 갖는 명사들을 같은 부류로 분류하였다. [Grefenstette 1993]는 직접목적어 뿐만 아니라 주어, 간접 목적어 관계를 모두 이용하여 유사 명사를 추출하는 실험을 하였다. 또한 [Hatzivassiloglou 1993]는 말뭉치에 나타나는 형용사 명사, 형용사-형용사 분포 정보를 이용하여, 의미에 근거한 형용사 분류 실험을 하였다.

본 연구에서는 품사가 태깅된 말뭉치로부터 목적어-서술어 관계에 있는 명사-동사의 분포 정보를 이용한다. 이 분포 정보는 명사 분포와 동사 분포, 2가지로 나타낼 수 있다. 명사 분포란 말뭉치내에서 같은 명사를 목적으로 취하는 동사들의 분포를 의미하며, 동사 분포란 말뭉치내에서 같은 동사의 목적으로 나타나는 명사들의 분포를 의미한다. 예를 들면 말뭉치내에서 동사 ‘들다’는 ‘점심, 가방, 예’를 목적으로 취하고, ‘점심’은 ‘들다, 먹다, 차리다’의, ‘가방’은 ‘들다, 열다, 뒤지다, 찾다’의, ‘예’는 ‘들다, 찾다, 제시하다’의 목적으로 나타난다고 할 때, 명사 ‘점심’에 대한 분포 정보는 (들다, 먹다, 차리다), ‘가방’에 대한 분포 정보는 (들다, 열다, 뒤지다, 찾다)가 되고, 동사 ‘들다’에 대한 분포 정보는 (점심, 가방, 예)가 된다.

### 3. 알고리즘

본 연구는 3 단계로 이루어진다. 첫번째 단계는 말뭉치로부터 언어적 정보를 추출하는 것으로, 품사가 태깅된 한국어 말뭉치로부터 목적어-서술어 관계에 있는 명사-동사 단어쌍을 추출한다. 이렇게 추출된 단어쌍으로부터 명사와 동사에 대한 분포를 만든다. 두번째 단계에서는 상대 엔트로피를 이용하여 주어진 명사 분포간의 유사도를 계산한다. 상대 엔트로피는 정보 이론에서 사용되는 것으로, 두 확률 분포의 차이 정도를 나타낸다. 세번째 단계에서는 목적으로 취하는 명사 집합의 유사 정도에 따라 동사 분포간 유사도 계산을 한다.

#### 3.1 1 단계: 목적어-서술어 관계의 명사-동사 단어쌍 추출

첫번째 단계에서는 말뭉치로부터 목적어-서술어 관계에 있는 명사-동사 단어쌍을 추출한다. 일반적으로 품사가 태깅된 한국어 말뭉치로부터 목적어-서술어 관계에 있는 명사-동사 단어쌍을 추출하기 위해서는 구문 분석이 선행되어야 한다. 그러나 첨가어라는 한국어의 특성을 살펴 조사와 어미를 이용하면, 구문 분석 없이 품사만 태깅되어 있는 말뭉치로부터 목적어-서술어 관계에 있는 단어쌍을 추출할 수 있다. 예를 들면, 조사 ‘을/를’은 일반적으로 목적격 관계를 나타내고, ‘을/를’과 같이 나타나는 명사는 어떤 동사의 목적어임을 알 수 있다. 따라서 다음과 같이 목적어-서술어 관계의 명사-동사 단어쌍을 추출하는 원칙을 세울 수 있다.

원칙 1> 말뭉치내에서 품사가 조사로 태깅되고 그 어휘가

명사	동사				
	쓰다	읽다	벗다	짚다	찾다
편지	0.3	0.5	0	0.1	0.1
안경	0.4	0	0.4	0	0.2

표 1: 명사 ‘편지’와 ‘안경’에 대한 분포

‘을’, ‘를’인 경우, 이는 목적격 관계를 나타낸다.

원칙 2> 목적어는 가장 가까이에 나타나는 동사에 붙는다.

이 원칙들은 오류를 포함하고 있다. 즉, <원칙 2>에 의하면 “짐을 달리는 기차에서 받는다”와 같은 문장으로부터 (짐-달리다)와 같은 잘못된 단어쌍을 추출하게 된다. 이러한 현상은 자동사(달리다)가 명사(기차)를 수식하는 문장에서 나타나는데, 본 실험에서 이용하는 분포 정보는 이와 같은 오류 정보를 포함한다. 그러나 실제 위의 원칙을 적용하여 목적어-서술어 관계의 명사-동사 단어쌍을 추출한 결과, 이와같이 잘못 추출된 단어쌍은 바르게 추출되는 단어쌍에 비해 그 수가 매우 적었다.

이렇게 추출된 명사-동사 단어쌍으로부터 명사 분포와 동사 분포를 만든다. 명사 분포는 각 동사가 그 명사를 목적으로 취하는 조건 확률로 나타난다. 표 1은 명사 ‘편지’와 ‘안경’에 대한 분포의 예이다.

동사 분포는 각 명사가 그 동사의 목적어로 나타나는가, 아닌가하는 정보만을 포함한다. 즉, 각 명사가 그 동사의 목적어로 나타난 적이 있으면 1의 값을 갖고 그렇지 않을 경우는 0의 값을 갖는다.

#### 3.2 2 단계: 명사 분포간 유사도 계산

두 명사 분포간의 유사도 계산은 상대 엔트로피를 이용한다. 상대 엔트로피는 정보 이론에서 두 확률 분포간의 유사도를 측정하는 할 때 사용한다. 두 확률 분포  $p, q$ 가 주어질 때 상대 엔트로피는 식 1과 같이 정의된다.

$$D(p \parallel q) = - \sum_x p(x) \log \frac{p(x)}{q(x)} \quad (1)$$

식 1은 식 2과 같이 풀어 쓸 수 있다.

$$D(p \parallel q) = - \sum_x p(x) [\log \frac{1}{q(x)} - \log \frac{1}{p(x)}] \quad (2)$$

식 2에서 알 수 있듯이, 상대 엔트로피는 확률 분포의 각 시점에서 확률 분포  $p, q$ 의 차이값의 평균값이다.

앞에서 명사 분포는 각 동사가 그 명사를 목적으로 취하는 조건 확률로 나타난다고 정의하였다. 이 정의에 의해 명사  $n$ 에 대한 분포,  $d(n)$ 을 식으로 표현하면 식 3, 4와 같다.

$$d(n) = \langle p(v_1|n), p(v_2|n), \dots, p(v_k|n) \rangle \quad (3)$$

$$p(v^i|n) = \frac{\text{freq}(v^i, n)}{\sum_{j=1}^k \text{freq}(v^j, n)} \quad (4)$$

	Answer should be Yes	Answer should be No
The system says Yes	a	b
The system says No	c	d

표 2: 결과 분석을 위한 분할표

식 4에서  $freq(v, n)$ 은 말뭉치에서 목적어-서술어 관계로 나타나는 빈도수를 의미한다.

명사 분포  $d(n_1), d(n_2)$ 가 주어질 때, 상대 엔트로피에 의한 두 분포간의 유사도는 다음과 같다.

$$D_1(d(n_1) \parallel d(n_2)) = - \sum_{i=1}^k p(v_i | n_1) \log \frac{p(v_i | n_1)}{p(v_i | n_2)} \quad (5)$$

### 3.3 3 단계: 동사 분포간 유사도 계산

3 단계에서는 명사 분포를 이루는 동사들 중 2 단계의 상대 엔트로피에 의한 유사도 계산에 참여하지 못한 것들에 대한 고려를 한다. 실질적으로 많은 경우, 자료 부족 현상에 의해 명사 분포들은 많은 수의 동사들을 공유하지는 못한다. 이것은 명사 분포만을 이용한 의미 분별 실험 결과로도 알 수 있다. 이와 같은 자료 부족 현상을 보완하기 위해 동사 분포 정보를 사용한다.

3 단계는 2 단계에서  $p(v_i | n_1) = 0$  혹은  $p(v_i | n_2) = 0$ 인 경우에 적용되며 그 과정은 다음과 같다.

1. 명사 분포  $d(n_1)$ 에서  $p(v_i | n_1) = 0$ 인 모든 동사  $v_i$ 에 대해, 각 동사의 동사 분포들에 대해 OR 연산을 수행하여 새로운 분포  $dn_1$ 을 만든다.
2. 명사 분포  $d(n_2)$ 에서  $p(v_i | n_2) = 0$ 인 모든 동사  $v_i$ 에 대해, 각 동사의 동사 분포들에 대해 OR 연산을 수행하여 새로운 분포  $dn_2$ 을 만든다.
3. 새로이 만들어진 분포  $dn_1, dn_2$ 에 대해 다음과 같이 내적의 곱을 취한다.

$$D_2(dn_1, dn_2) = dn_1 \cdot dn_2$$

4. 너무 많은 명사를 목적어로 갖는 동사들은 1,2 과정에서 제외된다.

따라서 최종적인 유사도는 2 단계에서 구한  $D_1$ 과 3 단계에서 구한  $D_2$ 에 의해 계산된다.

$$D = \alpha D_1(d(n_1) \parallel d(n_2)) + \beta D_2(dn_1, dn_2) \quad (6)$$

상수  $\alpha\beta$ 는 실험 상수이다.

## 4. 실험 및 평가요소

본 논문에서 동사 의미 분별 실험이란 말뭉치에서 사용된 의미 중의성을 지닌 동사들 그 쓰인 의미에 따라 분류하는 것이다. 즉, 말뭉치내에 ‘편지를 쓰다’, ‘글을 쓰다’, ‘안경을 쓰다’란 세 문장이 있을 때, 앞의 두 문장에서의 ‘쓰다’는 같은 의미의 ‘쓰다’이고, 마지막 문장의 ‘쓰다’는 이것과 다른 의미이다라고 분류하는 것이다.

한국어 동사 10개에 대해서 의미 분별 실험을 하였다. 표 5는 실험 대상이 되었던 동사의 의미들을 정리한 것이다.

실험에 사용한 말뭉치는 품사가 태깅된 것으로, 교과서, 소설, 기사 등 다양한 형태의 글로 구성되어 있다. 먼저, 말뭉치로부터 실험 대상이 되는 의미 중의성을 지닌 10개의 동사와 목적어 관계에 있는 명사 478개를 추출하였다. 말뭉치내에서 이 명사들을 대상으로 하여 목적어-서술어 관계에 있는 명사-동사 단어쌍을 추출하였다.

실험은 명사 분포만을 이용한 것(실험 I)과 명사 분포와 동사 분포를 모두 사용하는 것(실험 II)으로 나누어 하였다.

의미 분별 실험의 결과는 사전에 나온 동사의 의미 용례에 의거하여 평가하였다. 즉, 사전에 나와 있는 의미 용례에 맞게 분별한 것에 한해서 바르게 의미 분별된 것으로 결정하였다. 실험 결과 분석은 정보 검색 분야에서 사용하는 표 2와 같은 분할표(contingency table)를 이용한다.

표 2의 분할표에서 ‘Yes’란 같은 의미 부류인 경우를 의미하고, ‘No’란 서로 다른 의미 부류를 의미한다. 이 분할표에 따라 우리는 다음과 같이 recall, precision, fallout을 정의할 수 있다.

$$Recall = \frac{a}{a+c} \cdot 100\%$$

$$Precision = \frac{a}{a+b} \cdot 100\%$$

$$Fallout = \frac{b}{b+d} \cdot 100\%$$

이들을 각각 설명하면, 먼저 recall은 ‘Yes’이어야 하는 것 중 시스템이 ‘Yes’라고 대답한 것의 비율이다. precision이란 시스템이 ‘Yes’라고 한 것 중 진짜 ‘Yes’인 것의 비율이다. fallout이란 ‘No’이어야 하는데 시스템이 ‘Yes’라고 한 것의 비율이다.

일반적으로 recall과 precision을 이용하여 시스템을 비교할 때, 어느 한 값이 극단적으로 좋은 시스템보다는 두 값이 비교적 비슷하게 좋은 시스템을 선호하는 경향이 있다. recall과 precision이 어느 정도 비슷한가를 분석하기 위해 F-measure를 계산하였다. F-measure는 recall과 precision의 값이 비슷할수록 그 값이 크다. 7에서  $\beta$ 는 precision에 상대적인 recall의 비율을 의미하는데, 본 실험에서는  $\beta = 1$ 로 하여 recall과 precision을 같은 정도로 고려하였다.

Verb	Recall	Precision	Fallout	F-measure ( $\beta = 1$ )
치다	5.9%	100.0%	0.0%	11.1%
들다	8.2%	88.6%	0.6%	15.0%
쓰다	15.2%	50.0%	9.2%	23.3%
따르다	24.1%	100.0%	0.0%	38.8%
말리다	16.7%	100.0%	0.0%	28.6%
대다	16.7%	66.7%	3.0%	26.7%
감다	0.0%	0.0%	0.0%	0.0%
피우다	0.0%	0.0%	0.0%	0.0%
타다	0.0%	0.0%	0.0%	0.0%
풀다	10.0%	100.0%	0.0%	18.2%
평균	9.7%	60.5%	1.2%	16.2%

표 3: 명사 분포만을 사용한 실험(실험 I) 결과

Verb	Recall	Precision	Fallout	F-measure ( $\beta = 1$ )
치다	55.3%	79.7%	5.6%	65.3%
들다	64.4%	82.1%	8.3%	72.2%
쓰다	57.6%	51.4%	20.7%	54.3%
따르다	43.4%	85.7%	2.5%	57.6%
말리다	75.0%	66.7%	42.9%	46.2%
대다	66.7%	70.6%	10.0%	68.6%
감다	72.7%	80.0%	6.7%	76.2%
피우다	100.0%	80.0%	16.7%	85.7%
타다	42.9%	66.7%	24.5%	52.2%
풀다	70.0%	70.0%	13.0%	70.0%
평균	64.8%	72.8%	15.1%	64.8%

표 4: 명사 분포와 동사 분포를 사용한 실험(실험 II) 결과

$$F = \frac{(\beta^2 + 1) * precision * recall}{\beta^2 * precision + recall} \quad (7)$$

## 5. 결과 분석

표 3, 4는 앞에서 정의한 recall, precision, fallout, F-measure, 이 네가지 요소에 의해 실험 결과를 분석한 것이다. 표 3은 분포 정보 중 명사 분포 정보만을 사용하여 실험한 것의 결과이고, 표 4는 명사 분포 정보와 동사 분포 정보 모두를 사용하여 실험한 것의 결과이다. 동사 분포 정보를 함께 사용하여 실험한 결과, 평균 64.8%의 recall, 평균 72.8%의 precision을 얻었고, 실험 I에 비해 실험 II의 결과가 평균 recall과 precision이 더 높게 나타났으며, F-measure 역시 높게 나타났다.

그림 1, 2는 실험 I, II의 결과를 recall과 precision 별로 비교한 그래프이다. 또한 그림 3은 recall과 precision을 함께 비교한 것이다. 실험 I의 결과는 recall은 매우 낮고 precision은 비교적 높는데 반해, 실험 II의 결과는 recall은 실험 I에 비해 상대적으로 높고, precision도 비교적 높게 나타난 것을 그림 1, 2, 3으로 알 수 있다.

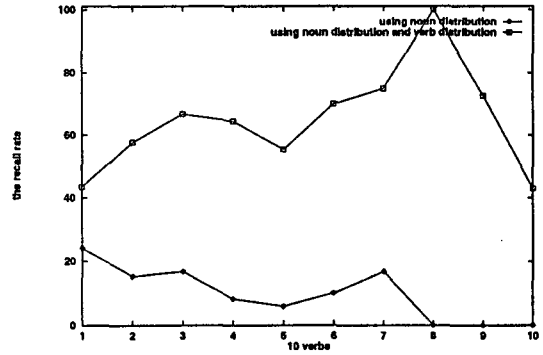


그림 1: 실험 I과 실험 II 결과의 recall 비교

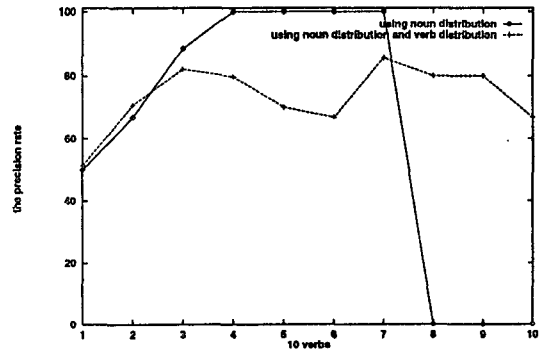


그림 2: 실험 I과 실험 II 결과의 precision 비교

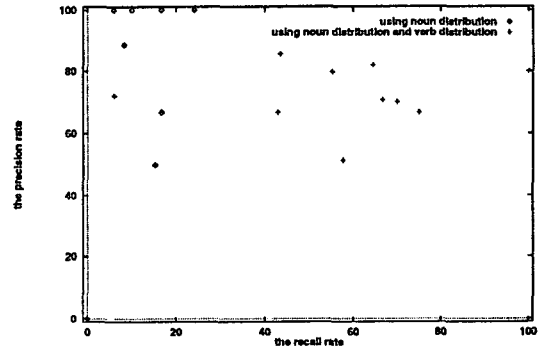


그림 3: 실험 I, II 결과의 recall과 precision 비교

## 6. 결론

본 논문은 말뭉치로부터의 분포 정보를 이용하여 의미 중의성을 지닌 한국어 동사를 사용 의미에 따라 분류하였다. 여기서 분포 정보란 말뭉치내에서 목적어-서술어 관계로 나타나는 명사-동사의 분포를 의미한다. 본 논문에서는 명사 분포 정보와 동사 분포 정보, 두가지를 이용하는데, 명사 정보만으로 실험하였을 때 나타나는 자료 부족 현상을 동사 분포 정보를 사용하여 어느 정도 해소하였다.

실험은 명사 분포 정보만을 이용한 것과 명사 분포 정보와 동사 분포 정보를 함께 이용한 것, 두가지를 수행하였다. 동사 분포 정보를 함께 이용한 실험 결과, 평균 64.8%의 recall, 평균 72.8%의 precision을 얻었다.

이와 같은 접근 방법은 의미 분별뿐만 아니라 명사 분류에도 마찬가지로 적용될 수 있을 것이다.

## 참고문헌

- [Brown 1991] Peter F. Brown, Stephen A. Della Pietra and Vincent J. Della Pietra. 1991. Word Sense Disambiguation using Statistical Methods. In *Proceedings of the 29th Meeting of the Association for Computational Linguistics*.
- [Brown 1992] Brown, P. F., Pietra, V. J. D., DeSouza, P. V., Lai, J. C., and Mercer, R.L. 1992. Class-based n-gram models of natural language. In *Computational Linguistics*.
- [Church 1989] K. Church and P. Hanks. 1989. Word association norms, mutual information, and lexicography. In *Proceedings of the 27th Meeting of the Association for Computational Linguistics*. Vancouver, B.C.
- [Gale 1992] Gale, William, K. Church and David Yarowsky. 1992. A Method for Disambiguating Word Senses in a Large Corpus. *Computers and Humanities*.
- [Grefenstette 1993] G. Grefenstette. 1993. SEX-TANT: extracting semantics from raw text: implementation details. In *Heuristics: The Journal of Knowledge Engineering*.
- [Hatzivassiloglou 1993] Vasileios Hatzivassiloglou, and Kathleen R. McKeown. 1993. Towards The Automatic Identification Of Adjectival Scales : Clustering Adjectives According To Meaning. In *Proceedings of the 31st Meeting of the Association for Computational Linguistics*. Columbus.

동사	의미
감다	위아래 눈시울을 맞붙이다 머리, 몸을 물로 씻다 실, 끈 따위를 무엇에 빙 두르다 다리틀 걸다
대다	무엇을 어디에 닿게 하다 돈이나 물건을 마련하여 주다 정보를 말하다 정해진 시간에 가서 닿다
들다	동인 물건을 잡아 위로 올리다 어떤 사실이나 예를 내보이거나 지적하다 손에 가지다 몸의 한부분을 쳐들어 올리다 '먹다'의 높임말
따르다	남의 뒤를 쫓다 관례나 법규 따위에 복종하다 나란히 같이 가다. 수반하다 그릇등을 기울여 액체를 흐리게 하다
말리다	젖은 것을 마르게 하다 남이 하는 행동을 하지 못하게 하다
쓰다	머리에 얹거나 얼굴에 걸다 사용하다, 이용하다 글을 짓다 남을 부리어 일을 시키다 소비하다, 낭비하다 무엇에 마음이나 힘을 들이다 억지나 고집을 심하게 부리다
치다	다른 물건을 때리다 소리가 나게 두드리다 악기를 연주하다 공격하거나 타격을 주다 적은 분량의 액체를 따르거나 가루 따위를 뿌려서 넣다 붓 등으로 점이나 선을 찍거나 긋다 무엇을 치르거나 격다 퍼서 벌이거나 늘어 드리다 소리를 힘있게 내다 어떤 행동을 힘주어 하거나 저지르다 가축을 기르다 동물이 새끼를 낳거나 까다
타다	탈것이나 짐승의 등 따위에 몸을 실다 위험하거나 어려운 곳을 겨우 오르거나 지나가다 어떤 자연 현상이나 기회, 기세 따위를 이용하다 많은 액체에 적은 액체나 고체 등을 섞다 재산, 월급, 상, 배급 따위를 받다 악기를 다루어 소리를 내다 시절, 기후의 영향을 쉬 받다
풀다	묶이거나 얽히거나 또는 합쳐진 것을 끄르거나 풀어지게 하다 어떤 감정이나 상태를 가라앉히거나 사라지게하다 액체에 다른 물질을 타서 섞다. 모르거나 복잡한 문제나 일을 알아내거나 해결하다 금지제한된 것을 터놓다 사람을 동원하다
피우다	담배를 빨아 연기를 입코로 내보내다 어떤 행동을 부리다. 사동의 의미로 피게 하다 냄새면 따위를 퍼뜨리거나 일으키다

표 5: 실험 대상 동사들의 의미

- [Hindle 1990] Donald Hindle. 1990. Noun Classification From Predicate-Argument Structures. In *Proceedings of the 28st Meeting of the Association for Computational Linguistics*.
- [Pereira 1992] Fernando Pereira and Naftali Tishby. 1992. Distributional similarity, phase transitions and hierarchical clustering. In *Working Notes, Fall Symposium Series. AAAI*.
- [Pereira 1993] Fernando Pereira, Naftali Tishby and Lillian Lee. 1993. Distributional Clustering Of English Words. In *Proceedings of the 31st Meeting of the Association for Computational Linguistics*. Columbus.
- [Resnik 1993] Philip Stuart Resnik. 1993. *Selection and Information: A Class Based Approach to Lexical Relationships*. Ph.D. Dissertation. University of Pennsylvania.
- [Schütze 1992] Schütze, H. 1992. Word Sense disambiguation with sublexical representations. In *Workshop Notes, Statistically-Based NLP Techniques. AAAI*.
- [Yarowsky 1992] David Yarowsky. 1992. Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora. In *Proceedings of the 15th International Conference on Computational Linguistics*.