

세종 원시 말뭉치에서 살펴본 ‘ 상황’ 의 의미 운율

임인빈
고려대학교

Abstract This paper examines the semantic prosody of Korean word ‘sanghwang’ based on Sejong Korean Raw Corpus. In dictionary, the meaning of ‘sanghwang’ is ‘the state of things [affairs]’, and the neutral semantic prosody is expected. Moreover, there are three example sentences; the first has the positive semantic prosody, the second has the neutral semantic prosody, and the last example has the negative semantic prosody. The number of concordances of ‘sanghwang’ in Sejong Korean Corpus is 3,857, and the semantic prosody of each concordance is examined. As a result, in the concordances of subcorpus ‘News(63%), Spoken, Human, Science, Society’, words of the negative semantic prosody is co-occured with ‘sanghwang’ frequently, and the concordances of ‘Educate(36%)’, are co-occured less than the former. The sub-corpora which contain writer’s opinion relatively stronger has more concordances with the negative semantic prosody, and the sub-corpora which contains writer’s opinion relatively small has less concordances with the negative semantic prosody.

1. 들어가기

의미 운율(Semantic prosody)¹을 모어 화자의 직관만으로 확인한다는 것은 불가능한 일이다. 모어 화자의 직관에는 의미가 긍정적이거나 부정적이지 않은 중립적인 단어의 경우, 긍정적 의미 운율과 부정적 의미 운율, 그리고 중립적 의미 운율이 모두 나타날 수 있다는 가정은 할 수 있지만 어느 정도의 비율로 나타나며, 이것이 어떤 단어와 함께 쓰여서 의미 운율을 나타내게 되는지에 대한 부분은 살펴볼 수 없다.

이 글에서는 의미 운율에 대한 연구 중 하나로 한국어에서 중립적인 의미를 지니는 단어 중에서 부정적 의미 운율과 긍정적 의미 운율, 그리고 중립적 의미 운율을 모두 지닐 수 있는 단어 중에서 ‘ 상황’ 이라는 단어를 선택하여 살펴봄으로써, 중립적 의미 운율이 예상되는 어떤 단어의 의미 운율이, 말뭉치에서 살펴볼 경우에, 특정한 의미 운율을 지닐 수 있음을 보이고자 한다.

국어사전²에서 상황이라는 단어는 ‘ 일이 되어 가는 형편이나 모양’ 이라는 의미를 가지고 있으며, 단어의 의미와 예문만으로는 중립적인 의미 운율을 예상할 수 있다. 또한 사전의 예문도 ‘ 상황이 유리하다’ (긍정적 의미 운율), ‘ 상황이 나빠지다’ (부정적 의미 운율), ‘ 상황에 따라 다르다’ (중립적 의미 운율)의 세 종류의 의미 운율 예문이 포함되어 있다. 사전만을 참고했을 때 ‘ 상황’ 은 중립적 의미를 지니고 있으며, 세 의미 운율을 지닐 수 있음을 알 수 있지만, 어느 정도의 비율을 보일지에 대해서는 알 수 없다. 그래서 세종 말뭉치에서 실제 의미 운율을 살펴보려 한다.

실제 데이터에서는 세 의미 운율이 그렇지 않을 것이라는 것을 직관적으로 짐작할 수 있다. 한국어 화자는 상황이라는 단어를 긍정적인 일 보다는 부정적인 일을 묘사할

¹ 의미 운율의 정의는 Bill Louw(1993)과 Tony Berber Sardinha(2000)를 참조.

² 엡센스 국어사전(샤프 전자사전), 민중서림.

때 자주 쓴다. 실제 데이터에서도 그런지 세종 원시 말뭉치를 통해 살펴보자. 특히 각 말뭉치의 특징에 따라 사용되는 양상이 다를 것이라는 가설³ 아래 각 하위 폴더 별로 나누어 조사하고 각 말뭉치에서 나타나는 양상을 조사해 비교해 보기로 했다.

2. 연구방법

의미 운율을 알아보기 위해 ‘상황+조사’의 형태로 된 용례를 찾는 것이 가장 먼저 해야 할 작업이므로, 세종 원시 말뭉치에서 글잡이 용례 검색 프로그램을 사용하여 ‘상황*’으로 용례 검색을 실시하였다. 그 결과 전체 말뭉치에서 3,857 건이 검색되었고, 각 하위 폴더 별로는 Spoken(687), Art(322), Docu(292), Educate(323), General(79), Human(609), Imaginary(268), Life(88), News(1132), Science(153), Society(591) 건이 각각 검색되었다.

이 용례 전체의 의미 운율을 모두 살펴보기에는 시간이 너무 부족하므로, 자료의 편집을 위해 글잡이 검색 프로그램에서 제공하는 메뉴를 통해 용례를 Text 파일로 저장하였다. 이 자료를 한글 2007에서 매크로 기능을 이용하여 각 하위폴더 말뭉치당 100건 안팎의 용례로 무작위적으로 줄인 후 각각의 용례의 의미 운율을 살폈다. 이 과정에서 용례를 줄이는 비율은 각 말뭉치에 따라 조금씩 달라졌다.⁴ 가장 많은 용례를 보인 News의 경우와 가장 적은 용례를 보인 General의 경우, 검색된 용례의 수에 많은 차이를 보였기 때문이다. 살펴본 용례를 100건 안팎으로 줄인 이유는 결과를 비교하기 쉽게 하기 위해서이다.

세종 말뭉치 전체에서 ‘상황’의 용례를 보기 위해 3,857건의 검색 결과를 매크로 기능을 사용하여 1/20로 줄여서 살펴보았다.⁵ 매크로를 이용하여 한 용례를 남기고 다음의 19개의 용례를 삭제하는 과정을 반복하도록 컴퓨터의 매크로 기능을 사용하였다. 또한 나머지 하위 폴더 별로 살펴보면 Spoken은 1/7, Art는 1/3, Docu는 1/3, Educate는 1/3로 줄였으며, General은 검색된 전체 용례가 79건이므로 줄이지 않고 전체 용례를 모두 살펴보았다. Human은 1/5, Imaginary는 1/3로 줄였고, Life도 General처럼 검색된 용례가 88건이므로 전체를 살펴보았다. 가장 많은 용례가 검색된 News는 1/12로 줄였으며, Science는 검색된 용례가 153건으로 줄이기에는 너무 적은 숫자이므로 전체를 살펴보았다. Society도 1/5로 줄인 후 조건을 살펴보았다.⁶

³ Hans Lindquist and magnus Levin(2000)에서는 연어를 중심으로 살펴보기는 했지만, 각 말뭉치마다 특징이 있는 것으로 나타났다. 그 특징은 고유한 것이므로 각각의 말뭉치의 특징을 사과와 오렌지처럼 다른 것으로 볼 수 있다.

⁴ 이 연구에서 말뭉치를 무작위적으로 줄이는 과정에서 각 말뭉치 별로 줄여든 비율이 다른 것은 이 연구에 아무런 문제가 되지 않으므로 이에 대한 특별한 조치를 취하지는 않았고 대신 전체 용례를 일정 비율로 줄여서 비교할 수 있도록 전체 용례를 1/20로 줄인 결과를 살펴보았다.

⁵ 강범모 (2003)에 용례가 많은 말뭉치를 매크로 기능을 사용하여 자동적으로 줄이는 방법에 대해 자세히 설명되어 있다.

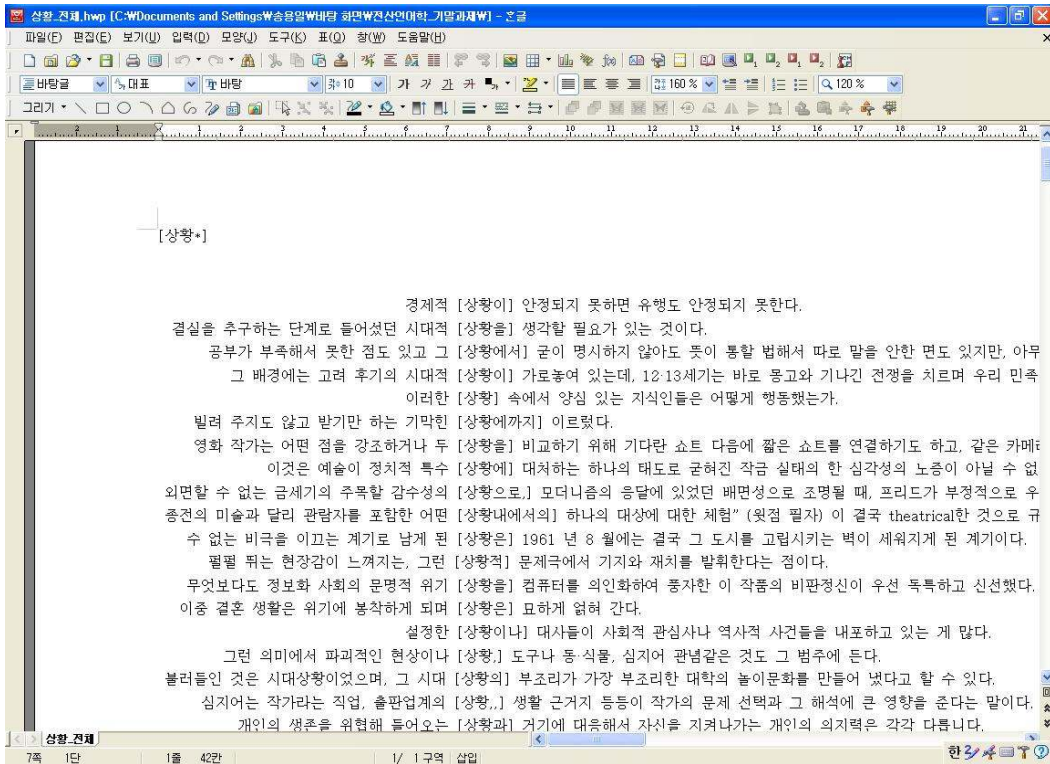
⁶ 각 말뭉치 별로 말뭉치의 크기와 용례의 차이가 있으므로 피할 수 없는 과정이며, 각 말뭉치 별로 비교할 수 있는 토대가 된다.

임인빈

구분	구어	Art	Docu	Edu	Gene	Hum	Imag	Life	News	Scien	Socie	전체
검색된 용례	687	322	292	323	79	609	268	88	1132	153	591	3857
살펴본 용례	91	105	98	107	78	131	88	85	99	153	108	195
줄인 비율	1/7 (14%)	1/3 (33%)	1/3 (33%)	1/3 (33%)	전체	1/5 (20%)	1/3 (33%)	전체	1/12 (8%)	전체	1/5 (20%)	1/20 (5%)

<표 1> 검색된 전체 용례와 살펴본 용례의 비율

각 말뭉치마다 100 개 정도의 용례를 살펴보고 서로 비교하기 위해 용례를 줄인 화면은 다음의 <그림 1>에서 볼 수 있다.



<그림 1> 매크로를 이용하여 줄인 용례의 캡처 화면

‘ 상황+ 명사’ 를 위시한 여러 부적절한 용례가 검색되어, 이런 경우 의미 운용의 어느 부분에도 속하지 않은 것으로 보고 이 연구의 조사 범위가 아니므로 제하고 나머지 부분만 살펴보았다.

의미 운용은 기준 단어의 좌측 4 단어와 우측 4 단어에서 부정적인 의미를 갖는 단어들이 있는지의 여부를 살펴보는 것으로 측정했으며⁷, 상황에 따라 5 단어나

⁷ John Sinclair(1996)의 연구 결과를 토대로 기준 단어의 좌우 4 단어에 필요한 정보의 대부분이 담겨 있을 것임을 가정하여 앞 뒤 4 단어를 살펴보았고, 몇 가지의 예외를 제외하고 대부분 경우에 앞 뒤 4 단어 안에 정보가 담겨 있었다.

6 단어를 살펴본 예도 있다⁸. 이 부분에서 조사자의 직관이 개입되었지만, 기본적으로는 데이터에서 나타나는 의미 운율과 부정적인 의미 운율을 보이는 단어의 종류와 수를 파악하는데 집중하였다.

이 연구에서 살펴보고자 하는 부분은 ‘상황’이라는 단어의 의미 운율을 살펴보는 것이며, 또한 부정적인 의미가 강할 것이라는 가정을 하였으므로 부정적인 의미 운율을 보일 부정어와 부정적인 의미를 지닌 단어의 수를 세었다. 또한 부정적인 의미 운율이 중립적인 의미 운율과 긍정적인 의미 운율에 비해 얼마나 자주 나타나는지를 알아보기 위해 부정적인 의미 운율이 나타나는 경우와 다른 두 종류의 의미 운율이 나타나는 경우의 수를 나누어서 세어보았다. 각 말뭉치 별 비교를 위해 전체 용례 중에서 부정적 의미 운율이 나타나는 비율을 측정하였다.

3. 연구결과 및 분석

위와 같은 방법으로 세종 말뭉치의 자료를 살펴본 결과를 살펴보자. 부정적 의미 운율을 나타내는 단어의 종류는 많았지만, 이를 모두 제시하는 것도, 그 수를 파악하기도 쉽지 않았다. 그러므로 빈도가 비교적 높은 어휘를 뽑아서 그 빈도 수를 세었다.

먼저 첫 3 개의 말뭉치에서 사용되는 부정적 의미 운율을 나타내는 단어를 모두 찾아서 빈도를 파악한 후 2 회 이상 나타나는 어휘 29 개를 선택하였다. 부정적 의미 운율을 나타내는 어휘를 앞에서 찾은 29 개의 어휘와 기타로 나누어 모든 말뭉치를 살펴보았다. 그 결과 29 개의 어휘 중 전체 말뭉치 용례에서 3 회 미만의 빈도를 나타내는 단어 12 개를 제외하였다. 이런 방법으로 부정적 의미 운율을 나타내는 단어들 중 비교적 빈도가 높은 어휘들은 다음의 17 개 단어였다.⁹

- (1) 극한, 급박한, 긴박한, 대처, 못하는, 문제, 아닌, 악화된, 앓은, 어려운, 없는, 위기, 절망적인, 절박한, 처한, 최악, 한계

위의 예문 (1)에서 가장 빈도가 자주 나타나는 단어는 ‘없는, 어려운, 앓은, 못하는, 아닌’의 순서로 나타났다. 수식어인 ‘어려운’을 제외한 단어들은 전형적인 부정어이므로 부정의 의미 운율을 나타내는 중요한 기준이 되는 단어들이다. 이러한 부정어들과 연어를 이루는 ‘상황’은 부정적 의미 운율을 갖는 것이 당연할 것이다.

또한 ‘어려운’과 함께 빈도가 높은 편에 속하는 ‘극한’, ‘급박한’, ‘긴박한’, ‘악화된’, ‘절망적인’, ‘절박한’, ‘처한’ 등의 형용사들과 ‘문제’, ‘위기’, ‘최악’, ‘한계’ 등의 명사들은 어휘의 의미 자체에 부정적 의미를 지니고 있으므로 부정적 의미 운율을 갖게 되는데 중요한 역할을 하게 된다. 이 또한 부정적 의미를 지닌 수식어들과 자주 공기하는 ‘상황’이 부정적 의미 운율을 갖게 되는 이유 중의 하나일 것이다.

한 가지 눈여겨보아야 할 사항은 각 말뭉치에서 빈도가 높게 나왔던 ‘아닌’과 ‘없는’은 전체 용례를 1/20 로 줄인 용례 집단에서는 비교적 출현 빈도가 낮은 편이었고 반대로 각 말뭉치에서 상대적으로 빈도가 낮았던 ‘어려운’의 경우에는 전체 용례를 줄인 집단에서 비정상적으로 높은 빈도를 보였다. 이는 각 말뭉치를

⁸ 이렇게 살펴본 예는 극히 일부 용례만 있으며, 용례와 의미 운율을 결정하는 단어 사이에 삽입된 수식어구가 긴 경우가 이에 해당된다.

⁹ 이 과정에서 첫 3 개의 말뭉치에서 1 회 출현했거나 출현하지 않았지만 나머지 말뭉치에서 많이 출현한 어휘가 있을 수 있으나, 조사자의 눈에 띄는 어휘는 발견되지 않았다.

임 인 빈

무작위적으로 줄이는 과정에서 우연히 일어난 것으로 볼 수 있다.¹⁰ 이러한 사실에도 불구하고 ‘상황’이 부정적 의미 운율을 지니고 있다는 사실에는 변함이 없고, 부정적 의미 운율을 갖게 하는 단어가 변한 것일 뿐이다. 부정적 의미 운율을 지니게 하는 단어를 찾는 것보다는 ‘상황’이 부정적 의미 운율을 지니고 있음을 밝히는 것이 이 연구의 첫 번째 목적이므로, 이 문제에 대해서는 더 이상 다루지 않기로 한다.

	출현 횟수	출현 비율
없는	75	5.6%
어려운	61	4.6%
않은	54	4.0%
못하는	45	3.4%
아닌	36	2.7%
문제	25	1.9%
악화된	21	1.6%
처한	20	1.5%
위기	17	1.3%
극한	12	0.9%
한계	10	0.7%
대처	9	0.7%
최악	7	0.5%
긴박한	6	0.4%
절망적인	6	0.4%
절박한	6	0.4%
급박한	5	0.4%
전체	1,338	

<표 2> 고빈도 부정적 의미 운율 단어

각 세부 하위 폴더 말뭉치에서 자주 공기하는 의미 운율 17 개의 단어를 살펴본 결과를 표로 제시하면 다음의 <표 3>과 같다.

¹⁰ 말뭉치의 용례를 줄이는 과정에서는 무작위적으로 한 것이므로 우연히 이러한 결과가 도출된 것이며, 연구 결과에는 큰 영향을 미치지 않았다.

세종 원시 말뭉치에서 살펴본 ‘상황’의 의미 운율

	구어	Art	Docu	Edu	Gene	Hum	Imag	Life	News	Scien	Socie	전체
극한		2				1	2	2	1	1		3
급박한	1	1	1						1			1
긴박한		1	1				1	1	1			1
대처	1			2		1		2			1	2
못하는	3	2	1	4	1	1		5	2	11	6	9
문제	1			4		3		1	3	9	2	2
아닌	4	4	4		3	4	2	2	5	4	2	2
악화된	2			1	1			1	4		1	11
않은	7	3	4	3	2	2	5	2	7	7	4	8
어려운	7	3		1	2	3	2	4	2	3	4	30
없는	6	3	7	1	7	10	7	6	9	10	5	4
위기	5	3			1	1		1	1	1		4
절망적인		1		1		1	1	1		1		
절박한		1		2		1	2					
처한		2	1	4	1	1	1	1	2	3		4
최악	1						1		5			
한계		1	2	2	2				1		2	
기타	9	14	18	14	14	41	20	11	18	31	32	63
부정	47	41	39	39	34	70	44	40	62	81	59	111
중립/ 긍정	44	64	59	68	44	61	44	45	37	72	49	84
총계	91	105	98	107	78	131	88	85	99	153	108	195

<표 3> 부정적 의미 운율을 보이는 어휘의 빈도수¹¹

<표 3>에서 볼 수 있듯이 Science 와 Society 의 말뭉치에서는 의미 운율을 지니게 하는 어휘가 상당히 다양해서 기타에 속하는 어휘가 많았으며, spoken 과

¹¹ 오른쪽의 ‘전체’ 칼럼은 각 말뭉치에서 나온 값의 합이 아니라 전체 용례를 1/20 로 줄인 195 개 용례를 살펴본 것이므로 결과는 각 말뭉치의 합과는 다르다.

life 의 경우에는 다양한 종류의 의미 운율을 나타내는 어휘들이 사용되었으며, 그 결과 기타에 속하는 어휘가 많지 않았다.

또한 News, Spoken, Human, Science, Society 의 하위 폴더의 용례는 부정/긍정적 의미 운율이 더 비교적 많이 나타났고, Imaginary 에서는 부정적 의미 운율이 반으로 나타났으며, Educate, Art, Docu, General, Life 의 경우에는 부정적 의미 운율이 50%보다 적게 나타났다.¹² 각 말뭉치 별 전체 말뭉치의 비율에 대한 부정적 의미 운율의 비율을 모두 측정하여 표로 제시하면 다음 <표 4>와 같다.

구분	구어	Art	Docu	Edu	Gene	Hum	Imag	Life	News	Scien	Socie	전체
부정	47	41	39	39	34	70	44	40	62	81	59	111
중립/ 긍정	44	64	59	68	44	61	44	45	37	72	49	84
총계	91	105	98	107	78	131	88	85	99	153	108	195
비율	52%	39%	40%	36%	44%	53%	50%	47%	63%	53%	55%	57%

<표 4> 부정적 의미 운율의 비율

<표 4>와 같이 전체 3857 개의 용례를 1/20 정도로 줄여서 살펴본 용례의 경우에는 부정적 의미 운율이 57% 정도 나타났다.¹³ 부정적 의미 운율이 가장 많이 나타난 말뭉치는 News 로 약 63% 정도로 비율이 높았으며, 가장 적게 나타난 말뭉치는 Educate 로 약 36%로 비율이 낮았다. 그러나 가장 낮은 Educate 의 경우에도 전체의 1/3 이 넘는 수치가 나타난 것으로 Educate 의 경우에도 ‘상황’ 이라는 단어는 부정적 의미 운율을 갖는 단어임을 알 수 있다.

부정적 의미 운율이 높은 말뭉치들은 News, Spoken, Human, Science, Society 등으로, 특징은 저자들의 개인적인 견해가 비교적 많이 들어가게 되는 말뭉치이며, 이러한 경우 부정적인 의미 운율이 많이 나타나는 것을 볼 수 있다. 반대로 부정적 의미 운율이 낮은 말뭉치들은 Educate, Art, Docu, General, Life 로 전자에 비해 객관적인 정보 전달의 기능이 높고 저자의 견해가 비교적 적게 들어가게 되는 말뭉치인 것을 알 수 있다.

4. 결론

한국어 화자의 경우 대부분 ‘상황’ 이라는 단어가 부정적 의미 운율을 가진 단어라는 것을 직관적으로 알 수 있을 것이다. 그러나 때로는 직관이 실제 데이터에서 잘못된 것으로 나타나는 경우도 많다. 코퍼스 언어학은 그러한 의미에서 화자의 직관을 확인하는 작업으로 의미 있게 사용될 수 있다.

¹² ‘상황’ 이 중립적 의미 운율을 지닌 단어라면 부정적 의미 운율과 긍정적 의미 운율, 그리고 중립적 의미 운율이 1/3 로 나와야 하는데 부정적 의미 운율이 가장 적은 말뭉치의 경우에도 전체의 40% 이상을 나타내는 결과를 통해 부정적 의미 운율이 강하게 나타남을 볼 수 있다.

¹³ 표에 나타난 데이터의 수치상으로 보면 전체 말뭉치에서 부정적 의미 운율의 비율은 절반 정도여야 하는데 훨씬 높게 나타난 이유는 부정적 의미 운율이 절반 넘게 나타난 말뭉치의 용례가 반대의 결과를 보이는 말뭉치의 용례보다 훨씬 많기 때문이다.

‘상황’의 경우는 직관과 일치하는 결과를 보이고 있음을 살펴보았다. 부정적인 의미 운율을 가졌다는 사실은 직관과 일치하지만, 화자의 직관은 거의 대부분 부정적 의미 운율로 사용될 것 같지만, 부정적인 의미 운율로 사용되는 용례의 빈도가 직관에 의한 판단보다는 긍정이거나 중립적인 의미 운율로 사용된 용례보다 훨씬 자주 사용됨을 알 수 있다. 직관적으로는 부정적 의미 운율이 거의 대부분일 것으로 생각했지만, 중립적 의미 운율과 긍정적 의미 운율도 직관에 비해 비교적 잦은 빈도로 나타나는 것을 볼 수 있다.

부정적인 의미 운율을 보여주는 가장 빈도가 높은 어휘는 ‘없는, 어려운, 얇은, 못하는, 아닌’으로 ‘어려운’을 제외한 나머지 어휘들은 모두 전형적인 부정어들이며, 나머지 어휘들도 의미가 부정적인 어휘이다. 이러한 부정어와 의미가 부정적인 어휘와 자주 공기하는 ‘상황’은 부정적 의미 운율을 갖는 단어일 수밖에 없다.

가장 부정적 의미 운율 어휘의 빈도가 높은 말뭉치는 News 이고, 가장 낮은 말뭉치는 Educate 이다. 전체적으로 부정적 의미 운율 어휘의 빈도가 높은 말뭉치는 저자들의 개인적인 견해가 비교적 많이 들어가게 되는 말뭉치라는 특징이 있었고, 빈도가 낮은 말뭉치는 비교적 객관적인 정보 전달의 기능이 높고 저자의 견해가 비교적 적게 들어가게 되는 말뭉치임을 확인하였다.

참고문헌

- 강범모. 2003. 언어, 컴퓨터, 말뭉치 언어학. 고려대학교 출판부.
- Lindquist, Hans and Levin, Magnus. 2000. ‘Apples and Oranges: On Comparing Data from Different Corpora’. in C. Mair and M. Hundt (eds.). *Corpus Linguistics and Linguistic Theory*. pp. 201-214.
- Louw, Bill. 1993. ‘Irony in the Text or Insincerity in the Writer? - The Diagnostic Potential of Semantic Prosodies’. in M. Baker, G. Francis and E. Tognini-Bonelli (eds.). *Text and Technology: in honour of John Sinclair*. pp. 157-176. www.benjamins.com
- Sardinha, Tony Berber. 2000. ‘Semantic Prosodies in English and Portuguese: A Contrastive Study’. *Cuadernos de Filología Inglesa*. 9. 1. pp. 93-110.
- Sinclair, John. 1996. ‘The Search for Units of Meaning’. in L. M. Barbaresi and J. Cinclair (eds.). *TEXTUS IX, 1, Special Volume: Lessico e Morfologia*. pp. 75-106.
- Tognini-Bonelli, Elena. ‘The corpus-driven approach’, in *Corpus Linguistics at Work*. www.benjamins.com. pp. 84-100.

임인빈
고려대 언어학과
성북구 안암동 5가
inbean75@korea.ac.kr