

## 코퍼스를 이용한 은유 표현의 추출과 어휘데이터베이스의 확장 연구\*

- 경제 기사를 대상으로 -

최운호\*\* · 강범모 · 차재은\*\*\*  
(고려대학교 · 고려대학교 · 경기대학교)

### <Abstract>

Choi Woonho · Cha Je-Eun · Kang Beom-mo. 2006. **The Extraction of Metaphorical Expressions from the Newspaper Corpus.** *Korean Semantics*, 20. This paper aims at developing methods of extracting figurative expressions from a domain-specific corpus and extending the multi-lingual lexical database entries for those expressions to be included. The nominal entries which can be used figuratively are selected from the annotated corpus, using the hierarchical semantic class of Sejong Electronic Dictionary. And from the raw corpus, the partially-parsed corpus are constructed automatically. To extract candidates for figurative expression, we fixed some syntactic patterns in advance, for example [NP+subj-marker NP+obj-marker V], and extract those patterns from partially-parsed corpus. After investigating the extracted expressions and classifying them whether figurative or literal, we concluded that some nominal entries are used more figuratively in specific syntactic patterns.

\* 이 연구는 2004년도 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구(연구과제번호 A00003)이다.

\*\* 주저자

\*\*\*교신저자

핵심어: 은유(metaphor), 비유적 표현(figurative expression), 다국어 어휘데이터베이스(multilingual lexical database), 코퍼스(corpus), 의미부류(semantic class), 의미위계(semantic hierarchy), 전자사전(electronic dictionary), 세종전자사전(Sejong Electronic Dictionary), 연어(collocation), 관용어(fixed expression)

### 1. 머리말

‘바닥을 치다, 황금 알을 낳는 거위’와 같은 언어표현은 우리의 일상 언어에서 흔히 사용되고 있다. 일상 언어생활에서 이와 같은 은유 표현이 차지하는 비중은 굉장히 크다.<sup>1)</sup> ‘바닥을 치다’와 같은 언어 표현이 은유적으로 사용된다면 ‘점수/등급/순위 따위가 최하위에 있다’라는 의미로 사용되지만, 우리 말 사전에서는 문자적 의미(literal meaning)에서 벗어나는 이러한 의미가 명확히 드러나지는 않는다.

(1) 가. 그는 성큼 그 안으로 들어가 무언가 살피는 듯했다. 그러다가 바닥에 떨어진 찢겨진 사이로 미세한 틈을 발견하고는 다시 손바닥으로 **바닥을 쳤다**. (오마이뉴스 “단장기”)

나. **주가가 바닥을 치고** 상승세로 전환했다는 데 공감하지만 단기급등에 따른 후유증이 우려되었던 것이다. (한겨레신문 2002/1/21)

(1.가)에서는 ‘바닥’이 문자적 의미로 사용되었지만, (1.나)에서 ‘바닥을 치다’는 비유적 표현(figurative expression)이다. 다음은 「표준국어대사전」에 풀이된 ‘바닥’의 의미이다.

- (2) 「표준국어대사전」 ‘바닥’의 뜻풀이
  1. 평평하게 넓이를 이룬 부분. 2. 물체의 밑부분. 3. 일이나 물건이 다 없어진 끝. 4. 지역이나 장소. 5. 피륙의 짜임새. 6. (광) 사금광의 감흙층 밑에 깔려 있

1) Lakoff & Johnson(1980)은 우리가 은유와 더불어 살고 있을 뿐만 아니라 은유를 통해서 세계를 이해한다고 말하고 있다. (김중도 2004)

는 굳은 층. 7. (경) 주가가 매우 많이 내려서 낮은 수준에 있는 상태. 또는, 주가가 하락하다가 상승 직전에 있는 상태.

「표준국어대사전」에서는 명사 ‘바닥01’의 의미를 7가지로 기술하고 있는데, 명사 ‘바닥’이 경제 분야에서 주가의 상태를 나타낼 때 사용된다고 설명되어 있다. 반면, 「연세한국어사전」에서는 공간의 의미에서 벗어나는 이러한 뜻은 설명되어 있지 않다. 또한 ‘바닥을 치다’라는 표현은 주요 사전들의<sup>2)</sup> 표제 단위(표제어, 부표제어, 관용어, 속담)로 올라 있지 않다.

(3) 국어사전별 ‘바닥’의 하위 표제항

ㄱ. 「표준국어대사전」

- 바닥(을) 긁다 「관용」
- 바닥(을) 누르다 「관용」
- 바닥(을) 보다 「관용」
- 바닥을 비우다 「관용」
- 바닥(을) 짚다 「관용」
- 바닥(이) 드러나다 「관용」
- 바닥(이) 질기다 「관용」
- 바닥 찢깨 「관용」
- 바닥 다 보았다

ㄴ. 「연세한국어사전」

- 바닥을 기다
- 바닥을 내다
- 바닥(이) 없다

(3)은 「표준국어대사전」과 「연세한국어사전」에서 ‘바닥’의 하위 표제항으로 등재되어 있는 목록들이다. (3. ㄱ)에서 ‘바닥(이) 질기다’라는 표현은 증권 거래에서 사용되는 「관용어」로 등재되어 있다. 그렇지만, 실제 경제 기사에서 많이 사용되고 있는 ‘바닥(을) 치다’와 같은 표현은 등재되어 있지 않다.

2) 본고에서는 ‘표준국어대사전, 연세한국어사전, 우리말 큰사전’을 주요 사전의 범주에 포함시켰다.

자연언어처리, 한국어 교육 등에 응용될 수 있는 어휘데이터베이스 구축을 목표로 할 경우, 이 표현들은 문어나 구어에서 매우 빈번하게 사용되고, 또한 의미의 불투명성과<sup>3)</sup> 연관되어 있기 때문에 이러한 목록을 확보하고 기술하는 것은 어휘데이터베이스 구축에 필수적이다.

그동안 우리말 은유 연구는 언어적 사실 자체보다는 수사적 기교나 인지의 의미론적 관점에서 주로 연구되어 왔다. 따라서 다양한 은유 표현을 확인하는 작업보다는 수사적 효과, 은유의 구성 원리 등을 규명한 논의가 상대적으로 많았다. 그러나 자연언어처리를 위해서는 먼저 어떤 표현이 은유인지 아닌지 판단하여 목록화하고 이를 대상으로 세부적인 분류와 해석을 하는 작업이 필요하다.<sup>4)</sup>

일상 언어에서 은유는 끝없이 생성되므로 은유 표현의 수는 무한하다. 그러나 ‘수필은 난이요, 학이요, 청초하고 옷맵시 날렵한 여인이다’와 같은 담화 차원의 은유는 문맥 의존적으로만 해석이 가능하므로 당장 어휘 데이터베이스에 수용할 수는 없을 것이다.<sup>5)</sup> 따라서 어휘 데이터베이스에서 수용할 수 있는 은유 표현 단위는 최소 ‘단어’에서 최대 ‘문장’까지가 될 것이다.

지금까지 은유 표현의 사전화 작업이나 어휘 데이터베이스화 작업은 ‘단어’ 차원에서만 이루어졌을 뿐<sup>6)</sup> ‘황금 어장, 황금 알을 낳는 거위’와 같은 단어 이상의 단위에 대해서는 본격적으로 시도된 바 없다. 그러나 구나 문장 단위의 은유가 실제로 많이 쓰이므로 이 은유 표현들을 찾아서 어휘 데이터베이스 안에 반영하는 일이 필요하다. 이 경우 어떤 표현이 최종적으로 은유

- 3) ‘바닥을 치다’와 같은 예는 개별 요소들의 결합으로 새로운 제3의 의미가 발생된다. 이처럼 은유 표현 중 일부는 관용어로 굳어진다.
- 4) 자연언어처리 과정에서 어떤 표현이 은유적으로 쓰였는지 여부를 판별해 내는 것과 이 연구에서 논의하고 있는 은유표현의 추출 과정은 별개의 문제이다. 이 연구에서는 자동/수동의 방법으로 어휘데이터베이스에 은유표현을 추가하는 방법론을 제시하고자 하는 것을 목적으로 하고 있다.
- 5) 물론 은유의 전모를 파악하기 위해서는 화용론적 차원의 논의를 포함한 다각적 검토가 필요하다. 그러나 현 단계에서, 은유에 관한 화용론적 차원의 지식이나 은유의 해석, 생성 과정에 관한 지식을 어휘 데이터베이스 안에 수용하기에는 시기상조로 보인다.
- 6) 종이사전에서 ‘-을 비유적으로 이르는 말’ 유형으로 뜻풀이 된 다의어가 여기에 속한다. 단어 단위의 은유 표현은 의미의 확장에 해당하므로 기존의 종이사전이나 전자 사전에서 다의어로 처리하는 것이 현실적인 방안이다.

인지 언어인지, 관용어인지에 대한 분류 기준을 설정하고 이를 어휘데이터베이스 목록 확장에 반영하는 것도 중요한 문제가 된다.

요컨대, 은유를 어휘 데이터베이스에 반영할 때는 먼저 은유 표현을 수집하고 다음으로, 이 표현을 일정한 분류 기준에 따라 관용구나 언어, 혹은 단어 단위의 은유로 판정하는 일이 중요하다. 본고에서는 이 두 가지 문제를 다룰 것이다.<sup>7)</sup>

## 2. 코퍼스를 이용한 은유표현의 추출

본 연구에서는 5만여 개의 표제어로 구성된 다국어 어휘데이터베이스를 구축하는 과정에서 ‘은유 표현을 중심으로 한 확장 의미’를 다국어 어휘데이터베이스<sup>8)</sup>에 반영하기 위한 자료 구축 방법론을 수립하고자 하였다. 코퍼스에서 은유표현을 추출하기 위해서 형태주석정보가 부가된 8,553,690 어절 규모의 경제 관련 신문 기사 코퍼스를<sup>9)</sup> 사용하였다.

먼저 대상 코퍼스에서 ‘바닥’을 검색하여 총 450개의 용례를 찾아내었다.<sup>10)</sup>

- 
- 7) 이 연구의 목적은 은유 표현을 어휘데이터베이스에 반영할 수 있도록 대용량 코퍼스에서 추출하는 방법을 모색해 보는 것이다. 부분구문분석을 이용하여서 중심어의 공기 관계를 추출하는 방법은 은유적 표현만이 아니라 비은유적 표현, 특히 언어 관계를 이루는 어휘 구성의 자동 추출에도 똑같이 활용될 수 있다. 이 연구에서는 은유적 표현이라고 판단할 수 있는 후보 목록의 추출을 위해서 21세기 세종계획의 의미부류 사전을 부분적으로 활용하였다.
- 8) 이 연구에서 구축하고 있는 다국어 어휘데이터베이스는 Choi et al. (2006)에 간략히 소개되어 있으며, 상세한 내용은 강범모 외(2005)에 소개되어 있다. 전체 표제어는 한국어 5만여 개의 표제어로 구성되어 있으며, 여기에 영어/중국어/일본어 항목이 대응되고, 각 표제어(한국어)에는 유로워드넷의 ILI가 대응되는 방식으로 구성되어 있다.
- 9) 본고에서 사용한 자료는 최운호(2005)에서 구축한 자료의 일부로 1999년 12월 ~ 2003년 1월까지 인터넷한겨레에서 서비스된 기사 6400만 어절 중 경제 분야 기사 850만 어절이다.
- 10) ‘바닥’은 임의적으로 추출한 대상어이다. 원래는 465개이나 수작업으로 검토하는 과정에서 일부 중복 예를 제외하여 450개가 되었다. 코퍼스를 대상으로 하였으므로 ‘바닥나다’나 ‘밑바닥’과 같은 예는 추출되지 않았으나 ‘바닥치다’는 용례로 검색되었다. 이 중 ‘바닥나다’는 『표준국어대사전』에도 한 단어로 처리되어 있으나 ‘바닥치다’는 한 단어로 등재되어 있지 않다. 본고는 ‘바닥치다’를 종이 사전을 기준으로 ‘바닥(을) 치다’와 같은 두 단어로 간주하였다.

다음으로 수작업을 통해 해당 용례를 은유적 의미를 가진 것과 그렇지 않은 것으로 나누어 본 결과 450개 중 은유 표현은 328개, 비은유 표현은 122개로 판정하였다.<sup>11)</sup> 이 결과는 일상 언어에서 은유 표현이 빈번하게 사용되리라는 예상을 확인시켜 주는 것이다. 다음으로 컴퓨터를 이용하여서 은유 표현을 추출하는 작업을 염두에 두고, 은유 표현과 비은유 표현을 통사 구조에 따라 분류하였다. 이때 해당 통사 구조는 ‘바닥’이 포함된 어절을 중심으로 좌우로 1어절씩만을 추출하여 판정하였고, 앞뒤 1어절만 보고 은유 표현인지 여부를 결정할 수 없는 경우에는 ‘t. 기타’로 표시하여 용례를 분류하였다.

다음 <표1>과 <표2>는 한겨레신문 경제 기사 코퍼스에서 ‘바닥’을 중심으로 검색하여 추출한 용례를 은유 표현과 비은유 표현으로 판정하고 그 결과를 통사 구조를 중심으로 분류하여 빈도순으로 제시한 것이다.

은유표현 (328회)	
통사 구조	빈도
<b>a. NP1-이 NP2(=‘바닥’)을 V</b>	194 (59.15%)
NP1 = 주가, 소비심리, 실질금리, 경기, 1분기, 자금, 재정, 공적 자금, 재원, 품질순위, 매출, 환율, 일본경제, 주택시장, 재고, ... V = 기다, 기록하다, 높다, 다지다, 드러내다, 던다, 땀들다, 면하다, 모르다, 벗어나다, 보이다, 지나다, 찍다, 찾다, 치다, 탈출하다, 통과하다, 헤매다, 확인하다	
<b>b. NP1(=‘바닥’) NP2</b>	33 (10.06%)
NP2 = 가격, 경기, 국면, 상태, 수준, 시점, 탈출, 통과, 형성	
<b>c. NP1-이 NP2(=‘바닥’)에/까지 V</b>	27 (8.23%)
NP1 = 경기, 주가, 자금, ... V = 가까워지다, 근접하다, 깔다, 다다르다, 도달하다, 이르다/떨어지다, 오다	
<b>d. NP1-이 NP2(=‘바닥’)이 V/A</b>	18 (5.49%)
NP1 = 재고, 유동성, 증시 V/A = 나다, 되다, 드러나다, 보이다, 확인되다	

11) 은유적 의미에 관한 판정은 필자들의 직관에 의존하였지만 분석 결과를 종합해 보면 결과적으로는 ‘바닥’이 지시하는 대상이 물리적/공간적 의미의 바닥인지 아닌지에 따라 결정된 것이라 하겠다.

은유표현 (328회)	
통사 구조	빈도
e. NP1-이 NP2(='바닥')이다	14
NP1 = 주가, 경기, 집값	(4.27%)
f. NP1-이 NP2(='바닥')으로 V	7
NP1 = 주가, 경기	(2.13%)
V = 곤두박질치다, 떨어지다, 추락하다	
g. NP1-이 NP2(='바닥')에서 V	7
NP1 = 경기, 한국경제	(2.13%)
V = 벗어나다, 빠져나오다, 탈출하다, 헤어나다	
h. NP1 NP2(='바닥')	4
NP1 = 경기, 단기	(1.22%)
t. 기타	24
	(7.32%)

<표1> '바닥'의 은유 표현

비은유표현 (122회)	
통사 구조	빈도
a. NP1(='바닥') NP2	54
NP2 = 높이, 단열, 마감재, 매트, 먼, 먼적, 부분, 색, 충격, 충격음, 충격도, 충격소음, 틈새, 포장	(44.26%)
b. NP1 NP2(='바닥')	23
NP1 = 1층, 3층, 건축물, 대리석, 등산화, 베란다, 샌들, 시루, 신발, 운동화, 중앙, 콘크리트, 표준, 플레이트	(18.85%)
c. (NP1-을) NP2(='바닥')에 V	8
V = 깔다, 설치하다, 세워놓다, 얹어지다, 주저앉다	(6.56%)
d. NP1(='바닥')-의 NP2	6
NP2 = 두께, 매트, 왁스, 접착부분, 틈, 홈	(4.92%)
e. NP1(='바닥')-이/도 V/A	6
V/A = 깊다, 고르다, 두툼하다, 평평하다	(4.92%)
f. V/A NP1(='바닥')	3
V/A = 갈라지다, 장식하다	(2.46%)

비은유표현 (122회)	
통사 구조	빈도
<b>g. NP1(='바닥'), NP2</b>	2
NP2 = 바닥, 벽면	(1.64%)
<b>h. NP1(='바닥')-과 NP2</b>	2
NP2 = 모서리, 옆면	(1.64%)
<b>i. NP1(='바닥')-을 V</b>	2
V = 다지다	(1.64%)
<b>t. 기타</b>	16
	(13.11%)

<표2> '바닥'의 비은유 표현

<표 1>과 <표 2>에서 몇 가지 특징을 찾아낼 수 있다.

첫째, 은유 표현과 비은유 표현은 선호하는 통사 구조에서 차이를 보인다. 은유 표현에서는 'NP1-이 NP2(=바닥)-를 V' 구조가 가장 고빈도인 반면, 비은유 표현에서는 'NP1(=바닥) NP2' 구조가 가장 많이 나타난다. 반면, 비은유 표현에서는 'NP1-이 NP2(=바닥)를 V'가 드물게 나타나며, 은유 표현에서는 'NP1 NP2(=바닥)' 유형이 거의 나타나지 않는다.<sup>12)</sup>

둘째, 은유 표현과 비은유 표현은 중심어 '바닥'과 인접하고 있는 선, 후행 어휘의 의미 부류에서 확연한 차이를 보인다. 동일한 통사 구조 'NP1(=바닥) NP2'에서, 은유 표현에 나타난 NP2는 '가격, 경기, 국면, 상태, 수준, 시점, 탈출, 통과, 형성' 등 [-구체성]을 가지고 있는 반면, 비은유 표현에 나타난 NP2인 '높이, 단열, 마감재, 매트, 면, 면적, 부분, 색, 충격, 충격음, 충격도, 충격소음, 틈새, 포장' 등은 [+구체성]을 가지고 있다.<sup>13)</sup>

셋째, 비은유 표현은 인접어의 어휘 의미 부류 정보만 가지고도 어느 정도 탐색이 가능하다. 반면, 은유 표현에서는 주어의 의미 부류가 중요한 역할을 한다. 즉, 주어가 '주가, 경기, 경제, 매출' 등일 경우에 은유로 해석된다. 이것

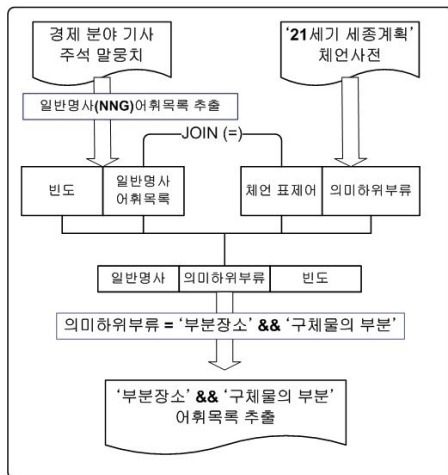
12) 남기심(1995)에서는 어떤 낱말이 전이 의미를 가지고 쓰일 때는, 공기 제약, 형태적인 제약, 통사적인 제약을 가지고 나타난다는 사실을 밝히고 있다. 이 논의에 따르면, 은유는 의미 전이의 주요 기제이므로 은유 표현과 비은유표현이 선호하는 통사 구조가 다르게 나타날 수밖에 없다.  
 13) 본고의 의미 자질 판정은 세종사전의 의미 부류 체계에 따른 것이다.

은, 'NP1이 NP2(=바닥)를 V'와 같은 구조에서 동사가 가진 선택 제약이 은유의 성립에 중요한 역할을 한다는 의미이다.

이러한 특징들은 은유 표현을 추출하는 과정에서 중요한 단서를 제공한다. 본고에서는 이상의 관찰을 토대로 코퍼스에서 은유 표현을 (반)자동으로 추출하는 과정을 다루어 보려 한다. 3장에서는 대상이 되는 핵심어를 추출하는 방법에 대해 설명하고, 4장에서는 핵심어와 코퍼스를 이용한 은유 표현의 추출 과정을 다루어 본다.

### 3. 핵심어 선정

은유 표현의 추출을 위한 핵심어 선정을 위해서는 두 가지 자료를 이용하였다. 하나는 형태 주석이 첨가된 한겨레신문 경제분야 기사 말뭉치이고, 다른 하나는 "21세기 세종계획"에서 구축한 체언사전이다. 21세기 세종계획의 체언사전에 기술된 정보 중에서 어휘 의미부류 정보를 코퍼스에서 추출한 어휘목록에 대응시켜서, 핵심어 후보들을 구성하였다. <그림1>은 이 연구에서 은유표현 추출을 위한 핵심어 선정의 절차를 보여준다.



<그림1> 은유 표현 추출을 위한 핵심어 선정 과정

먼저 한겨레 말뭉치의 경제분야 기사에서 일반명사만을 추출하여 빈도를 구하고, 이 자료를 “21세기 세종계획”의 체언사전 표제어와 비교하였다. 이 과정에서 “21세기 세종계획”에 등록된 의미 부류가 존재하는 경우, 한겨레말뭉치에서 추출한 일반명사 빈도 목록에 “21세기 세종계획”의 의미부류를 추가하였다. 구체적으로는 2장의 어휘 ‘바닥’이 지닌 특성, 즉 ‘구체물의 부분’과 ‘부분 장소’라는 속성을 동시에 지니고 있는 어휘들을 추출하여 상위 8개의 핵심어를 선정하였는데 결과는 <표3>과 같다.

순위	표제어	빈도	순위	표제어	빈도
1	창구	980	5	복판	25
2	바닥	465	6	상단	18
3	줄기	139	7	칸	10
4	목	115	8	가장자리	4

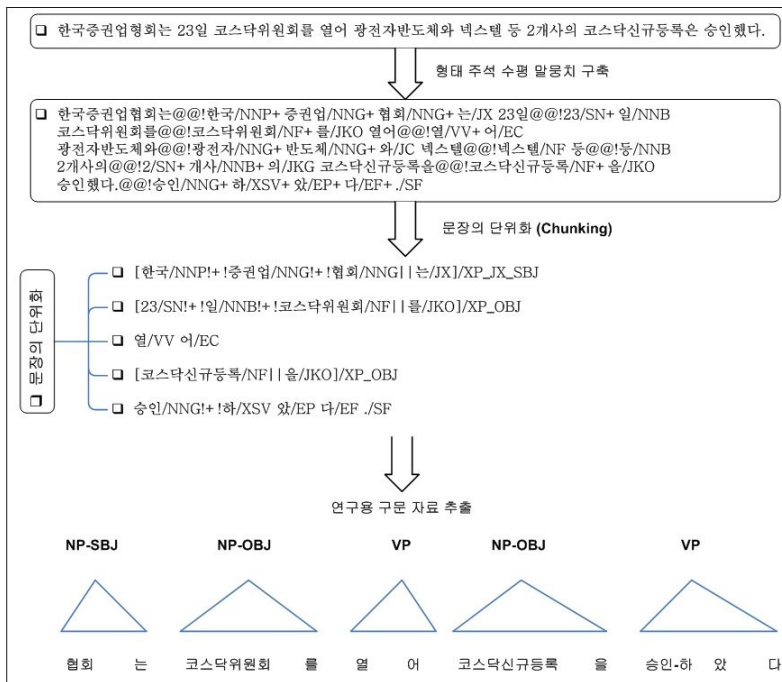
<표3> 핵심어 목록

<표3>을 보면 2장에서 임의적으로 선정한 ‘바닥’이 포함되어 있음이 흥미롭다. 3장에서 추출한 핵심어 목록을 이용해서 경제분야 기사 코퍼스에서 특정 구문 유형을 추출하여서 은유적 표현과 비은유적 표현이 어떻게 사용되는지를 4장에서 보여주도록 한다. 그리고 5장에서는 언어·관용어를 어휘데이터베이스에 반영하기 위해서 판정하는 기준을 ‘바닥’을 대상으로 설명하도록 한다.

#### 4. 핵심어와 코퍼스를 이용한 은유 표현의 추출

핵심어가 포함된 은유 표현을 추출하기 위해서 한겨레 말뭉치의 경제분야 기사를 분석하였다. 구문 분석 말뭉치를 이용할 수 있다면 특정 어휘의 구문 용례를 추출하는 것이 가능하지만, 대용량으로 이용할 수 있는 말뭉치가 아직 존재하지 않기 때문에 이 연구를 위해서 형태분석 말뭉치를 대상으로 부분구문분석 말뭉치<sup>14)</sup> 구성하기로 한다. 부분구문분석 말뭉치 구축 방

식은 다음 <그림2>와 같다.<sup>15)</sup>



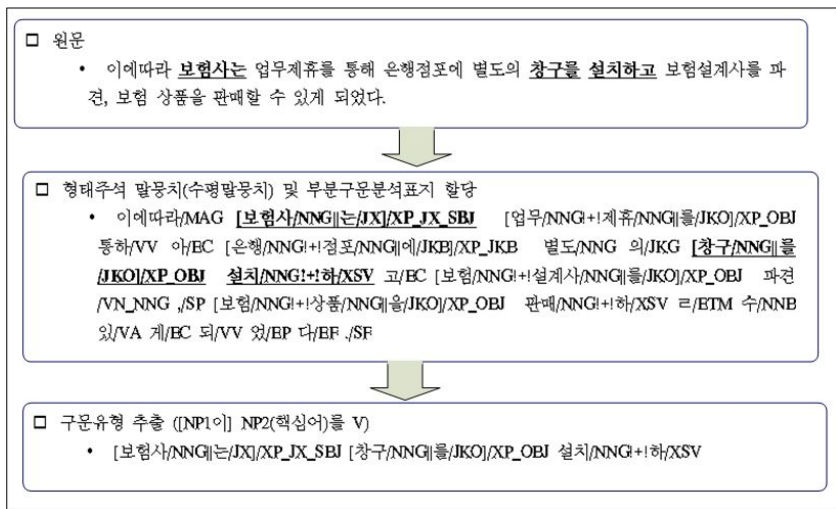
<그림2> 은유 표현 추출을 위한 부분구문분석 말뚝치 구축 과정

<그림2>의 과정을 통해서 구축된 부분구문분석 말뚝치에서 <표 3>의 관찰 대상 어휘가 포함된 자료를 각각 추출하여 개별 어휘마다 특정 구문 유형 별로 분류된 자료를 구성하였다. 개별 어휘마다 추출한 구문 유형은 <표1>에서 제시된, NP-이 NP-을 VP, NP-이 NP-에/까지 VP, NP-이 NP-에서 VP 등의 구성이고 그 중에서 NP의 머리어가 관찰의 대상으로 삼은 어휘들인 경우를 다시 걸러서 자료를 구축하였고, 다음과 같은 결합관계를 보이며

14) 부분구문분석(partial parsing)은 문장의 구문구조를 완전하게 분석하지 않고 간략한 분석을 통해서 활용하고자 할 때 이용하는 방법이다.  
 15) <그림 2>에서 소개하고 있는 부분구문분석 말뚝치 구축 및 특정 구문 유형 추출 방법은 최운호(2005)의 방법을 따랐다.

나타나는 구문유형들을 각각 1단계 추출 대상 유형과 2단계 추출 대상 유형으로 삼았다.

- 1단계<동사구, 문장>: 동사와의 결합관계<sup>16)</sup>  
 [NP1-이] NP2(핵심어)-를 V  
 [NP1-이] NP2(핵심어)-에/까지 V  
 [NP1-이] NP2(핵심어)-에서 V  
 [NP1-이] NP2(핵심어)-이 V/A  
 [NP1-이] NP2(핵심어)-로 V  
 [NP1-이] NP2(핵심어)-에서 V
- 2단계<명사구>: 명사와의 결합 관계  
 NP1(핵심어) NP2  
 NP1 NP2(핵심어)  
 NP1-의/JKG NP2(핵심어)



<그림3> 구문 유형 추출 과정

16) [NP1-이]에서 주어의 표지는 {이/가}, {은/는}을 기본적으로 설정하였고, 각각의 조사가 주격 표지로 사용되었는지 여부는 자료의 관찰을 통해서 수정하였다.

자료의 추출 과정을 ‘원문 → 형태주석 말뭉치(수평말뭉치) → 부분구문분석 표지 할당 → 구문유형 추출’로 간략히 예시하면 <그림 3>과 같다. 이와 같은 구문유형 추출 과정을 통해서 구성된 자료를 일부 제시하면 다음과 같다.

(4) ‘창구’

금융노조-는 대화창구-를 열-어 ([NP1이] [NP2를] V)

대화-의 창구 (NP1-의 NP2)

로비-의 창구 (NP1-의 NP2)

마케팅-의 창구 (NP1-의 NP2)

창구 지도 (NP1 NP2)

(5) ‘줄기’

정부-는 정책 줄기-를 바꾸-지 ([NP1이] [NP2를] V)

소문-이 음모론-의 줄기-를 이루-는다 (NP1-의 NP2)

통화정책-의 줄기 (NP1-의 NP2)

(6) ‘목’

특수사업-이 권씨-의 목-을 죄다 ([NP1이] [NP2를] V)

수사관들-이 목-을 조르다 ([NP1이] [NP2를] V)

경제-가 현대-에 목-을 매다 ([NP1이] [NP2를] V)

아기-가 목-을 젖히-다 ([NP1이] [NP2를] V)

각각의 구문유형별 자료는 세종사전의 용언사전 및 체언사전의 ‘선택제약’ 정보와 ‘의미하위부류’ 정보와 비교하면서 은유표현 여부를 판가름한다. ‘창구’의 의미는 표준국어대사전과 세종계획 전자사전에서 각각 다음과 같이 기술되어 있다.

(7) ‘창구(窓口)’의 의미 (표준국어대사전)

「1」창을 내거나 뚫어 놓은 곳. 「2」사무실이나 영업소 따위에서, 손님과 문서·돈·물건 따위를 주고받을 수 있게 조그마하게 창을 내거나 대(臺)를 마련하여 놓은 곳. 「3」단체나 기관 따위에서 서로 협조하거나 외부 사람과 직접 접촉하기 위하여 마련하여 놓은 연락 부서를 **비유적으로 이르는 말**.

(8) 세종계획 전자사전에서 ‘창구(窓口)’의 의미하위부류

- 의미 1: 구체물의 부분  
길쪽으로 나 있는 조그만 ~가 죄수들이 외부와 연락을 할 수 있는 유일한 통로였다.
- 의미 2: 부분장소  
고객들은 은행 ~에 갈 필요 없이 언제든지 은행 CD/ATM기에서 공과금을 납부할 수 있게 되었다.
- 의미 3: 경로  
이 회사의 홈페이지는 고객과 직접 커뮤니케이션 할 수 있는 ~가 되었다.

(7)과 (8)의 자료를 보면, 표준국어대사전의 뜻풀이와 세종계획 전자사전의 의미하위부류 구분은 동일하다고 판단할 수 있으며, 표준국어대사전에서 ‘비유적으로 이르는 말’로 분류되어 있는 뜻풀이가 세종계획 전자사전에서는 ‘경로’라는 의미부류로 할당되어 있는데, 세종계획 전자사전의 의미하위부류에서는 비유적인 표현과 문자적 의미가 구분되지 않고 있다는 것으로 볼 수 있다. 세종계획 전자사전에서 기술된 ‘창구’의 의미하위부류를 의미부류체계에서 각각 찾아보면 다음과 같은 위계를 보이고 있다.

(9) ‘창구’의 의미하위부류 위계

- 구체물 > 관계구체물 > 부분 > 구체물의 부분
- 장소 > 부분장소
- 추상적대상 > 경로

‘창구’의 의미부류에서 ‘경로’라는 의미부류는 이미 문자적인 의미를 벗어난 것으로 보고, ‘구체물의 부분’과 ‘부분장소’라는 속성에 해당하는 것을 축자적인 의미로 사용된 것으로 보고, ‘경로’라는 의미로 사용된 것을 은유적 의미로 판단하기로 한다. ‘창구’가 머리어로 사용된 명사가 “[NP1이] [NP2(= ‘창구’)를] V”에서 사용된 예는 186회이다. 이 중에서 ‘창구’가 문자적 의미가 아닌 의미로 사용된 경우는 모두 51회이다. 다음은 그 예이다.

(10) [NP1이] [NP2(='창구')를] V 용례 (186회)

ㄱ. 문자적 의미 (135회)

응답자의 56.6%가 은행창구를 이용하다, 비씨카드는 기업회원진담창구를 운영하다, 영업점창구를 방문하다, 수용자들은 은행창구를 이용하다, ...

V = 개설하다, 늘이다, 단일화시키다, 드나들다, 마련하다, 바꾸다, 방문하다, 설치하다, 열다, 운영하다, 이용하다, 점검하다, 제공하다, 찾다, 찾아가다, 통하다, 확보하다, 활용하다

ㄴ. 은유적 의미 (51회)

재계는 경험단일창구를 만들다, 금융노조는 대화창구를 열다, 공식대화창구를 홍보팀으로 일원화하다, 정부책임자급이 대화창구를 만들다, ...

V = 뚫다, 마련하다, 만들다, 바꾸다, 복원하다, 빛다, 설치하다, 열다, 운영하다, 일원화하다, 정하다, 추진하다, 통일하다, 통하다, 합의하다, 활용하다

(10)에서 보듯이 명사 '창구'를 목적어로 취하는 동사는 다양하게 나타나며, 문자적 의미와 은유적 의미로 사용될 때 모두 사용되는 동사들도 상당수 있다.

한편, "[NP1의] [NP2='창구']"와 같은 구성에서는 '창구'의 의미가 은유적으로 사용되는 경향이 많은데, 예를 보면 다음과 같다.

(11) [NP1의] [NP2='창구']

ㄱ. 문자적 의미

감독 기관의 창구 지도

ㄴ. 은유적 의미

북한쪽의 창구 부재, 대화의 창구, 민간분야 교류 활성화의 창구 역할, 정·관계 로비의 창구, 마케팅의 창구

(11)의 구성에서 문자적 의미로 사용된 경우는 '지도'라는 명사의 의미가 [장소]라는 의미부류로 사용된 경우이다. 그 이외에는 대부분 은유적 의미로 사용되었다.

명사 '줄기'의 경우는, 실제 사용된 용례 대부분이 '배아줄기세포', '성체줄

기세포'와 같은 복합어의 일부로 사용되었지만, 동사의 목적어 구성으로 사용되거나 'NP1-의 NP2(=줄기)' 구성으로 사용된 경우가 있다.

(12) [NP1이] [NP2(='줄기')를] V 용례 (5회)

- 정부는 정책줄기를 바꾸다, ~ 것이 줄기를 이루다, 큰 줄기를 잡지 못하고 사소한 자구 수정에 매달리다, 소문이 음모론의 줄기를 이루다, ~ 기본 줄기를 벗어던지다

명사 '줄기'의 의미를 [식물의 부분], [구체물의 부분]이라는 의미로 한정한다면, 경제분야에서는 명사 '줄기'가 동사의 목적어로 사용되는 경우 대부분 은유적 의미로 사용되는 것으로 볼 수 있다. 그 이외의 구문 유형으로는 [NP1의] [NP2='줄기']가 있는데, 이 유형으로 사용된 경우에도 은유적인 표현으로 사용되고 있는 것을 볼 수 있다.

(13) [NP1의] [NP2='줄기']

- 통화정책의 줄기

명사 '목'은 [구체물의 부분], [신체부위], [부분장소] 등의 의미로 사용되고 있는데, [NP1이] [NP2(='목')를] V 구문유형의 경우 모두 27회의 용례 중에서 13회의 용례가 은유적인 의미로 사용되었다.

(14) [NP1이] [NP2='목']를 V (27회)

ㄱ. 문자적 의미 (14회)

수사관들이 목을 조르다, (목걸이를 감아) 긴 목을 강조하다, (상처를 감추거나) 긴 목을 강조하다, 목을 길게 빼다, 목을 보이다, 좋은 목을 차지하다(부분장소), 아기가 목을 젖히다, 아기가 목을 가누다, 안전띠는 아기의 목을 조이다, 목을 보호하다, 유아가 목을 지탱하다, 목을 따뜻하게 하다, 목을 돌리다, 투자자가 목을 주무르다

ㄴ. 은유적 의미

특수사업이 권씨의 목을 조르다, 경제가 목을 매다, 채권단이 일거수 일투족에 목을 매달다, 최고 경영자의 목을 내놓다, 경제가 ~ 부문에 목을

매다, 삼성생명이 생보사 상장에 목을 매다, 지원 대책에 목을 매다, 둘다 미국 경제 향방에 목을 매달다, 기업의 목을 죄다, 나라 바깥 상황에 목을 매다, 광고들이 장면 연출에 목을 매다, 완성차 업체에 목을 매다, 현대전자도 반도체 경기에 목을 걸다

## 5. 은유 표현과 연어, 관용어 판정

4장에서는 핵심어의 의미 부류를 통해서 반자동적으로 은유 표현을 추출하는 과정을 살펴보았다. 이렇게 추출된 은유 표현은 궁극적으로 어휘 데이터베이스 내에서 연어나 관용어로 귀속될 것이다. 5장에서는 2장의 ‘바닥’ 관련 은유 표현을 중심으로 연어, 관용어로의 판정 과정에 대해 살펴본다.

### 5.1. ‘바닥’의 확장 의미

‘바닥’ 관련 은유 표현을 연어/관용어로 판별해 내려면 먼저 이러한 은유 표현에 나타난 ‘바닥’의 어휘 의미를 파악할 필요가 있다. 이를 위해서 <표 1> ‘바닥’의 은유 표현을 중심으로 공통 의미를 확인해 본다.

<표 1>에서 ‘바닥’은 ‘주가, 집값, 경기 등 등수화될 수 있는 것, 혹은 단계를 가지는 것 중 최하위 상태’라는 어휘 의미를 가지고 있다. 즉, ‘바닥’은 [공간], [장소]라는 구체적 의미에서 [상태]라는 추상적 의미로 의미의 확장을 겪고 있다. 참고로 이러한 확장 의미가 기존 사전에 어떻게 반영되어 있는지 살펴보면 다음과 같다.

<input type="checkbox"/> 연세한국어 사전 ‘바닥’의 의미
① (사람이 서든가 물건이 놓일 수 있도록) 평평하고 넓게 이어진, 공간의 아래쪽 평면 ② (건축에서) 건물 안의 아래쪽. ③ (개울, 호수, 강 따위가) 물 아래의 땅과 닿는 쪽. ④ 병이나 그릇의 속의 가장 아래쪽. ⑤ 겉으로 드러나 있지 않는 가장 깊은 부분. ⑥ ‘길/장 바닥’의 꼴로 쓰여, 그러한 곳의 땅에 닿는 공간. ⑦ 신이나 버선 같은 것의 밑을 이루는 평평한 공간. ⑧ (지역 이름 따위에 붙어 그 이름의) 지역.
<input type="checkbox"/> 표준국어대사전 ‘바닥’의 의미
① 평평하게 넓이를 이룬 부분(책상의 바닥). ② 물체의 밑부분(구두 바닥). ③ 일이나 물건이 다 없어진 끝. ④ 지역이나 장소 ⑤ 피륙의 짜임새 ⑥ 사금광(砂金鑛)의 감흙층 밑에 깔려 있는 굳은 층. ⑦ 주가가 매우 많이 내려서 낮은 수준에 있는 상태
<input type="checkbox"/> 21세기 세종계획 전자사전 ‘바닥’의 의미 (사용 예)
① 바닥에 깔다<공간> ② 솔바닥의 누룽지<공간> ③ 좁은 시골 바닥에서 살지만 마음은 편하이.<장소> ④ 이 천은 바닥이 좀 굵다.<상태> ⑤ 그의 성적은 바닥에서 논다.<상태>

<표4> 기존 사전의 확장 의미의 처리

<표4>를 보면 「연세한국어사전」에는 [공간]과 [장소]만 나와 있는 반면, 「표준국어대사전」과 「21세기 세종계획 전자사전」에는 [상태]라는 추상적 의미가 반영되어 있음을 알 수 있다. ‘바닥’의 확장 의미가 반영된 「표준」과 「세종」 두 사전에서도 ‘바닥’이 은유에 의한 의미 확장을 보인다는 정보는 반영되어 있지 않은데, 이러한 정보 역시 어휘 데이터베이스에 반영될 필요가 있다. <표1>에서 명사구인 b, h를 제외한 언어 단위들은 다음과 같은 틀을 가진다.

- (15) [집값, 주가, 자금...(등수화 될 수 있는 것, 혹은 단계를 가지는 것)]-이 바닥-  
을/에/(으)로/에서 어찌하다.

여기서 ‘어찌하다’에 속하는 대부분의 동사들은 행위자로서 유정물을 전제하므로 ‘바닥’의 확장 의미를 중심으로 전체 문장의 의미를 해석하려면 NP1에 대한 정보를 이용하는 일이 중요하다. 즉, 문장 의미가 개별 어휘의 합의 합이라고 할 경우 이러한 구문의 해석에는 ‘바닥은 상태이다’라는 은유 정보와 ‘NP1(=집값, 주가, 자금, …)-은 유정물이다’라는 은유 정보가 요구된다. 이러한 내용을 전제로 다음 절에서는 연어, 관용어 판정의 문제를 다루어 본다.

## 5.2. 연어 · 관용어 판정

먼저 연어와 관용어의 개념에 대해 간단히 정리해보자. 연어에 대한 정의는 연구자마다 조금씩 다르지만 대체로 광의의 연어는 어휘들의 단순한 공기 현상으로, 협의의 연어는 특정 어휘가 다른 어휘를 요구함으로써 발생하는 어휘소들 간의 제한적 공기 관계로 규정할 수 있다. 또 연어는 구성 요소의 의미 반영 면에서는 자유 결합과 가깝고, 빈도수나 심리적 현저성에서는 관용어에 가깝다는 특징을 가지고 있다.<sup>17)</sup>

반면 관용어는 두 의미 구성 요소가 모여 질적으로 다른 새로운 의미를 낳게 되는 언어 단위이다(A+B=C). 관용어는 비유적, 수사적 차원에서 생성된 구성이 사은유화를 거친 결과로 나타나며, 해당 관용 의미와 개념적으로 동일한 대응어를 상정할 수 있다는 특징을 가지고 있다.

본고에서는 단순 공기, 구성소 의미 유지, 대응어 상정이라는, 비교적 명시적인 세 가지 기준에 따라 ‘바닥’ 관련 은유 표현을 판정하도록 한다. 이 기준에 의하면 <표1>의 은유 표현 ‘바닥을 기다, 바닥을 벗어나다, 바닥을 치다, 바닥 가격, 바닥에 근접하다, 바닥이 드러나다, 바닥이다, 바닥으로 떨어지다, 바닥에서 탈출하다, 바닥 경기’의 판정 결과는 다음과 같다.

17) 김진해(2000: 29) 참조

은유 표현	제한적 공기	구성소 의미 유지	대응어 상정	결과
바닥을 기다	X	△	○	관용어
바닥을 벗어나다	X	○	X	연어
바닥을 치다	X	△	○	관용어
바닥 가격	X	○	X	연어
바닥에 근접하다	X	○	X	연어
바닥이 드러나다	X	○	X	연어
(무엇이) 바닥이다	X	○	X	연어
바닥으로 떨어지다	X	○	X	연어
바닥에서 탈출하다	X	○	X	연어
바닥 경기	X	○	X	연어

<표5> 연어·관용어 판정 결과

<표 5>의 은유 표현은 ‘시치미를 떼다’와 같이 제한적 공기 관계를 가지지 않으므로 모두 광의의 연어에 해당한다. 다음으로 구성소의 의미 유지라는 기준으로 보면, ‘바닥을 기다’와 ‘바닥을 치다’의 경우는 ‘바닥 - 상태’의 의미가 유지된다고 보기는 어렵다. 다시 말해 ‘최하위 상태’로서의 ‘바닥’과 은유의 과정을 설정한다고 해서 의미적 투명성이 모두 보장되는 것은 아니다. 예를 들어 ‘바닥’과 ‘어찌하다’가 결합할 경우 ‘떨어지다, 벗어나다’에 비해서 ‘기다, 치다’ 등은 의미적으로 투명성의 정도가 좀 떨어진다고 할 수 있다.<sup>18)</sup>

구성소 의미 유지 문제는 대응어 상정의 문제와도 연결된다. 즉, ‘바닥으로 떨어지다’나 ‘바닥에서 벗어나다’의 경우는 ‘최하위 상태로 떨어지다’, ‘최하위 상태에서 벗어나다’로 해석되고 이에 대한 대응어를 찾지 않아도 되지만, ‘바닥을 기다’나 ‘바닥을 치다’의 경우는 ‘최하위 상태를 기다’나 ‘최하위 상태를 치다’로 해석하면 어색하다. 이 경우는 ‘정도나 수준이 최하위로 떨어지다’와

18) 물론 의미적 투명성의 판정 과정에도 어느 정도 주관의 개입될 수 있다. 실제로 이 문제에 대해서는 사전마다 처리 방법이 다르다. 가령, 세종 사전에서는 ‘바닥을 기다’와 ‘바닥을 치다’를 모두 연어로 처리한 반면, 연세 사전에서는 ‘바닥을 기다’만을 관용어로 등재하였다. 반면 표준사전에서는 두 가지 모두 표제 단위로 올라 있지 않다. 이런 점에서 본고와 같은 논의가 더 활발히 이루어져야 할 것이다.

같은 대응 표현을 설정하는 것이 더 자연스럽다. 요약하면, ‘최하위 상태’로서의 ‘바닥’이라는 비유적 뜻을 인정하되, ‘바닥을 기다/치다’는 따로 관용구로 처리하여 등재하는 방법이 효율적이라는 것이다.

## 6. 맺음말

이 연구에서는 코퍼스와 세종전자사전의 의미부류를 이용하여 은유 표현을 반자동적으로 추출하고, 이렇게 추출된 구, 문장 단위의 은유 표현을 연어와 관용어로 판정하는 과정을 살펴보았다. 은유표현의 추출 과정에서 중요한 것은 핵심어의 선·후행어의 의미 부류와 선택제약정보이다. 또, 은유 표현을 연어와 관용어로 판정하는 과정에서 중요한 것은 구성소의 의미 유지 가능성임을 확인하였다. 이 연구에서는 경제분야 신문기사 코퍼스만을 대상으로 하였지만, 이러한 연구는 여러 유형의 코퍼스로 확대될 필요가 있다. 이와 같은 작업을 통해서 기존 사전에서 처리하지 못한 유용한 언어 단위들을 찾아내어 어휘 데이터베이스에 반영할 수 있을 것이며, 세종전자사전의 의미부류의 보완점도 발견할 수 있을 것이다.

## 참고문헌

### 1. 論著

- 강범모·이유선·차재은(2005), “다국어어휘데이터베이스 구축방법론 연구 및 모형 개발[2],” 고려대학교 민족문화연구원.
- 김중도(2004), 은유의 세계, 한국문화사.
- 김진혜(2000), 연어연구, 한국문화사.
- 남기심(1995), “어휘 의미와 문법,” 동방학지 88, 연세대학교 국학연구원, 157-179.
- 박영순(2000), 한국어 은유 연구, 고려대학교 출판부.
- 유혜원(2004), 한국어 정보 처리의 이론과 실제, 서울: 제이앤씨.
- 최경봉(1997), “은유 표현에서의 단어의 의미론적 역할,” 한국어의미학 1, 한국어 의미학회, 171-190.
- 최경봉(2002), “은유 표현에서 어휘체계의 의미론적 역할,” 한국어학 15, 한국어학회,

285-306.

최운호(2005), 한국어 처리에서 '구류음'을 위한 명사의 특성 연구, 서울대 언어학과 박사학위논문.

한정환(1990), "국어 비유어 연구," 고려대 석사학위 논문.

Azchary J. Mason(2004), CorMet: A Computational, Corpus-Based Conventional Metaphor Extraction System, *Computational Linguistics* 30-1.

Choi, Woonho, Beom-mo Kang and Hochol Choe(2006), Research of Multi-Lingual Information Processing Methodologies and Database Development for Globalizing Korean Studies, *Proceedings of the Third International WordNet Conference*, 309-310. Global WordNet Association.

Fellbaum Christiane(1998), *WordNet: An Electronic Lexical Database*. Cambridge, Ma: The MIT Press.

Goatly, Andrew(1997), *The Language of Metaphors*, London: Routledge.

Lakoff, George and Mark Johnson(1980), *Metaphors We Live By*, Chicago and London: The University of Chicago Press.

## 2. 資料

국립국어원(1999), 표준국어대사전, 서울: 두산동아.

연세대학교언어정보개발연구원(1998), 연세 한국어사전, 서울: 두산동아.

한글학회(1989), 우리말큰사전, 서울: 어문각.

## 최운호

서울시 성북구 안암동 5가

고려대학교 민족문화연구원 중점연구소 A-204호  
136-701

전화 번호: (02) 3290-2503

전자 우편: dowland@paran.com

## 차재은

서울시 서대문구 충정로 2가 71

경기대학교 한국동양어문학부 국어국문학전공

전화 번호: (02) 390-5160

전자 우편: jecha3@hanmail.net

강범모

서울시 성북구 안암동 5가

고려대학교 문과대학 언어학과

136-701

전화 번호: (02) 3290-2173

전자 우편: bmkang@korea.ac.kr

원고접수일: 2006년 7월 10일

게재확정일: 2006년 8월 8일

K C I