

사회적 가치와 관련된 명사의 사용 양상과 네트워크 분석*

이 영 제·김 흥 규·강 범 모**

(고려대학교)

<Abstract>

Lee, Young-je, Kim, Heung-kyu and Kang, Beom-mo. 2011. A Study of the Network and the Use of Nouns Related to Social Values. *Korean Semantics 34*. This paper aims to analyze the network and use of Korean nouns related to social values while making use of co-occurring words of statistical significance. Based on network analysis, we can observe changes of the use of value nouns in recent years. We assume the complex system theory which is reflected in linguistic usage. The corpus used in this paper is Trends 21 Corpus, a large-scale Korean newspaper corpus compiled at Korea University. Cases of the usage change of words include the following. The use of 'myeongpum'(masterpiece) has been changed. It had not been long before that this word was used to mean "the higher" or "the highest". The use of 'jungsancheung'(the middle class) had been related to the use of 'buyucheung'(the highest class) in the early 2000s. Most recently we can find that the use of 'joongsancheung' is more related to the use of 'bingoncheung'(the poor class).

핵심어: 네트워크(network), 복잡계(complex system), 공기어(co-occurring word), t-점수(t-score), 가치 명사(value noun), 사회적 가치(social value), 물결21 코퍼스(Trends 21 Corpus)

* 이 연구는 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2007-361-AL0013). 자세하게 논평을 해 주신 심사자들에게 감사드린다.

** 교신저자

1. 들머리

본고에서는 사회적 가치와 관련된 의미를 가진 명사(이하 가치 명사)와 공기하는 단어들의 단어/의미 네트워크를 분석하여 언어, 사회, 문화 관련 지식을 발견하는 데 목적이 있다. 이를 위해 먼저 가치 명사와 공기하는 단어들이 서로 관련성을 맺고 있는 복잡계(complex system)의 특성을 지니고 있음을 확인한다. 단어의 공기 관계는 전체적으로는 불규칙적이고 혼돈스러운 복잡한 사용 양상을 보이지만 부분으로 세분화해 보면 규칙적인 양상을 포착할 수 있다는 점에서 복잡계라 할 수 있다.¹⁾ 부분에서의 규칙들로부터 새로운 일련의 특성들을 확인하여 언어 현상의 변화나 사회, 문화적 현상의 흐름까지 해석할 수 있다는 것이 본고에서의 입장이다.

특히 사회 구성원들이 인식하고 있는 ‘가치’의 변화는 불특정 다수의 구성원들이 매우 복잡한 요인들로부터 영향을 받아 산출하는 결과물이라는 점에서 복잡계의 특성이 잘 드러난다고 볼 수 있다. 이러한 가치의 변화는 언어 현상에 그대로 반영이 되며 본고에서는 이들을 사회적 가치와 관련된 명사(약칭 가치 명사)로 통칭한다. 이들은 언중들의 사회, 문화적 가치에 대한 인식을 보여주는 명사들이므로 인식이 변화하면 이를 반영하여 사용 양상이나 의미가 변하게 된다. 특히 어떤 단어의 사용 양상 측면에서 보면 이들이 공기하는(co-occur) 다른 단어들과의 쓰임에 변화가 나타난다. 만약 이들이 특정한 패턴으로 변화하는 규칙이 발견되면 그것은 일정한 규칙이나 패턴이 창발(emergence)되는 현상을 보여주는 것이라 할 수 있다.

본고에서는 우선 연구 대상인 가치 명사를 의미적 기준과 연구 방법론과

1) 복잡계의 네트워크 분석 이론에서는 이를 작은 세상의 특성(small world property)라고 명명한다(윤영수·채승병, 2005). 우리가 살고 있는 사회는 다양한 요인들이 반영되어 변화하는 복잡계의 특성을 지닌다. 이 때 사회의 변화는 전체적으로는 혼란스러운 양상처럼 보이지만 우리는 세부적으로는 일정한 규칙성을 띠는 변화들을 포착할 수 있다. 즉 불안정하고 불확정적인 혼란 속에서도 특정한 현상과 현상 간의 관계를 설명할 수 있는 패턴을 찾아내고 이를 규칙화 할 수 있다는 것이다. 윤영수 외(2006)에 의하면 이를 설명하기 위한 복잡계 이론은 프랙탈(fractal) 이론, 혼돈(Chaos) 이론, 네트워크 분석(Network Analysis) 이론 등이 있다. 언어 현상의 복잡계적 특성에 대해서는 고창수(2007), 채현식(2007), 강범모(2010)에서 이미 자세히 다루었으며 2.2에서 간략히 상술한다.

의 적합성 등을 기준으로 일부만을 선별한다. 그리고 약 3억 어절 이상의 대규모 신문 자료인 [물결21]의 코퍼스를 대상으로²⁾ 가치 명사와 유의미한 공기 관계를 이루는 일반명사들을 추출하여 중심어(keyword)와 공기어(co-occurring word)³⁾ 간의 네트워크가 시간의 흐름에 따라 변화하는 양상을 분석한다. [물결21] 코퍼스의 특성상 2000년부터 2008년까지 9년치 신문 기사를 대상으로 하며 이들의 통시적인 변화에 초점을 맞춘다. 그리고 중심어와 공기어들의 사용 양상의 변화가 단지 우연한 것인지, 실제 유의미한 것인지를 검증하기 위해 통계 검증 기법을 적용한다.

이 연구는 대규모 신문 코퍼스를 기반으로 하여 장기간의 언어·사회·문화의 변화 흐름을 분석하는 연구의 일환이다. 언어 네트워크 분석 연구는 이미 강범모(2010)에서 구체적으로 연구 방법을 제시한 바 있다. 강범모(2010)에서는 공기하는 명사들을 관련어로 하여 의미/개념의 연관성 네트워크를 구성하기 위한 방법론을 제시하였으며 본고 역시 동일한 관점을 견지한다. 그리고 김일환 외(2010b)에서는 감정명사를 대상으로 네트워크를 분석하였다.⁴⁾

이처럼 네트워크 분석은 다양한 영역의 단어까지 확장이 가능하며 또한 신어처럼 시간의 흐름에 따라 유의미한 차이를 보이는 범주에도 적용할 수 있다. 따라서 공기어의 네트워크 분석은 앞으로 더욱 다양한 분야로 확대해 나아가는 것이 가능하다.

-
- 2) [물결21] 코퍼스는 고려대학교 HK사업단에서 진행하고 있는 물결21 사업의 결과물로 이에 대해서는 2.2절에서 상술한다. 김홍규 외(2010)에서는 이 사업의 기반이 되는 자료에 대한 설명부터 태깅(tagging) 방식과 세부 지침까지 덧붙이고 있다. 태깅 방식은 21세기 세종 국어 기초자료구축의 형태 분석 방식과 유사하다.
- 3) 본고에서의 공기어(co-occurring word)란 문단 내에서 공기 관계를 이루는 단어를 의미하며 연구 대상은 명사로 한정한다. 이는 일종의 언어(collocation)로 볼 수도 있으나 본고에서의 공기어는 문단 내에서 함께 공기하는 단어를 총칭하기 때문에 선형적 연속성을 보이는 문법적 언어 부류와는 무관하며 따라서 이는 좁은 의미의 언어와는 다르다. 이 연구에서는 하나의 사건 기술이 하나의 생각 단위의 기술이라는 점을 고려하여 문단 단위를 사용하였다. 강범모(2010)에서는 이처럼 문단 내에서 공기하는 단어를 관련어(related word)로 명명하고 언어와 연관된 개념임을 명시한 바 있다.
- 4) 이밖에 인접 분야에서는 주로 사회 연결망을 분석한 연구가 활발한데 김용학 외 (2006) 등이 있으며 최근에는 의미의 연결망(semantic network)에 대한 연구 등으로 관심 영역이 확대되고 있다. 그리고 내용 분석(content-analysis) 연구도 어휘 개념의 역사적 변화를 분석한다는 점에서 공통의 관심사에 해당한다고 볼 수 있다.

2. 연구 대상과 분석 방법론

2.1. 사회적 가치와 관련된 명사의 정의

사회적 가치와 관련된 명사란 의미적으로 사회적 가치나 평가를 반영하고 있는 명사들을 포괄적으로 가리킨다. 여기서의 가치란 긍정 또는 부정의 가치를 의미할 수도 있고, 정치·경제·사회·문화 등 다양한 분야에서 평가되는 가치들을 의미하기도 한다.⁵⁾ 본고에서는 이들이 문단 내에서 다른 단어와 공기하는 사용 양상을 통해 해당 문맥을 추론하거나 단어의 의미 변화가 창발되는 현상을 확인하고자 한다.⁶⁾ 그리고 가치 명사에 대한 연구는 언어의 변화뿐만 아니라 관련된 사회, 문화적 변화의 흐름을 분석하는 데에도 객관적이고 과학적인 근거를 제공하게 될 것이다.

우선 가치 명사의 존재론적 정의를 살펴보자. ‘가치’를 평가하는 의미를 담고 있는 단어들은 먼저 개념적으로 실체와 양식의 구분에서 양식에 해당한다. 가치의 평가는 구체적인 지시 대상을 갖기보다 추상적인 양식에 관련한 것이기 때문이다. 양식은 다시 사태와 관계로 구분될 수 있는데 전자가 주로 가치에 관한 의미를 포괄하고 있으며 사태는 다시 사건과 상태로 구분된다 (최경봉, 1998:75-77).

- (1) 양식 - 사태 - 사건
 - 상태 - 현상
 - 추상

-
- 5) 사회적 가치와 관련된 명사 또는 가치 명사라는 개념에 대해 기존에 언어학적으로 정의된 바는 없으며 본고에서도 이를 상식적인 수준에서의 개념으로 사용한다. 익명의 심사자께서는 이러한 점들을 보다 엄밀히 정의해야 할 것을 강조하면서 이를 어휘가 가리키는 개념의 사회/역사적인 함의로 볼 것을 제안한 바 있다. 본고 역시 단어를 통해 개념을 연구하는 시도로서 심사자의 지적에 동의하는 바이나 다만 단어와 개념을 동일시하지는 않고 서로 대응하는 것으로 간주한다. 나인호(2010:279-280)에서 언급한 대로 개념이 해석의 대상이라는 점에는 공감하나 결국은 이 또한 정의의 문제로부터 자유로울 수 없다는 점에서 단어와 개념이 서로 대응하는 것으로 보아야 할 것이다.
- 6) 문장보다 넓은 범위의 관련어 추출에 문단이 유용한 이유에 대해서는 강범모(2010:4)에서 설명한 바 있다. 문장은 관련어의 추출이 제한된 반면, 하나의 화제와 내용적 응집성을 가지고 전개되는 문단에는 보다 많은 수의 관련된 단어들을 추출할 수 있어 보다 유리하다. 이는 또한 단어들에 지시하는 사물, 나아가 개념들의 네트워크를 구축하는 것이라 할 수 있다.

- (2) a. **양식-사태-사건-자동/타동**: 각성, 감동, 느낌, 뜻, 발달, 발전, 부패, 사랑, 승리, 전쟁, 실수, 유학, 통일, 투쟁, 표현, 향상, 확인, ...
- b. **양식-사태-상태-추상**: 가치, 간계, 경륜, 독특성, 명예, 목숨, 생명, 세월, 수단, 역사, 운명, 일반, 진심, 청춘, 특권, 현실, 혼, 힘, ...
- c. **양식-사태-상태-현상**: 강성, 교만, 다행, 영구, 웅장, 위대, 자유, 지혜, 진실, 탁월, 투명, 평화, 풍부, 행복, ...

본고에서는 (2)의 분류 중 ‘양식-사태-사건-자동/타동’, ‘양식-사태-상태-추상’, ‘양식-사태-상태-현상’에 해당하는 명사 가운데 사회적, 문화적 가치를 평가하는 의미를 담고 있거나, 긍정과 부정 등 이분법적 대립을 이루는 부류의 단어 중에서 상위 빈도의 일반명사부터 차례로 검토하여 연구 대상을 선정하였다. 이 때 연구의 목적이 존재론적 의미 분류에 해당하는 모든 단어들의 양상을 확인하고자 하는 것이 아니고 특징적인 변화 양상을 포착하고자 하는 것이기 때문에 연구 대상은 정치, 경제, 사회, 문화 등의 요인들과 관련 되면서 대표성⁷⁾을 띠는 26개의 가치 명사만으로 제한한다.⁸⁾

이러한 의미론적 기준을 가지고 [물결21] 코퍼스에서 추출한 일반명사 목록을 검토하여 연구 대상 목록을 선정하였다. 기준에 해당하면서 상위 빈도로 나타나거나, 가치 명사의 의미를 고려할 때 유의미한 관련성이 예상되는 단어들을 선별하여 다음과 같이 연구 대상을 추출하였다. 이들은 네트워크 분석에서 중심어에 해당한다.

7) 이 때의 대표성은 주로 빈도를 기준으로 하나, 주목할 만한 사회적 가치를 담고 있는 단어도 포함하였다. 또한 형태소 분석의 한계로 인하여 동형이의어를 구분할 수 없으므로 중의성이 있는 단어들은 기준에 적합하더라도 연구 대상에서 제외하였다.

8) 익명의 심사자께서는 개인적 가치로 볼 수도 있는 ‘행복, 불행’ 등과 경제적 계층을 나타내는 ‘부유층, 중산층’ 등, 그리고 단어 의미 성분에 가치의 평가가 반영되긴 하지만 실제 사회적 가치를 반영하고 있는지는 확인할 수 없는 ‘명품’ 등을 동계로 묶어 ‘가치 명사’라는 정의를 내린 본고의 논의에 대해 의문을 제기하였다. 본고에서의 가치 명사는 논의의 성격상 ‘일정한 가치를 개념적으로 가지고 있는 단어’와 ‘일정한 가치를 내포하고 있는 단어’를 모두 다루는 포괄적인 개념으로 수용하였다. 그리고 ‘행복’ 등의 개인적 가치는 결국 전체 사회적 가치와 무관할 수 없을 것이고, ‘부유층’을 비롯한 몇 가지 중심어들은 어휘의미적인 경제적 가치 뿐만 아니라 계층과 관련된 여러 가지 사회적 가치까지 포함하고 있다고 볼 수 있다. ‘명품’의 경우도 심사자께서 지적하신 대로 가치의 평가를 담고 있는 것이 기본적인 의미이지만 실제 사용 양상이 사회적 가치를 반영하고 있다는 점에서 용어 개념의 포괄적인 적용이란 전체 하에 사회적 가치와 관련된 중심어로 다룰 수 있을 것이다.

b. 중심어가 3개인 부류

중심어	절대빈도	중심어	절대빈도	중심어	절대빈도
갈등	68,182	개혁	151,355	성장	116,486
통합	97,219	안정	58,301	복지	71,834
화해	23,553	변화	139,634	분배	11,306

c. 중심어가 2개 이하인 부류

중심어	절대빈도	중심어	절대빈도	단어	절대빈도
성공	127,553	위기	139,634	평등	13,005
실패	65,018	기회	100,945	차별	27,659
좌파	7,887	명품	18,637		
우파	5,497				

이들을 중심어의 개수에 따라 3가지 부류로 구분하였다. 이상의 26개 단어를 대상으로 본고에서는 각각의 단어의 사용 양상이 연도별로 변화하는 현상이나 여러 공기하는 단어의 군집이 특징적인 변화의 패턴을 보이는 현상들을 네트워크의 관점에서 분석하고 이를 통해 여러 가지 언어, 사회, 문화적 흐름의 변화를 살펴볼 것이다.

2.2. 네트워크의 분석

2.2.1. 복잡계로서의 언어와 네트워크

언어가 복잡계의 특성을 보인다는 점은 이미 Bybee(1995) 등에서 언급한 바 있다. Beckner et al(2009)에서는 구체적으로 언어의 복잡계적 특성을 다음과 같이 6가지 속성으로 제시하였다.

- (5) a. 배분된 통제와 집합적 창발
 - b. 본질적인 다양성
 - c. 영속적인 역동성
 - d. 요인들의 확장과 경쟁을 통한 적응

- e. 비선형성과 국면의 전환
- f. 연결망 구조에 대한 의존성과 민감성
- g. 변화는 국부적이다.

언어의 속성에 따라 규정한 이러한 원리들을 바탕으로 고창수(2007), 채현식(2007) 등에서는 빈도 효과나 어휘부의 자기조직화 원리 같은 분석의 틀에 적용해 국어에서도 복잡계적 특성이 나타난다고 보았다. 고창수(2007)에서는 언어 연구에서 비선형적이고 다분화된 방법론이 필요함을 주장하면서 정상과학으로부터 떨어진 최소주의 이론을 대체할 수 있는 대안을 복잡계 이론으로부터 찾고자 하였다. 그는 언어학이 정상과학이라는 전제하에 음운론적 생성과 소멸이나 단어의 창발 등이 모두 복잡계적 속성을 보여주는 것이라 정의하면서 다음과 같은 복잡계의 특성을 제시하고 있다(고창수 2007:13).

(6) 복잡계의 특성

- | | |
|------------------|--------------|
| a. 비선형적 관계, | b. 회귀적 관계, |
| c. 열린 체계, | d. 통시적 체계, |
| e. 독립된 체계, | f. 모호한 경계, |
| g. 다양하고 역동적인 관계, | h. 창발 현상의 촉발 |

채현식(2007)에서도 복잡계로서의 어휘부를 설정하고 단어들이나 틀의 자기조직화(self-organization) 원리를 제시하고 있다. 채현식(2007)에서는 생성형 태론의 규칙 중심 접근법이 지닌 한계(심리적 실재성 부족, 규칙의 과잉생성과 과잉 분석)를 지적하면서 어휘부도 복잡계라고 보면 어휘부를 설명할 때 창발이나 어휘부 규칙성 같은 것들을 잘 설명할 수 있다고 주장한다. 그리고 언어학의 사용-기반(usage-based) 모델을 근거로 해서 어휘부의 규칙성 역시 저장된 단어들로부터 독립해서 존재하지 않고 고도로 조직화된 어휘부에 저장된다고 본다. 그리하여 어휘부를 복잡계로 보았을 때 어휘부가 지닌 두 가지 이질적 성격, 곧 단어들의 저장소로서의 어휘부와 규칙성을 지닌 구조로서의 어휘부를 통일된 관점에서 설명할 수 있다고 주장한다.

본고에서는 이러한 언어 현상의 복잡계적 특성을 설명하기 위해 빈도 효과

과(frequency effect)에 기반한 공기 관계의 네트워크 분석 방법을 적용한다.¹³⁾ 즉, 자주 함께 출현하는 단어들이 군집화하여 이들이 하나의 네트워크를 이루고 있는 것을 확인하고 이 네트워크 안에서 새로운 정보(information)를 찾는 것이다. 이러한 네트워크는 고정된 것이 아닌 동역학적 네트워크이므로 시간선상에서의 변화 흐름 속에 놓이게 된다. 여기서는 표면적인 네트워크의 변화 양상을 통해 변화가 일어나는 동인이나 결과 현상들을 해석하는 데 초점을 둔다. 그리고 네트워크 분석을 통해 언어 내적인 지식(knowledge)뿐만 아니라 언어 외적인 사회, 문화적 지식까지 관찰해 보고자 한다.

2.2.2. 네트워크의 구성

[물결 21]의 대규모 신문 코퍼스는 주요 4대 일간지의 2000년 이후 신문기사 텍스트 전체를 가공하여 언어 자원 데이터베이스로 구축한 것이다(김홍규 외 2010:1).¹⁴⁾ 본고에서는 이 중 2000년부터 2008년까지 9개년에 걸친 4대 일간지의 신문 텍스트 전체를 대상으로 복잡계 이론을 모델로 한 네트워크 분석을 시도한다. 이 때 전체 어절 수는 약 348,261,978 어절이고, 기사 수는 1,763,581 개이다.¹⁵⁾

신문 기사 텍스트로부터 정밀한 어휘 사용 양상을 밝히기 위해서는 원전 기사에 언어학적 정보를 주석(tagging)하는 작업을 수행할 필요가 있다(김홍규 외 2010). 본고에서의 연구 대상 코퍼스도 형태소 분석기를 이용하여 품사 정보를 부착한 코퍼스(tagged-corpus)이다. 품사 표지 및 부착 기준, 세부 지침

13)빈도 효과라는 용어는 원래 생물학, 심리학 등에서 사용되던 것인데, 언어 현상을 설명하기 위해 Bybee & Hopper(2001) 등에서 도입한 바 있다. 복잡계에서의 창발(emergence)이 언어 현상에서는 빈도의 증가로부터 비롯되며 이것이 빈도 효과와 관련된다. 구체적으로는 공기 관계의 빈도가 일정 수준을 넘어서면 빈도 효과로 인해 새로운 현상이나 규칙이 창발되는 것을 말한다.

14)[물결 21]의 구축 대상인 4대 일간지는 동아, 조선, 중앙, 한겨레신문이다. [물결 21]에서는 다양한 이론 모델과 방법론으로 21세기의 언어·문화·사회적 특성과 그 변화를 분석하여 인문 사회과학 연구 전반에 걸친 토대 연구를 이루고자 하며 본고 역시 그 일환으로서의 연구이다.

15)전체 코퍼스는 광고, 단신 등을 제외한 기사, 사설, 칼럼, 독자 투고 등의 신문 기사를 대상으로 한다. 단 한겨레신문의 경우에만 2007년 신문 기사까지를 대상으로 하였다. 그리고 이 중 명사(NNG)에 해당하는 어절은 총 223,794,143 어절로 전체에서 가장 많은 비중을 차지한다. 자세한 내용은 김홍규 외(2010:83)을 참고.

등은 21세기 세종계획 국어 기초자료 구축의 성과물인 형태 분석 코퍼스를 기반으로 하여 [물결 21]에서 수정, 보완한 SJ-RIKS 코퍼스의 방법을 따른다(김일환 외 2010a).

다음 (7)은 [물결 21]의 신문 코퍼스에서 ‘중산층’과 주로 공기하는 단어들을 문단 단위로 추출한 예이다.

(7) 한 문단 내에서 ‘중산층’과 주로 공기하는 명사의 예¹⁶⁾

정 후보는 또 “근로소득세에 물가 연동제를 도입, 2000만원 이하 저소득 근로자에게는 물가인상을 반영, 사실상 감면에 준하도록 세금을 줄이겠다”고 말했다. 정 후보는 또 “근로소득세 부과 구간 ‘8000만원 이상’도 ‘8000만원 이상 1억3000만원 미만’과 ‘1억 3000만원 이상’으로 세분화해 <중산층>과 고소득층의 세금 부담을 차등화하겠다”고 말했다. 자영업자의 소득세도 “장사가 안되면 세금을 탄력적으로 적용하겠다”고 했다.

2.2.3. 네트워크의 통계 검증 기법

본고에서는 빈도 효과를 입증하기 위한 통계 검증 기법으로 t-점수를 이용한 t-검증(t-test) 기법을 적용한다.¹⁷⁾ 아래는 t-점수를 구하는 식을 도시한 것이다(Barnbrook 1996, 강범모 2010).

$$(8) \text{ t-점수} = \frac{O - E}{\sqrt{O}} \quad (O \text{는 관찰값, } E \text{는 기댓값})$$

16) 밑줄 친 명사 중 Font의 크기가 가장 큰 것은 ‘중산층’과 문단 내에서 더 자주 공기하는 명사이다. 즉, Font의 크기가 클수록 자주 공기하는 명사이다.

17) t-검증이 단어와 단어 간의 공기 관계 분석에서 유의미성을 입증하는 데 적합한 이유에 대해서는 홍중선 외(2001), 박병선(2005) 등에서 자세히 언급하고 있다. Manning & Schutze(1999)에 따르면 이밖에 텍스트의 통계적 검증을 위한 기법으로 z-점수나 카이제곱(χ^2) 검증, MI(mutual information) 등이 있으나 홍중선 외(2000)에서는 두 단어가 공기하는 정도를 수치화하는 데 가장 적합한 검증 기법은 결과적으로 t-점수라는 결론을 내리고 있다. 본고에서는 이를 토대로 가치 명사와 공기 관계를 이루는 단어들의 연관 관계의 정도를 t-점수에 따라 파악한다.

(8)의 수식에서는 관찰값(observed frequency)이 크고 기댓값이 작을수록 t-점수(score)는 커지며 더 유의미한 값이 된다. 즉, 두 단어가 공기하는 횟수가 많으면서도 공기어가 전체 코퍼스에서 출현하는 빈도가 적다면 더 유의미한 공기 관계라는 것을 나타낸다. 어떤 단어에 대한 관찰값은 연구 대상인 ‘중산층’과 같은 중심어와 한 문단 내에서 공기하는 횟수를 빈도로 수치화한 것이다.¹⁸⁾

그러나 본고에서는 [물결21] 코퍼스의 특성상 단어 간의 공기 관계가 t-분포를 따르고 자유도가 무한대일 때를 가정하여 기준값을 상정한다.¹⁹⁾ Manning & Schutze(1999:163)에서는 토큰수 1,430만 개의 코퍼스에서 ‘new companies’가 공기하는 것이 통계적으로 유의미한 것인가를 t-검증을 통해 보여주고 있는데, 이 때 자유도가 30이상일 때는 공기 관계의 분포가 대체로 정규분포를 따른다고 보았다. 따라서 t-분포에서 자유도가 무한대일 때 신뢰구간을 99.5%로 설정하면 t-점수로 2.576을 넘어야 영가설을 기각할 수 있는 점수가 된다.²⁰⁾

아래 (8)의 표는 (7)에서 제시한 수식에 따라 t-점수를 계산한 결과값을 도표로 보인 것이다. 본고에서는 네트워크상의 공기어를 t-점수를 기준으로 상위 50~100개를 선정하거나, 표준정규분포를 따를 시 99.5%의 신뢰도로 유의미한 t-점수 2.576이상에 해당하는 공기어를 순위에 따라 적용해 네트워크상에 구현하였다.

18) 이 때 단어의 공기 관계를 계량화하기 위한 단위는 문단을 기준으로 하였다. 이를 텍스트 단위로 확대할 경우 여기서와는 다른 네트워크가 만들어져 새로운 지식(knowledge)의 해석이 가능할 것으로 예상된다.

19) 자유도가 무한대인 t-분포는 대체로 표준정규분포와 같다.

20) 인문사회과학 분야에서는 일반적으로 통계적으로 유의미한 것인가의 여부를 95%(t-점수: 1.645)를 기준으로 삼지만 여기서는 대규모 코퍼스를 대상으로 한 분석이라는 특성상 Manning, C & H. Schutze(1999)처럼 99.5%(t-점수: 2.576)에 해당하는 구간을 신뢰 구간으로 한다.

(9) '갈등'과 공기하는 명사(t-점수를 기준으로 내림차순으로 정렬하여 20개만 제시)

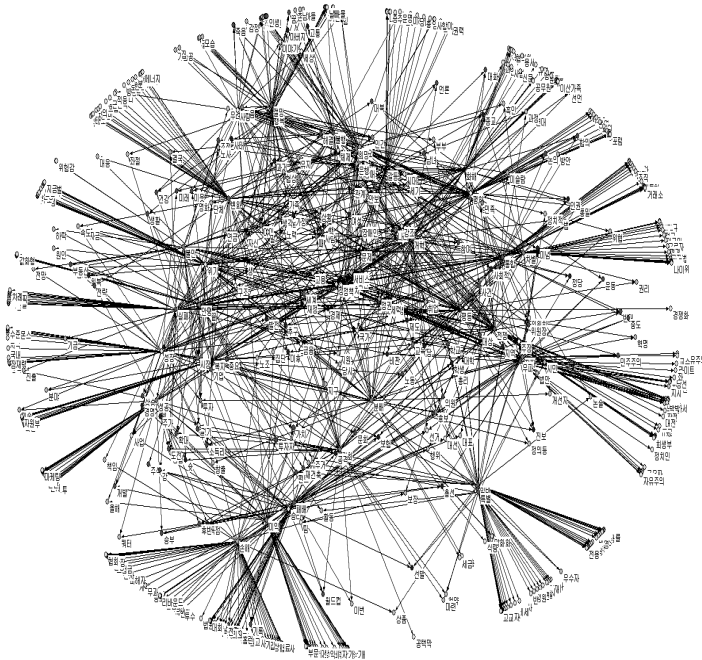
순위	중심어	공기어	t-점수
1	갈등	문제	95.034
2		사회	92.15
3		해결	66.475
4		관계	64.261
5		정부	63.915
6		사이	60.049
7		대립	59.014
8		노사	58.445
9		대통령	57.523
10		정치	55.055
11		사회적	54.251
12		해소	52.597
13		정책	52.045
14		내부	49.741
15		분열	46.304
16		세력	46.248
17		국민	45.893
18		이념	45.086
19		통합	44.875
20		노조	44.538

2.2.4. 네트워크의 실제

이 절에서는 네트워크의 실제 모습을 시각화하여 살펴보고 네트워크의 구성요소들이 시각화된 그림에 어떻게 반영되어 있는지 확인할 것이다. 네트워크 분석을 위해서는 통계적 분석과 시각화를 동시에 수행하는 도구가 필요한데 이 때 유용한 프로그램이 Pajek이다. Pajek은 김용학(2003), Nooy et al(2005), 강범모(2010) 등에서 자세히 소개하고 있는데, 많은 수의 결점을 가진 연결망 자료를 축소하여 일정한 규칙성을 찾아내 그래픽으로 표현하는 데

매우 탁월한 기능을 가지고 있다(김용학, 2003:178).

Pajek에서는 *.net이라는 확장자 파일을 이용해 네트워크 분석을 통계적으로 처리하며, 연결망의 결속 분석을 위해 연결 정도(degree), 밀도(density), 회귀(loop), 연결 정도 중앙성(centrality), 중심화지수(centralization) 등의 값을 수치화해서 보여준다. 이 중에서 우리가 필요로 하는 값은 주로 밀도(density)와 중심화지수(centralization)로 이 값들은 집단의 크기에 반비례한다.21) 아래 <그림 1>은 가치 명사 30개 전체와 그 공기어를 각 50개를 네트워크로 구현한 것이다.



<그림 1> 가치 명사 30개의 전체 네트워크

21) 연결 정도(degree)는 한 결점(node)이 관계를 맺고 있는 다른 결점의 숫자이며 연결 관계의 흐름에서 핵심적인 역할을 한다. 그리고 연결망 밀도(density)는 가능한 총관계 수 중에서 실제로 맺어진 관계 수의 비율, 즉 $1/(n(n-1)/2)$ 로 정의된다. 밀도는 일반적으로 집단의 크기와 반비례한다(김용학, 2003:38, 94). Pajek에서 사용하는 용어와 프로그램에 대한 자세한 설명은 Pajek 프로그램 홈페이지를 참고. (<http://vlado.fmf.uni-lj.si/pub/networks/pajeks>)

이 그림의 결점 간 연결 관계를 보여주는 밀도(density)는 0.002이고, 평균 연결 정도(average degree)는 4.244, 중심화지수(centralization)는 0.037이다.²²⁾ 이 그림에서는 네트워크의 중앙일수록 더 복잡한 양상을 띠는데 우리는 매우 복잡해 보이는 이 그림 안에서도 일정한 규칙과 연결 관계를 확인할 수 있다.

우선 화살표가 출발하는 결점(node)은 중심어이며 이 결점들은 본고에서의 연구 대상인 가치 명사들이다. 그리고 화살표가 도달하는 결점은 공기어이며, 이 때 공기어는 하나 이상의 중심어로부터 연결될 수 있다. 전체 그림은 매우 복잡하고 가시적으로 구분하기 어려운 연결 관계처럼 보이지만 이와 같은 구성을 이해하면 특징적인 규칙과 네트워크를 확인할 수 있다.

다음으로 아래 <그림 2>는 가치 명사 중 ‘행복’과 ‘불행’ 두 단어만을 중심으로 해서 각 중심어별로 100개의 공기어를 추출해서 네트워크로 표현한 것이다. 이 때 ‘행복’과만 공기하는 단어는 45개이고, ‘불행’과만 공기하는 단어도 45개이며, 두 중심어 모두와 공기하는 단어는 55개이다. 두 중심어 간에도 서로 공기하는 양상이 나타난다. 여기서 두 중심어 모두와 공기하는 단어는 그 링크(link)의 수에 따라 ‘2-링크’ 관계로 지칭한다.²³⁾

<그림 2>의 네트워크의 밀도는 0.009, 평균 연결 정도는 2.739, 중심화지수는 0.343으로 전체 가치 명사의 네트워크에 비해 중심화지수에서 높은 값을 보인다.²⁴⁾ 이 네트워크에서 행복과만 공기하는 명사들은 ‘자녀, 웃음, 건강, 기쁨, 부부’ 등 주로 긍정적 의미의 단어들이 나타난다. 반면 불행과만 공기하는 단어들은 ‘절망, 자살, 폭력, 상처, 비극’ 등 부정적 의미의 단어들이 주로 나타난다.

또한 <그림 1, 2>에서는 결점A와 결점B의 거리는 밀도나 중심화지수와 무관하며 연결 관계만 보여줄 뿐 상대적으로 밀접한 관계인지 여부를 보여주지 않는다. 즉, 연결 관계를 보여주는 링크에는 t-점수의 높고 낮음은 반영되지 않으며 결점 간의 연결 관계와 방향성만을 보여준다. 한편, 링크의 값이 연결선의 길이에 반영되도록 선택사항을 직접 결정할 수도 있다.

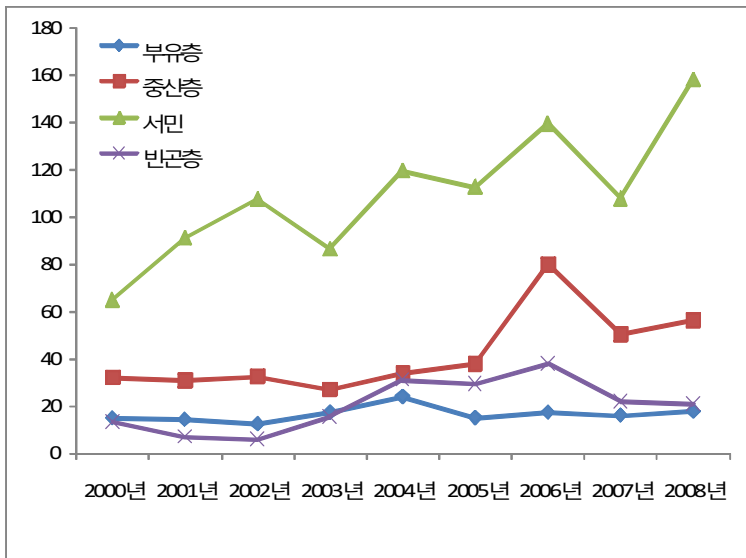
22)Pajek에서 이 값을 구하는 방식은 김용학(2003:183-189)을 참고.

23)링크(link)는 결점과 결점을 연결하는 선을 말한다.

24)밀도는 집단의 크기에 반비례한다.

신문기사 전체를 대상으로 한 중심어와 공기어의 네트워크이다. 그리고 이를 통해 중심어들 간의 관계도 구명해 보고자 한다. 복수의 중심어 네트워크는 우선 4개의 중심어 네트워크를 먼저 보이고 이어서 3개와 2개의 중심어로 이루어진 네트워크를 각각 제시한다.

먼저 소득이나 사회적 계층을 기준으로 대립 관계를 보이는 4개의 중심어 ‘부유층 : 중산층 : 서민 : 빈곤층’ 각각의 상대 빈도가 변화하는 추이를 보면 다음 <그림 3>과 같다.²⁵⁾



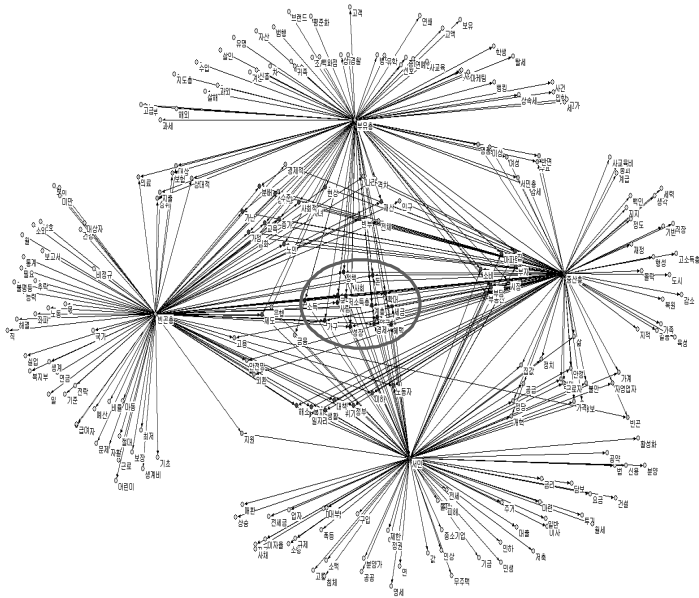
<그림 3> ‘부유층 : 중산층 : 서민 : 빈곤층’ 네 단어의 상대빈도

<그림 3>에서 서민과 중산층은 2003년 이후 상대빈도가 평균적으로 높아

25) 중심어 ‘서민’은 다른 중심어와의 관련성을 고려한다면 ‘서민층’이 되어야 할 것이다. 그러나 ‘부유, 중산, 빈곤’ 등이 계층의 의미를 지시하지 못하는데 비해 ‘서민’은 서민층을 지시하는 의미를 포괄하고 있고 대표적으로 쓰이기 때문에 실제 쓰임을 기준으로 대립 관계를 고려하면 ‘서민층’보다는 ‘서민’으로 중심어를 삼는 것이 더 적절할 것이라고 판단하였다. 또한 ‘서민층’의 실제 출현 빈도는 매우 낮아서 ‘중산층’ 등의 중심어와 관련해서는 ‘서민’이 더 많이 활용됨을 알 수 있다.

지는 것을 확인할 수 있다. 이 그래프로만 보면 명사 ‘중산층’이나 ‘서민’의 사용 양상이 확대된다는 추정을 할 수 있지만 이것이 중산층에 해당하는 이들이 늘어나서 빈도가 증가된 것인지, 여러 가지 문제들이 중산층 계층에 집중되어서 빈도가 증가되는 것인지 확인할 수 없다. 다만 중산층에 대한 사회적 관심이 높아진 것만은 분명하다. 네트워크로 변화양상을 살펴보면 그 원인과 결과를 객관적이고 합리적으로 설명할 수 있는 근거들을 찾을 수 있다.

그러면 다음의 ‘부유층 : 중산층 : 서민 : 빈곤층’과 공기하는 명사의 네트워크를 살펴보자.²⁶⁾



<그림 4> ‘부유층 : 중산층 : 서민 : 빈곤층’의 네트워크

<그림 4>에서 가운데 모여 있는 단어들이 4-링크 관계의 단어이다. 이들은 주로 ‘돈, 세금, 경제, 성장, 소득, 계층’ 등의 경제 용어들이다.²⁷⁾ 네 가지 중

26) 이 네트워크 밀도는 0.0066368, 중심화지수는 0.20524이다. 중심이별 공기어는 시각화 도구에 서의 제약으로 인해 t-점수를 기준으로 상위 100위로 한정하였다.

심어 중에서 ‘중산층’의 쓰임이 비교적 특이한데 ‘중산층’과 다른 세 중심어와의 관계를 보면 가장 복잡한 네트워크를 이루고 있다. 즉 ‘부유층 : 중산층 : 빈곤층’, 또는 ‘부유층 : 중산층 : 서민’, 그리고 ‘중산층 : 서민 : 빈곤층’의 세 단어와 3-링크 관계를 이루는 공기어는 다양하게 나타나는 반면 ‘부유층 : 서민 : 빈곤층’ 세 단어 간에 3-링크 관계를 이루는 공기어는 ‘제도’뿐이다. 또한 ‘부유층 : 서민’ 간에 2-링크 관계를 이루는 공기어도 ‘금융’, ‘은행’ 두 단어뿐이어서 상대적으로 ‘부유층 : 서민’ 두 단어만을 고려하면 링크가 거의 만들어지지 않는 것을 확인할 수 있다.²⁸⁾

이러한 결과들을 통해 알 수 있는 것은 ‘중산층’의 공기어와 다른 중심어의 공기어가 일치하는 비율이 상대적으로 높다는 점이다.²⁹⁾ 따라서 사회적 계층을 주제로 다루는 신문 기사는 전반적으로 다른 계층에 비해 중산층을 많이 포함하게 된다. 이것이 실제 중산층이 더 두터워졌다가거나 하는 사회적 현상을 그대로 반영한 것이라고 볼 수는 없다. 이를 관련된 언어 표현의 증가로 한정해야 한다는 점은 분명하지만 그렇더라도 ‘중산층’ 관련 언어 표현들이 증가했다는 것이 사회 구성원들의 관심도가 증가했음을 의미하는 것은 명백하다.³⁰⁾

한편 다른 의미 부류의 가치 명사들을 네트워크로 분석한 결과를 살펴보자(네트워크 그림은 생략). ‘행복 : 불행 : 희망 : 절망’의 네트워크는 밀도 0.015, 평균 연결 정도 3.478, 중심화지수는 0.02로 나타났다. 평균적인 네트워크

27) 4-링크 관계는 중심어 4개와 링크가 연결된 관계를 의미한다. 3-링크, 2-링크, 1-링크는 각각 중심어와 링크가 연결된 개수를 기준으로 설정된다.

28) 이 네트워크는 하나의 중심어가 시간의 흐름에 따라 공기어 네트워크를 어떻게 변화시켜 가는지 확인하기 위한 것이다. 이 절에서는 2절에서 제시한 26개의 가치 명사 중에서 특징적인 흐름이 나타나는 일부의 네트워크를 통해 이러한 정보들을 분석한다. 우선 하나의 중심어를 2000년부터 2008년까지 연도별로 9개의 중심어로 나누고 각 연도별로 공기어하는 명사 중 상위 50개를 공기어로 제시하였다.

29) <그림 4>에 나타난 나머지 공기어들의 경우 일반적인 경향을 따르고 있어서 여기서 더 이상 언급하지 않는다. 예를 들어 ‘부유층 : 중산층’ 두 중심어의 공기어는 ‘명품, 여성, 감세’ 등이며, ‘고용, 지원, 삶, 근로자’와 같은 단어와는 공기어하지 않고 있다. 다만 상위 100위에 해당하는 공기어만을 대상으로 한 연구이므로 한계가 있을 수 있다는 점을 밝혀두고자 한다.

30) 또한 3.2절에서는 연도별로도 이러한 연결 관계가 연도에 따라 다르게 나타난다는 점을 확인한다. 이러한 정보는 단순히 절대비율로 나타나는 공기 관계만을 가지고는 알 수 없는 것들이며 네트워크 분석의 과정을 통해서만 확인할 수 있다.

크에서는 공기하는 단어를 더 많이 포함시킬수록(집단이 커질수록) 밀도와 중심화지수가 작아지는 것이 일반적이며 이 네트워크 역시 이러한 경향을 따르고 있다.

이들 네 중심어는 인간의 삶에서 중요한 지표가 되는 단어들이다. 여기서 중심어 모두와 공기하는 4-링크의 공기어로 ‘인간, 사람, 삶, 자신, 아이, 마음’ 등이 나타나고 있는데 이는 ‘행복, 불행, 희망, 절망’의 중심어들이 인간, 사람, 삶, 자신 등과 관련된 개념을 주의미로 하고 있음을 파악할 수 있다.

다음으로 3개의 중심어 ‘갈등 : 통합 : 화해’의 네트워크에서는 각 중심어 중 하나만 공기하는 1-링크 단어의 비율이 유사한 비율로 나타났다.³¹⁾ 하지만 2-링크 관계로 공기하는 명사는 ‘갈등 : 화해’ 또는 ‘갈등 : 통합’의 두 단어 간의 네트워크에서 주로 많이 나타났다. 즉, ‘갈등 : 화해’에서는 주로 ‘정권’, ‘대북’, ‘종교’, ‘평화’, ‘화합’, ‘외교’ 등의 단어가 공기하는데 이는 2002년의 정권 교체기나 북한과 관련된 기사 또는 외교, 역사 문제와 관련하여 쓰인 것이다. 반면 ‘갈등 : 통합’에서는 ‘지역’, ‘당’, ‘개혁’, ‘조정’, ‘발전’ 등의 단어가 상위에 제시되는데, 주로 국내의 정치적 이슈에 해당하는 단어들이라는 점을 확인할 수 있다.

‘개혁 : 변화 : 안정’의 네트워크는 중심어 3개와 모두 공기하는 3-링크 관계의 단어들에서 특징적인 부분이 포착되었다. 세 중심어 모두와 공기하는 단어는 ‘정치, 사회, 경제, 정부, 국민’ 등 사회의 각 분야의 주체들이 상위로 나타난다. 한편 개혁과 변화의 2-링크 관계를 각각 보면 ‘조직, 규제, 제도, 국가, 기업, 금융’ 등 2000년 이후 개혁과 변화의 대상이었던 단어들이 주로 많이 나타나고 있다. 그러면서도 ‘시장, 금융, 재정, 규제’ 등에 대해서는 ‘안정’도 함께 언급되고 있음을 확인할 수 있다. 이는 안정을 통한 변화를 추구해야 하는 분야의 특성이 반영된 것이라 할 수 있다.³²⁾

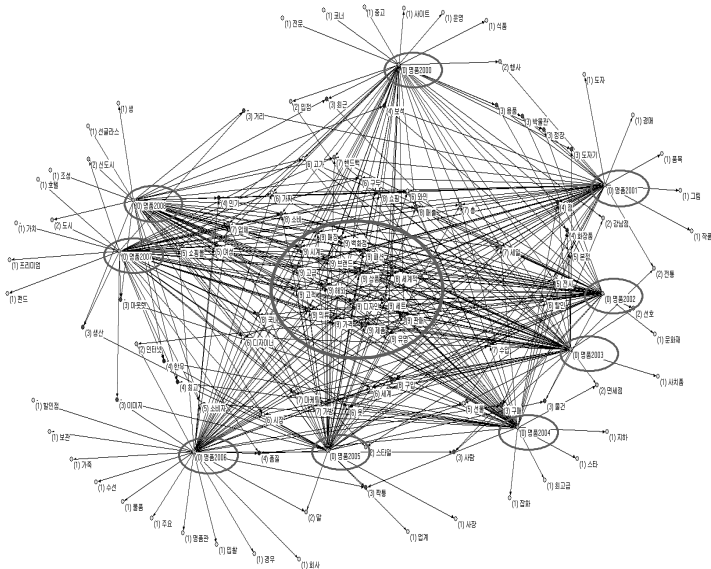
31) 이 네트워크의 밀도는 0.005, 평균 연결 정도는 2.489, 중심화지수 0.007으로 나타났다.

32) 이밖에도 ‘성장 : 복지 : 분배’의 네트워크에서는 ‘성장 : 분배’ 또는 ‘성장 : 복지’가 의미적 대립 관계를 이루는 만큼 이들이 상대적으로 많은 공기어를 2-링크 이상 연결하고 있을 것이 예상되었다. 그러나 ‘성장 : 복지’의 네트워크에서는 예상과 달리 2-링크 관계가 상대적으로 적게 나타났다. 한편 2개의 중심어인 ‘통합 : 화해’의 네트워크에서는 두 중심어 간에 공기어를 함께 공유하는 경우가 비교적 적었다. 따라서 ‘통합’, ‘화해’ 두 중심어의 공기어들은 상대적으로 구분된다고 볼 수 있다.

3.2. 중심어의 통시적 네트워크

이 절에서는 중심어가 9년간의 통시적 흐름 속에서 연도별로 공기어들을 달리하는 양상을 하나의 네트워크 안에서 표현한다. 그리고 이를 통해 단어 사용의 통시적 변화들을 확인하고자 한다.33)

□ ‘명품’ 2000년 ~ 2008년



<그림 5> ‘명품’의 2000~2008년 네트워크

명사 ‘명품’의 사전적 의미는 “뛰어나거나 이름난 물건 또는 그런 작품”이다. 그리고 불과 몇 해 전까지만 해도 주로 도자기, 그림 등의 문화재나 골동

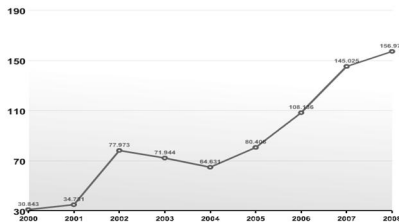
33) 3.1의 네트워크 그림은 여러 개의 중심어 간의 관계를 보인 것이라면 3.2의 네트워크 그림은 하나의 키워드를 각 연도별 중심어로 두고 공기어를 어떻게 공유하고 있는지를 보인 그림이다. 이 때 <그림 5>의 각 중심어 ‘명품2000’, ‘명품2001, ...’ 등은 각각을 별개의 단어로 취급한 것이 아니며 각 연도별로 공기어가 다르게 나타나는 것을 관찰하기 위해 임시로 중심어로 설정한 것이다.

품, 미술품 따위를 가리키는 의미로 쓰였다. 그러나 <그림 8>의 공기어 양상을 볼 때 현재 시점에서는 세계적 고급 브랜드의 시계, 의류, 핸드백, 구두 등 패션에 관한 제품을 의미하는 것이 명사 ‘명품’의 주의미라고 할 수 있을 것이다. 그림의 가운데 9-링크 관계의 공기어를 보면 ‘고급, 시계, 의류, 패션, 백화점, 디자인’ 등 대부분 패션 관련 상품에 관한 단어들 다수를 차지하고 있다.

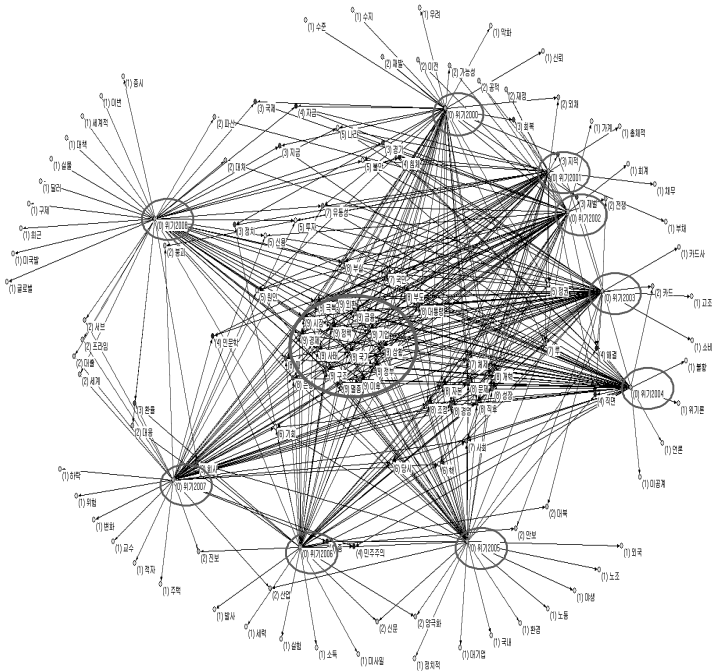
한편 명사 ‘명품’은 2000년대 후반 들어서 ‘편드, 호텔, 신도시, 한우’ 등과도 공기하면서 점차 “고급, 최고”에 해당하는 의미로 그 영역을 확장하고 있다. 2003년도에 ‘사치품’이 상위의 공기어로 나타나는 등 불과 수년 전까지만 해도 부정적 공기어가 상위에서 확인되는 반면, 2007년 이후에는 “고급, 최고”의 의미로 쓰이는 경우가 상위로 나타난다. 이러한 의미 영역의 변화는 필연적으로 빈도의 증가를 동반한다.³⁴⁾ 상대적으로 2000년, 2001년에 ‘박물관, 도자기, 경매, 그림, 작품’ 등과 공기하고 있는 점을 보면 이러한 특징이 잘 드러난다.³⁵⁾

□ ‘위기’ 2000년 ~ 2008년

34) 명사 ‘명품’의 연도별 상대빈도 변화(명사 단어 100만개 당 출현빈도)



35) <그림 5>의 네트워크는 밀도 = 0.031, 평균 연결 정도 = 7,500, 중심화지수 = 0.044의 값을 갖는다.



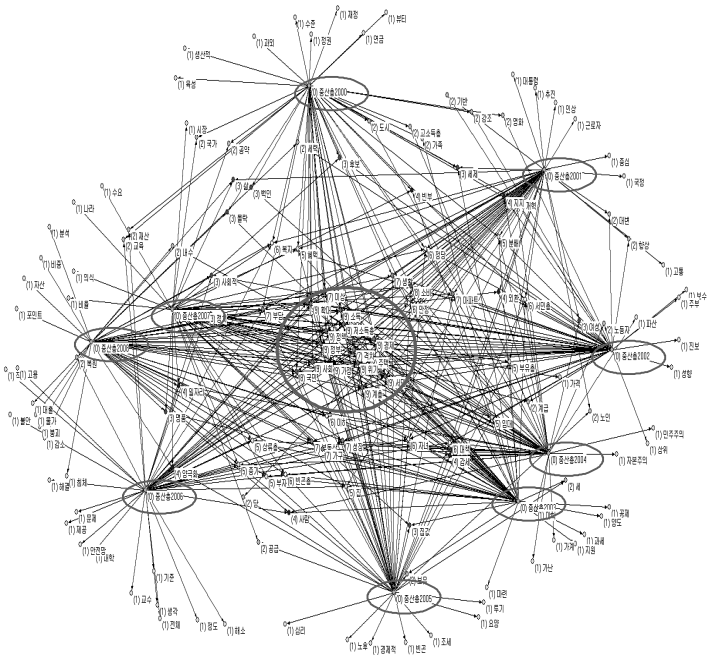
<그림 6> ‘위기’의 2000~2008년 네트워크
 (밀도 = 0.025, 평균 연결 정도 = 6.766, 중심화지수 = 0.042)

9개의 원으로 된 표시는 명사 ‘위기’를 2000년부터 2008년까지 각각의 중심으로 네트워크 상에 제시한 것이다. 9개년도 모두 ‘위기’와 공기하는 단어인 9-링크 공기어는 ‘기업, 금융, 외환, 시장, 정부, 국가’ 등의 단어가 상위의 공기어로 쓰이고 있다. 연도별로만 공기하는 단어들 중 2000년에는 ‘외채’나 국제 수지에 대한 우려 등이 반영되어 있으며, 2002년과 2003년에는 이라크 전쟁과 카드 사태에 관련된 단어들이 ‘위기’와 공기한다. 2005년과 2006년에는 양극화나 핵 위협에 따른 안보 문제 등과 관련된 단어들이 주로 나타나고, 2007년과 2008년에는 서브 프라임 주택 담보 대출로 인한 세계 경제 위기와 관련된 단어들이 주로 나타난다.

특히 2007, 2008년의 경제 위기시에는 과거의 경제 위기시에 쓰였던 단어

들이 아닌 새로운 공기어인 ‘서브 프라이م, 글로벌, 달러, 대출’ 등이 나타나는 것을 확인할 수 있다. 이는 시대별로 순환적으로 위기가 반복되더라도 명사 ‘위기’가 공기하는 단어들은 그대로 반복되는 것이 아님을 보여준다. 따라서 그 위기의 원인이나 결과도 다를 것이라는 추론이 가능하다.

□ ‘중산층’ 2000년 ~ 2008년



<그림 7> ‘중산층’의 2000~2008년 네트워크

명사 ‘중산층’에 대하여 살펴보면 2000년대 초반에는 ‘위기’나 ‘격차’, ‘몰락’ 등의 공기어가 나타나긴 하지만 상대적으로 다양한 분야의 공기어들이 나타난다.³⁶⁾ 먼저 ‘세금, 양도, 조세’ 등에 대한 관심이나 ‘아파트’, ‘부동산’, ‘명품’, 그리고 ‘민주주의, 자본주의, 진보’ 등 정치적 이슈에 이르기까지 다

36) 네트워크의 밀도 0.018, 평균 연결 정도는 5.806, 중심화지수는 0.039

양한 단어들이 상위의 공기어로 등장하고 있다. 9-링크 관계의 공기어를 제외하고는 대체로 경제 외적인 분야나 정치, 교육 등에 관련된 단어가 상위의 공기어로 나타난다.

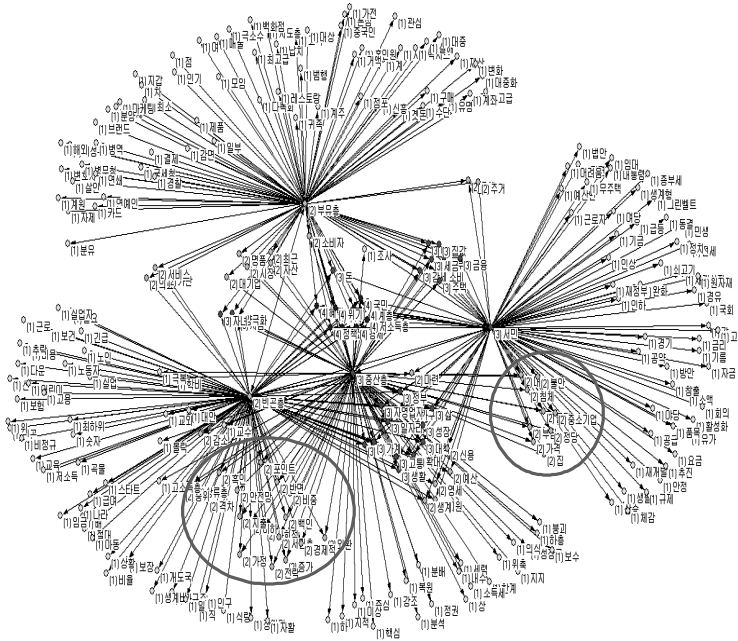
그리고 2007, 2008년에 이르러서는 ‘일자리, 대출, 물가, 붕괴’ 등도 관찰된다. 이에 반해 ‘서민’의 연도별 네트워크에서는 ‘임대, 부담, 은행’ 등의 경제 관련 명사들이 9-링크 공기어로 등장한다는 점에서 ‘중산층’과 차이점이 있다. 아래 (10)은 2000년부터 2008년까지 매해 각각의 중심어와 공기한 단어들을 제시한 것이다.³⁷⁾ 중산층에 비해 서민의 공기어가 상대적으로 생활 경제에 관련된 용어의 비중이 높다. 중산층은 경제 분야 이외에도 ‘가정, 사회, 사람, 계층’ 등이 9-링크 공기어로 나타나는데 이에 대한 관심도가 상대적으로 높다는 것을 의미한다.

(10) ‘중산층, 서민’과 2000년부터 2008년까지 모두 공기하는 단어

공기어(9-링크) : 가나다순	
중산층의 공기어	서민의 공기어
가정	경제
감세	금리
경제	금융
계층	대책
사람	대출
사회	부담
서민	안정
세금	은행
소득	임대
소비	정부
위기	정책
정부	주택
정책	중산층
주택	집

37) 지면의 제약으로 ‘서민’의 연도별 사용 양상에 의한 네트워크는 생략한다.

표시된 부분이 ‘중산층 : 빈곤층’의 2-링크 공기어이다. 전자는 후자에 비해 유형도 많고 내용도 다양하다. 그러나 6년 후인 아래 <그림 9>를 보면 그 양상이 많이 다르다. ‘중산층 : 서민’의 2-링크 공기어가 급격히 줄어들면서 상대적으로 ‘중산층 : 빈곤층’이나 ‘중산층 : 서민’의 2-링크 공기어의 유형이 많이 증가하였다.



<그림 9> ‘부유층 : 중산층 : 서민 : 빈곤층’의 네트워크 (2008년)

그리고 위의 2008년도 네트워크에는 ‘중산층 : 서민’의 2-링크 공기어에는 ‘침체, 불안, 부담’ 등 2002년에는 나타나지 않았던 단어들이 포함된다. 그리고 ‘중산층 : 빈곤층’의 공기어는 2002년에는 ‘노인, 심화, 증가, 위기’ 네 단어뿐이었지만 2008년에는 유형의 수도 훨씬 많아지고 ‘경제적, 격차, 외환, 안전망’ 등 다양한 의미의 명사들로 확대되었다. 이처럼 우리는 사회 계층과 관련한 가치를 내포하고 있는 ‘중산층’이라는 명사가 지금까지 상대적으로

‘서민’과 고빈도로 공기하였으나 이제는 ‘빈곤층’과의 쓰임이 더욱 늘어나고 있다는 점을 확인할 수 있었다. 이는 단순히 공기하는 단어만 확인해서는 포착할 수 없는 것이며 시간의 흐름에 따라 네트워크에서 공기하는 단어가 변화해 가는 양상을 검토해 보아야지만 알 수 있는 것이다. 아래에는 2002년과 2008년도 각각의 네트워크에서 중심어와 공기하는 단어 중 변화 양상이 뚜렷한 2-링크 관계의 공기어들을 표로 제시하였다.³⁸⁾

(10) 중심어 ‘중산층’과 공기하는 단어의 목록 비교 (가나다 순)

공기어(2-링크)					
중산층과 부유층		중산층과 서민층		중산층과 빈곤층	
2002년	2008년	2002년	2008년	2002년	2008년
부모	대기업	근로자	가격	노인	격차
상위	명품	대책	부담	심화	경제적
세금	소비자	대통령	불안	위기	반면
소비	시장	부담	정당	증가	비중
	자산	분배	중소기업		사회적
	최근	안정	집		안전망
		이미지	침체		외환
		정치			전략
		주거			증가
		지지			지출

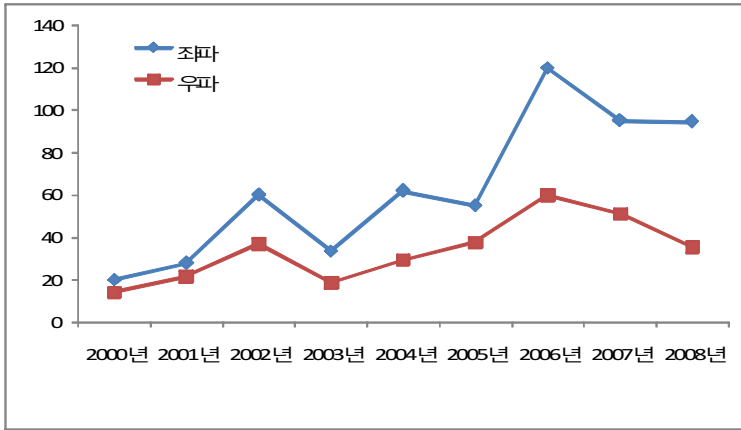
□ 우파, 좌파

‘우파’와 ‘좌파’의 네트워크에서는 앞서의 네트워크 분석과 다르게 t-점수 2.576 이상의 통계적으로 유의미한 공기어 일체를 분석 대상으로 하였다.³⁹⁾

38) 특별히 2-링크 공기어를 선택한 이유는 다른 수의 링크 관계인 공기어들보다 시간의 흐름에 따른 변화 양상이 뚜렷하게 나타나기 때문이다. 자세한 차이는 지면의 제약상 생략한다.

39) 앞서의 네트워크 분석에서와 다른 방식으로 관찰할 수 있는 점이 있음을 보이기 위해 공기어를 상위 몇 위로 구분하지 않고 통계적으로 유의미한 t-점수의 절대값을 만족하는 모든 공기어를 분석 대상으로 하였다.

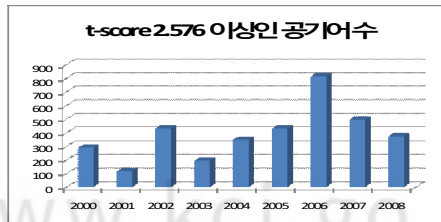
결과는 공기하는 명사의 수가 많아지면 네트워크의 크기도 따라서 증가하였고, 그 역도 성립하는 것으로 나타났다.⁴⁰⁾ 중심어 ‘좌파’와 ‘우파’ 각각의 코퍼스 내 상대빈도는 다음과 같다.



<그림 10> 중심어 ‘우파’, ‘좌파’의 연도별 상대빈도 변화

‘좌파’는 평균적으로 상대빈도가 증가하는 양상인 반면, ‘우파’는 최근으로 갈수록 감소 추세를 보인다. 단순히 보면 <그림 10>을 통해 우리는 이 두 중심어 모두가 대중의 관심이 늘면서 사용 빈도가 증가하게 되었고, 이로 인해 다양한 영역에서 쓰임이 증가하였을 것이란 점을 추측할 수 있다. 그러나 네트워크 분석의 결과는 이러한 추측이 잘못되었음을 보인다. 우리는 네트워크를 통해 두 중심어 중에서 특히 ‘좌파’가 급격한 변화 양상을 보인다는 점을

40) 명사 ‘우파’, ‘좌파’와 통계적으로 유의미하게 공기하는 명사 수의 변화



과’ 두 중심어의 출현빈도는 약 2배 차이밖에 안 되지만 단독으로 공기하는 단어의 유형은 수십 배 차이가 나고 있는 것이다. 이러한 현상은 앞서 <그림 10>에서 제시한 바와 같이 단순히 중심어의 빈도 변화만을 가지고 분석하는 기존의 연구 방법론으로는 발견하기 어렵다.

이처럼 <그림 13>에서는 두 중심어 모두와 공기하거나 ‘좌파’와만 공기하는 명사의 유형만이 급격하게 증가하였다. 만약 ‘좌파’와 ‘우파’의 대립이 첨예한 시기였다는 시대 상황을 반영한 것이라면 이 두 중심어 각각과 공기하는 단어도 증가해야 했겠지만 실체는 달랐다.

이를 해석하면 우선 ‘우파’의 공기어는 거의 대부분 ‘좌파’도 공기하는 문단에서 쓰여 대응을 이루지만, ‘좌파’는 ‘우파’와 공기하지 않더라도 많은 맥락에서 유의미하게 나타나고 있으며 따라서 언중들이 굳이 우파와 대립시키지 않더라도 어휘 의미를 명확하게 인식하는 것이라고 할 수 있다. 한편 기존에 부정적이었던 명사 ‘좌파’의 의미가 중립적이거나 긍정적인 것으로 변화되어 보다 다양한 맥락에서 쓰이게 된 것이라고 볼 수도 있을 것이다.⁴³⁾

4. 맺음말

지금까지 26 개의 가치 명사를 대상으로 이들의 공기어를 통계적으로 유의미한 범위 내에서 일부 추출하여 상호 연결 관계를 이용한 네트워크를 구축하고, 특징적인 네트워크를 분석하는 과정을 통해 가치 명사의 사용 양상과 변화 흐름을 살펴보았다. 이 연구의 의의는 단어의 의미를 결정하는 중요한 요소로 공기 관계를 이루는 단어를 택했다는 점, 그리고 통계적으로 유의미한 공기 관계의 명사들만을 택했다는 점에서 주관적인 직관에 의한 분석이 아닌 객관적인 자료와 방법론에 의한 분석이라는 점에 있다 하겠다. 특히 문

43) 코퍼스 검토 결과 우파를 대체할 수 있는 단어인 ‘우익’으로 인해 ‘우파’의 쓰임이 줄어들었을 가능성은 거의 없는 것으로 보인다. 실제 출현 빈도도 낮고 ‘우익’은 대부분 ‘좌익’과 대응되는데, 이들도 사용되는 맥락이 달리 나타나고 있었다. 익명의 심사자들께서는 이에 대해 좌파라는 단어의 의미가 중립적인 것을 넘어서 긍정적인 것에까지 나아갔을 가능성이거나 단어의 개념이 변화했을 가능성을 제시하였다. 필자 역시 이 두 가능성 모두에 동의하는데 다만 이 가능성에 대한 보다 자세한 논의는 지면의 제약으로 인해 후고를 기약하고자 한다.

장 차원에서 벗어나 문단 내에서 공기하는 단어들의 관계를 기반으로 서로 연관되는 공기어들이 어떻게 변화해 가는지에 대한 관찰을 하고, 이를 통해 기존의 언어 관계나 문장 내의 공기관계와는 다른, 새로운 현상을 관찰할 수 있었다.

이 연구는 언어 현상이 복잡계로 이루어진 네트워크라는 가설에 의해서 시작되었다. 복잡계 연구는 혼돈으로부터의 질서를 찾는 연구이다. 언어 현상 역시 매우 복잡한 현상 속에서 일부분을 이루는 작은 세계의 속성들을 포착할 수 있다.⁴⁴⁾ 이러한 현상은 네트워크로 표현된 시각적 결과물이나 수학적 수치로만 확인할 수 있다.

주목할 만한 발견은 일부 단어의 의미가 10년 미만의 짧은 시간 속에서도 통시적으로 변화하는 양상을 보인다는 점이다. 특히 ‘명품’은 공기하는 단어가 연도별로 눈에 띄게 달라지는 양상을 확인할 수 있었다. ‘명품’의 공기어가 ‘골동품, 도자기’와 같은 역사적 의미를 띤 사물에서 점차 시간이 흐를수록 고가의 의류, 잡화류 등이나 ‘한우’, ‘신도시’ 등에 이르기까지 고급스런 이미지를 요구하는 공기어들과 네트워크를 이루면서 쓰임이 변화하고 있었다.⁴⁵⁾ 그리고 ‘중산층’ 역시 중심어인 ‘부유층’, ‘서민’과 연결된 공기어의 비중이 줄고, 반면에 ‘빈곤층’과 연결된 공기어가 증가하였다. 이는 의미적 상관 관계에 있는 단어들의 네트워크 분석을 통해 사용 양상이 변화하는 현상을 보여준 것이다.

한편 ‘좌파’와 ‘우파’는 2000년대 초반까지는 공기어의 수도 적고 공기어의 의미도 제한적인 경향으로 나타났으나, 2000년대 후반으로 갈수록 공기어의 유형이 늘어나고 의미도 다양해졌다. 이는 ‘좌파’의 의미 영역이 통시적으로 변화하고 있음을 보여주는 객관적인 증거라 할 수 있다. 이밖에 ‘갈등’도 2000년대 초반에는 주로 ‘화해’와 공통된 공기어들이 유의미하게 공기하고 있었지만, 2000년대 중반 이후부터는 ‘화해’와 연관된 공기어 비중은 줄고, ‘통합’과 연관된 공기어 비중이 늘어났다.

이처럼 우리는 경험적으로 막연히 알고 있는 사실을 네트워크 분석을 통

44) 이는 반대로 작은 세계의 속성들이 모여 매우 복잡한 전체를 생성하는 것으로 볼 수 있다.

45) 익명의 심사자께서는 이 논의에서 ‘명품’의 인접 단어가 변화하는 양상을 관찰할 필요가 있음을 언급하였는데 전적으로 동의하며 후고에서 자세히 다룰 예정이다.

해 계량적 자료로부터 객관적으로 입증할 수 있고, 또한 직관으로 생각하기 어려운 부분에 대한 관찰을 가능하게 한다는 점에서 네트워크 분석이 언어 현상 연구에 중요한 연구 방법론이라는 사실을 확인할 수 있었다.

참고문헌

- 강범모(2003), 『언어, 컴퓨터, 코퍼스 언어학』, 고려대학교 출판부.
- 강범모(2010), “공기 명사에 기초한 의미/개념 연관성의 네트워크 구성,” 『한국어 의미학』 32, 한국어 의미학회, 1-32.
- 강범모 외(2001), “한국어 명사 의미 부류 체계의 구축과 활용,” 제 13회 한글 및 한국어 정보처리 학술대회 논문집, 247-251.
- 고창수(2007), “언어의 인터페이스란 무엇인가,” 『한국어학』 37, 1-18.
- 김용학 외(2006), “온라인 사회 연결망의 구조분석,” 『한국 사회학회 사회학대회 논문집』, 225-227.
- 김용학(2007), 『사회 연결망 분석』, 박영사.
- 김일환·이도길·강범모(2010a), “SJ-RIKS Corpus: 세종 형태의미 분석 코퍼스를 넘어서,” 『민족문화연구』 52, 민족문화연구원, 373-403.
- 김일환·이도길·강범모(2010b), “감정명사의 사용 양상과 네트워크 분석,” 『한국어학』 49, 한국어학회, 117-148.
- 김흥규 외(2010), 『물결 21: 신문텍스트 기반의 장기간 언어·사회·문화 연구』, 고려대학교 민족문화연구원.
- 나인호(2010), “코젤렉의 개념사 연구,” 『내일을 여는 역사』, 41호, 271-293.
- 박병선(2005), 『한국어 계량적 연구 방법론』, 도서출판 역락.
- 신우봉·김일환·김흥규(2010), “신문텍스트에서 나타나는 공간명사의 사용 양상과 네트워크 분석,” 『텍스트 언어학』 29, 텍스트 언어학회, 221-250.
- 신호필(2009), 『언어학과 통계모델』, 서울대학교 출판부.
- 윤영수·채승병(2005), 『복잡계 개론: 세상을 움직이는 숨겨진 질서 찾기』, 서울: 삼성경제연구소.
- 채현식(2007), “어휘부의 자기조직화,” 『한국언어문학』 64, 한국언어학회, 137-155.
- 최경봉(1998), 『국어 명사의 의미 연구』, 태학사.
- 최호섭·옥철영(2002), “한국어 의미망 구축과 활용-명사를 중심으로-,” 『한국어학』 17, 한국어학회, 301-329.
- 한정환 외 역(2004), 『유로워드넷』, 한국문화사.
- 홍재성 외(2000), 『21세기 세종계획 전자사전개발 연구보고서』, 국립국어원.
- 홍종선·강범모·최호철(2000), 『한국어 언어 관계 연구』, 월인.

- Bybee, J. and P. Hopper (Eds).(2001), *Frequency and Emergence of Linguistic Structure*, John Benjamin, 449-70.
- Beckner, C. et al.(2009), "Language is a Complex Adaptive System: Position Paper," *Language Learning*, Dec 2009 Supplement 1, Vol. 59, Oxford: U.K., Blackwell.
- Fellbaum, C.(1998), *Wordnet: An Electronic lexical database*. Cambridge, Mass: The MIT Press.
- Firth, J. R.(1957), 'Modes of meaning', In J. R. Firth, *Papers in Linguistics* 1934-1951, London: Oxford University Press.
- Nooy, W., Mrvar, A. & Vladimir. Batagelj.(2005), *Exploratory social network analysis with pajek*, Cambridge, Cambridge University Press.
- Sinclair, J.(1991), *Corpus, concordance, collocation*, Oxford: Oxford University Press.
- Vaughan, L. & You, J.(2006), "Comparing business competition positions based on Web co-link data: The global market vs. the Chinese market," *Scientometrics*, 68(3), 611-628.
- Manning, C. & H. Schutze.(1999), *Foundation of Statistical Natural Language Processing*, Cambridge, Mass: The MIT Press.

이영제

서울시 성북구 안암동 5가
고려대학교 민족문화연구원
136-701
전화 번호: 02)3290-1617
전자 우편: kukl@korea.ac.kr

김홍규

서울시 성북구 안암동 5가
고려대학교 국어국문학과
136-701
전화 번호: 02)3290-1966
전자 우편: gardener@korea.ac.kr

강범모

서울시 성북구 안암동 5가
고려대학교 언어학과
136-701
전화 번호: 02)3290-2173
전자 우편: bmkang@korea.ac.kr

294 이 영 제·김 홍 규·강 범 모

원고 접수일: 2011년 2월 21일

원고 수정일: 2011년 4월 6일

게재 결정일: 2011년 4월 13일