

친족명사의 공기어 양상과 네트워크 분석*

정유진 · 강범모**

(고려대학교)

Chung, Eugene & Kang, Beom-mo. 2011. A Network Analysis of Family Nouns. *The Linguistic Association of Korean Journal*. 19(2). 209-235. This study aims to explore how family nouns and their co-occurring words are portrayed in a Korean newspaper collection, the *Trends 21* corpus (Kim *et al.*, 2010). We adopt network analysis using co-occurring words with target words, thirty family nouns. The co-occurring words are regarded as not only semantically related but also mentally associated with target words. A network is a graph-based structure made up of entities which are connected by specific types of relationship. We have used Pajek (Nooy, Mrvar & Vladimir, 2005) to visualize the relationship between a target word and its co-occurring words. The nodes are labeled as target words and co-occurrences; the links are indicated with the t-score in this study. Based on the network analysis, we identify uses of co-occurring words among target family nouns as well as the structure of networks by connection and centrality value. In addition, we can find out distinctions among semantic groups of family nouns; for example, 'couple' vs 'core-family'.

Key Words: co-occurring words, family nouns, network analysis, newspaper, corpus, Pajek

1. 서론

언어는 사람들 사이에서 일어나는 공동의 활동이며 그 의미는 언어의 다양한 사용에서 찾아볼 수 있다(비트겐슈타인: 김진철, 정병찬 1985).¹⁾ 어떤 단어의 의미를 이해하기 위해서

* 이 논문은 2007년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (KRF-2007-361-AL0013).

** 제1저자-정유진; 교신저자-강범모

1) 비트겐슈타인이 창시한 “언어게임” 이론은 언어의 의미는 다양한 사용, 즉 문맥 내에서 찾아야 하는 언어

단지 사전적 정의보다 실제 맥락/문맥에서 사용되는 여러 요인을 고려해야 한다. Leech(1974)가 분류한 여러 의미 유형 가운데 언어적/배열적 의미(collocative meaning)와도 상응하는 부분이다.²⁾ Leech(1974)가 정의한 언어적 의미란 “한 표현이 문맥에 나타나 있는 다른 표현의 의미에 영향을 받아 가지게 되는 의미”로 한 단어가 그것과 공기하는 단어들의 의미 때문에 획득하는 연상(association)들로 구성된다. 한편 공기어는 대상어와 관련이 되는 단어이자 사물을 표상하고 더 나아가 연관되는 개념을 표상한다. Ogden & Richards(1923)이 제시한 단어, 개념, 사물 사이의 의미 삼각형은 의미론의 심리학적 기반을 제시했다. 예를 들어 ‘선생’의 경우 사용되는 문맥 속에서 함께 나타나는 관련 단어(공기어, co-occurring word)를 고려하면 그 의미가 분명해진다. ‘선생’이 ‘학교, 학생’ 등과 공기할 경우 가르치는 사람의 의미로 사용이 된 것이고 ‘안중근, 김구’와 같은 역사적 인물과 공기할 경우 사람에 대한 존칭의 표현을 위해 함께 쓰인다.

이 연구에서 우리는 대규모 코퍼스를 기반으로 문맥 내에서 공기하는 단어들을 추출하여 연구 대상이 되는 단어의 실제 사용 양상을 살펴보고자 한다. 특히 인물명사 가운데 이 사회 구성의 기본이 되는 대상어로 선정하여 10년 동안(2000년~2009년)의 신문 텍스트에서 공기하는 단어들을 살펴본다. 이를 통해 지난 10년간 우리의 생활을 서술하는 신문 기사에서 친족명사가 어떤 단어들과 사용되는지 알아본다. 다음으로 단어들의 공기 관계를 바탕으로 한 공기어 네트워크를 이용하여 단어들 사이의 사용양상을 시각화한다. 이 공기어 네트워크를 통해 친족명사들 사이에 어떤 의미영역이 형성되는지, 여러 대상어 사이에서 나타나는 중첩 공기어, 그리고 특정 대상어와만 사용되는 배타적 공기어를 알아보고 네트워크의 구조적인 특성을 관찰한다.

이 논문의 구성은 다음과 같다. 다음 장에서 연구 대상이 되는 친족명사와 공기어 추출 방법에 대하여 개관한다. 3장에서는 연구 방법으로 채택된 네트워크와 네트워크를 구조적으로 설명하는 여러 지표에 대해 설명한다. 4장에서 대상어와 추출된 공기어를 t-점수를 이용하여 시각화한 후 네트워크의 구조적 특성을 분석한다. 5장에서는 이 연구를 요약하고 그 의의를 논한다.

2. 연구 대상

이 논문은 고려대학교 민족문화연구원에서 10년간의 신문 기사를 기반으로 구축한 [물결

사용에 초점을 두고 있다. 즉 언어가 외적 세계에 대한 그림이 아니라 다수의 참여와 일련의 규칙 준수가 전제되어야 하는 게임처럼 언어도 일종의 게임처럼 볼 수 있다는 것이다.

2) Leech(1974)는 개념적 의미, 내포적 의미, 문체적 의미, 정서적/감정적 의미, 반사적/반영적 의미, 언어적 의미, 그리고 주제적 의미와 같은 7가지 의미 유형을 제안하였다.

21]3) 코퍼스에서 대상어와 한 문단에서 같이 나타나는 단어들을 추출해 친족명사들이 의미 부류를 형성하는데 공기어가 어떤 역할을 하는지 알아보려고 한다.

신문 기사의 궁극적인 목표는 정보 전달이며 이를 위해서 문자 언어의 양식으로 기록하고 있다. 또한 기사는 사람들이 속해있는 집단이 관련 되거나 관심을 두고 있는 사건을 기술함으로써 그 시대의 문화적, 사회적 실체를 구현한다. 일상생활 속의 단어와 사물 그리고 개념은 언어를 사용하는 공동체에 의해 그 연관성이 성립이 된다(강범모, 2010)는 점에서 공동체의 관심을 기록하는 신문기사는 단어의 의미를 보다 구체적으로 설명하는 중요한 언어자료로 사용될 수 있다. 이 연구는 내용어 가운데 사건기술의 대상이 되는 명사만을 고려한다. 특히 명사 가운데 인물명사는 사건의 기술에 있어 주체 또는 객체가 되어 어떤 사건을 인식하는 데에 중요한 역할을 하게 된다. 어휘 의미를 파악할 때 어휘 자체의 의미와 함께 다른 어휘와 함께 사용되어 구체화되는 의미 관계와 의미영역을 알아보기 위하여 한 문단 내에서 출현하는 공기어도 살펴보고자 한다. 따라서 이 논문에서는 인물명사 특히 사회 구성의 가장 기본이 되는 친족명사를 기반으로 그 의미 양상을 인식할 수 있는 단초를 제공하고자 한다.

2.1. 친족명사 어휘 선정

연구의 대상이 되는 한국어의 인물명사⁴⁾는 세종전자 사전의 의미부류 체계에서 구축된 어휘 의미부류를 따른다. 세종 전자사전의 의미부류에서 인물 명사는 <인물>, <기관>, <단체>, <모임>, <조직>, <직업>, <직위>, <집단>, <친족>, <호칭> 등과 같이 보다 세밀한 의미영역을 갖는 부류들로 나누어진다(이성현 2004). 어휘의 특성상 하나의 명사는 여러 의미 부류에서 나타날 수 있다. 예를 들어 ‘아버지’란 명사는 <친족>의 부류에 속하는 동시에 <호칭>의 부류에서도 찾아볼 수 있다. 최규일(1986)은 친족에 대하여 다음과 같이 정의를 하고 있다.

3) [물결 21] 코퍼스는 신문 텍스트를 통해 장기간의 언어, 사회, 문화 현상을 연구하기 위해 추진 중인 ‘물결 21(Trends 21)’ 사업에 의해 구축된 언어 자원 데이터 베이스이다. 이 코퍼스는 동아일보, 조선일보, 중앙일보, 그리고 한겨레신문의 2000년 첫 10년 동안(2000년부터 2009년까지) 간행된 신문 기사를 모아 놓은 것으로 규모는 약 4억 어절에 이른다(김홍규 외, 2010).

4) Nida(1975), Lyons(1977)과 최경봉(1998)에서는 실제적이며 관찰 가능한 명사류에 대해 특정한 의미 부류를 설정하고 각각 ‘사람’ 또는 ‘인간’이라는 의미부류를 할당한다. 세종의미부류 체계(이성현, 2004)에서는 논항의미 자질로서 ‘인간(인물)’이라는 부류를 제시한다. 코어넷(한정환 외, 2008)에서 제시하는 [사람] 개념을 나타내는 어휘의미체계는 크게 [인간], [준인간], 그리고 [사람(직업/지위/역할)]로 나뉘어 모두 7 단계로 분류되어 총 367개의 개념에 5806 가지의 의미항목으로 구성되어 있다. [인간]은 다시 [인칭]과 [인간(생물학적 특징)], [인간(친족관계)], [인간(대인관계)], [인간(사회집단)], [인간(능력/성향)]에 따라 구분된다. 구체적으로 [남자]와 [여자 개념을 나타내는 범주인 [인간(남녀)는 [인간(생물학적 특징)]에 속하는데, 여기에는 [인간(연령)]과 [인간(신체적 상황)] 개념도 포함되어 있다.

- (1) 사회의 가장 최소 단위인 가족의 범위를 넘어서 형성된 하나의 집단, 즉 가족이 확대된 집단임
- (2) 혈족을 주축으로 하여 인척 중에서 가장 가까운 관계를 유지하고 있는 사람들을 포함한 집단
- (3) 혈연관계와 인척관계에 의하여 맺어진 사람들을 의미한다. 그리하여 ‘혈족’은 혈연관계를 맺고 있는 사람이요, ‘인척’은 혼인에 의하여 나와 관련된 사람을 말함

이와 같이 친족은 가족을 단위로 하여 형성되는 사회구조로 부부관계, 친자관계, 형제관계 등 세 가지 관계가 친족의 가장 기본적인 요소이며 사회 구성원으로서의 역할을 결정한다. 친족은 사회를 구성하는 가장 기본 단위로서 사회 현상의 주체가 되어 기사에 언급되어진다.

친족어는 계열에 따라 아버지를 통해 형성되는 친족인 부계친(父系親), 어머니로 인해 형성되는 모계친(母系親), 남편을 통한 친족인 시계친(媳系親, 또는 夫系親), 아내를 통한 친족인 처계친(妻系親) 등에 쓰이는 지칭어와 호칭어로 분류된다. 친족어휘는 지칭어가 동시에 호칭어로 사용되기도 하며 어휘에 따라서는 지칭어로만 사용되기도 한다.

이 연구의 대상이 되는 친족명사는 가족 구성의 기본 요소인 부부, 친자, 형제 관계를 바탕으로 부계, 모계, 시계친족 어휘와 친족과 관련된 ‘가족, 남매, 부부, 사돈, 친정’과 같은 어휘 가운데 중의성과 빈도를 고려하여 30개로 한정하여 선정되었다.⁵⁾ 구체적으로 부계친에 속하는 ‘아버지’, ‘어머니’, ‘할아버지’, ‘할머니’, ‘형’, ‘형수’, ‘오빠’, ‘올케’, ‘누나’, ‘매형’, ‘언니’, ‘형부’, ‘동생’, ‘제수’, ‘매제’, ‘아들’, ‘딸’, ‘며느리’, ‘사위’, ‘조카’, ‘손자’, ‘손녀’ 등이 포함된다. 그리고 ‘아버지’의 유의어인 ‘아빠’, ‘어머니’의 유의어인 ‘엄마’, 시계친 어휘로 ‘시아버지’, ‘시어머니’가 선택되었다. 이밖에 친족과 관련된 ‘사돈’, ‘남편’, ‘아내’, ‘시댁’, ‘친정’ 등의 어휘 등도 연구 대상이 된다.

[물결 21]⁶⁾ 코퍼스를 기반으로 하여 선정된 어휘 항목, 절대빈도 그리고 상대빈도(비율)은 아래 표1에서 제시되고 있다.⁷⁾

5) 의미 분석 없이 형태소 분석만 된 코퍼스의 특성상 동음이의어와 같은 중의성을 가진 어휘는 배제하였다.

6) [물결 21] 코퍼스는 신문 텍스트를 통해 장기간의 언어, 사회, 문화 현상을 연구하기 위해 추진 중인 ‘물결 21(Trends 21)’ 사업에 의해 구축된 언어 자원 데이터베이스이다. 이 코퍼스는 동아일보, 조선일보, 중앙일보, 그리고 한겨레신문의 2000년 첫 10년 동안(2000년부터 2009년까지) 간행된 신문 기사를 모아 놓은 것으로 규모는 약 4억 어절에 이른다(김홍규 외, 2010).

7) 표1은 연구 대상어 사이에 의미적으로 연관된 단어를 중심으로 임의적으로 배열이 되어 있다.

표 1. 친족명사의 순위와 빈도

단어(순위)	절대빈도 ⁸⁾	상대빈도(% ⁹⁾	단어(순위)	절대빈도	상대빈도(%)
가족 (190)	184,428	0.078	시댁 (12036)	1,717	0.0007
부모 (577)	85,837	0.036	친정 (6135)	5,384	0.002
아버지(507)	96,723	0.041	식구 (4234)	8,939	0.004
아빠 (665)	29,088	0.012	부부 (811)	60,658	0.026
어머니(684)	74,887	0.032	남편 (698)	72,105	0.030
엄마 (722)	72,053	0.005	아내 (940)	53,831	0.023
아들 (473)	100,553	0.043	남매 (4679)	7,830	0.003
딸 (792)	63,638	0.027	자매 (4114)	9,233	0.004
시아버지(14119)	1,324	0.0006	언니 (3564)	11,617	0.005
시어머니 (6975)	4,385	0.002	오빠 (4069)	9,486	0.004
며느리 (5039)	7,050	0.003	누나 (4928)	7,439	0.003
할아버지 (2302)	5,384	0.009	동생 (1599)	30,897	0.013
할머니 (1271)	39,634	0.017	사돈 (12052)	1,911	0.0008
손자 (4491)	8,545	0.004	사위 (5916)	5,631	0.002
손녀 (8411)	3,369	0.001	삼촌 (9528)	2,725	0.001

일반명사의 총 어휘수는 236,564,751이며 최고빈도를 나타내는 어휘가 '말'(0.771%)임을 고려하면 대상 어휘가 전체 명사에서 차지하는 비율을 가늠해 볼 수 있다.

2.2. 친족명사의 공기어 추출

공기관계를 분석하기 위해서는 대상어(target word)와 함께 나타나는 어휘들이 적당한 크기의 문맥 범위에서 추출되어야 한다. 언어와 문맥의 관계에서 문맥은 말을 하게 하는 실제적인 상황뿐만 아니라 언어 사용자를 둘러싸고 있는 심리적, 사회적, 문화적, 그리고 역사적 요인까지도 포함한다(비트겐슈타인: 김진철, 정병찬 1985).

이 논문은 하나의 기사문 내에서 동일 사건을 기술하기 위해 사용된 단어들은 하나의 주제를 형성하기 위해 상호 작용한다(Ide & Vernois, 1998)는 전제 아래 대상어와 같은 문단 내에서 공기하는 명사들을 추출한다.

8) [물결 21] 코퍼스의 전체 어휘 수는 403,523,934개이다.

9) 제시된 백분율은 전체명사 가운데 해당 어휘가 차지하는 비율이다.

공기 명사 연어(collocation)는 일반적으로 주변 몇 단어, 혹은 한 문장으로 범위를 제한한다. 이것은 명사들 사이의 통사적 연관관계에 중점을 두는 연구에서 많이 실행된다. 본 연구의 목적은 실제 신문에 사용되고 있는 명사들 사이의 연관성을 포착하려는 것이다. 이것은 명사들 사이의, 나아가 명사들이 지시하는 세계(사회, 문화) 속의 사물들 사이의 관계를 파악하는 일이다. 그리고 그러한 의미들의 연관성은 통사적인 것이 아니며, 일관된 문맥 속에서의 공기 관계가 중요하다. 문단은 일관성이 보장되는 언어의 단위이다. 의미의 연관성은 단지 내용이나 뜻만이 아니라 상황, 맥락, 레지스터가 반영되어서 개별 단어가 표시하는 논리의 차이도 반영하는 것이 바람직하다. 동일 주제 및 사건 기술에 나타난 단어의 연관성을 보이기 위한 본 연구의 목적을 위하여 우리는 하나의 문단을 문맥으로 정하여 대상어와 동일 문단에서 등장하는 명사를 연구대상으로 한다.

다양한 문맥 범위 가운데 문단이나 기사 단위는 단어 사이의 공기 관계를 이용하여 같은 주제를 나타내고 있다는 점이 기존의 연구에서 이미 모형화된 바 있다(Ide & Vernois 1998). 의미의 연관성 연구를 위하여 Veling & van der Weerd(1999)는 하나의 기사를 문맥 단위로 설정하였다. 이들은 로이터 신문의 21,578개 기사에서 공기하는 단어들을 주목하여 일관된 주제를 기술하고 있는 어휘 항목들을 추출하였다. 예를 들어 {water, rainfall, dry} 혹은 {cola, PEP, pepsi, Pepsi-cola, Pepsico} 등은 동일한 의미 영역에 속하는 어휘이며 이들은 같은 문맥 내에서 하나의 주제를 기술하고 있다고 한다. 이들의 연구는 일반적으로 공기 빈도가 높은 어휘들은 같은 주제나 개념을 서술하기 위해 함께 사용되는 경향이 강하며 그 어휘들을 동일한 의미영역으로 할당하는 것이 바람직하다고 제시하였다.

신문은 소설이나 수필과 같은 다른 텍스트와 비교해 볼 때 한 문단의 어절 수와 문장 수가 비교적 적은 수로 구성되어 있다. 그리고 명사는 내용어(content word)로서 사건을 기술하는데 중요한 역할을 하며 같은 품사의 단어들이 비슷한 맥락에서 나타날 수 있다(강범모, 2010)는 점에서 이 논문은 명사에 한정하여 공기어를 추출한다.

어휘들의 공기 관계는 단어들 간의 관계를 연구하기 위해 전산 언어학이나 정보추출과 같은 분야에서 많이 사용되고 있다. 예를 들어 단어들 사이의 연관도(word association)를 측정하기 위해 공기 관계를 사용한다. Spence & Owens(1990)은 공기 빈도가 높은 어휘들은 그 단어 간의 연관도가 강하다고 실험을 통해 밝혔다.

이 연구에서는 단순히 빈도가 높게 나타나는 공기어들을 보는 것이 아니라 공기하는 단어들의 상관관계를 검증하는 통계적 유의성 지수 중 t-점수를¹⁰⁾ 통하여 공기어 쌍을 추출한다. t-점수는 공기되는 어휘가 코퍼스 상에서 어휘 자체가 고빈도 어휘이기 때문에 대상어와

10) t-점수는 공기되는 어휘들의 빈도를 계산하고 각각의 어휘가 출현하는 빈도를 기초로 하여 두 어휘의 연관성을 측정한다. t-점수는 관측빈도와 기대빈도를 기초로 계산이 된다. 관측빈도는 문단에서 대상어와 함께 나오는 단어(w)의 빈도를 세는 것으로 구해진다. 기대빈도는 전체 코퍼스크기(C), 전체 코퍼스에서 단어($f_c(w)$)의 빈도, 그리고 대상어와 함께 나온 문단의 어절수(p)의 비율로 계산을 한다.

의 공기 빈도가 높게 나오는 것과 저빈도 어휘이지만 대상 어휘와 유의미하게 함께 사용되는 경우를 구분하여 공기어를 추출시킬 수 있는 장점이 있다 (강범모, 2003; 박병선, 2005; Church et al., 1991).

친족명사 가운데 ‘아버지’의 공기어 양상을 살펴보면 아래 표2과 같이 빈도에 기반하여 추출된 공기어와 t-점수에 의해 나오는 공기어가 다르게 나타나는 것을 확인할 수 있다.

표 2. ‘아버지’의 공기어

공기빈도			t-점수		
아버지	말	24,086	아버지	아들	147.304
아버지	때	23,028	아버지	어머니	140.553
아버지	아들	22,979	아버지	때	108.899
아버지	어머니	20,707	아버지	딸	103.483
아버지	사람	13,028	아버지	가족	91.944
아버지	일	12,760	아버지	집	82.935
아버지	딸	11,512	아버지	아이	82.868
아버지	뒤	11,221	아버지	일	78.974
아버지	가족	10,676	아버지	자신	76.658
아버지	아이	10,456	아버지	말	73.349

단순 공기 빈도에 의해 공기어를 추출할 경우 ‘말’, ‘때’, ‘사람’, ‘뒤’¹¹⁾처럼 각 명사의 독립적인 빈도 자체가 높아서 많이 나오는 경우가 함께 제시되는 반면, t-점수를 통해 걸러진 공기어는 통계적으로뿐만 아니라 우리의 직관상으로도 유의미성을 가진다. 즉, 친족명사 ‘아버지’의 경우 t-점수로 추출된 공기어 중 가장 높은 유의도를 가진 어휘는 ‘아들’인 반면 단순 빈도로 추출된 공기어는 ‘말’이다. 일반명사인 ‘말’은 전체 일반명사 빈도 순위 중 1위에 해당하는 고빈도 어휘이다. 그렇다 하더라도 명사의 수가 월등히 많으면 공기명사도 아주 많아질 수밖에 없고 이것이 t-점수에 반영된다. ‘때’, ‘일’, ‘말’ 같은 명사들이 그렇다.

t-점수는 이와 같이 두 단어가 동일 문맥 내에서 빈도의 고저에 의해 나오는 즉, 기본적

$$E_w = f_c(w) \times \frac{p}{C}$$

$$t-score = \frac{O-E}{\sqrt{O}}$$

E: Expected Frequency (기대빈도)

O: Observed Frequency (관측빈도)

11) ‘말’, ‘때’, ‘사람’ 등은 전체 일반명사 빈도 순위 10위 안에 드는 고빈도 어휘들이다.

으로 코퍼스에서 많이 출현해서 공기어로 추출되는 것을 방지하는 기능을 하고 있다. 이 지표는 공기 빈도는 낮다 하더라도 대상어와 함께 유의미하게 많이 나오는 어휘를 공기어로 선정한다.

연구 대상이 되는 언어 자료인 [물결 21] 코퍼스의 기본 통계 정보는 다음과 같다.

표 3. [물결 21] 코퍼스 통계 정보

대상	구분	수량
코퍼스 전체	어절 수 ¹²⁾	403,523,934
코퍼스 전체	기사 수	1,938,014
코퍼스 전체 중 일반명사	유형	540,868
코퍼스 전체 중 일반명사	빈도	263,564,751

대상 코퍼스는 2000년부터 2009년 사이 신문 기사 가운데 광고, 사진, 일기예보, 인사, 동정 등을 제외한 기사, 사설, 칼럼 등의 텍스트 전체를 대상으로 KMAT(Korean Morphological Analyzer and Tagger, 김홍규 외, 2010)을 이용해 형태소 분석과 품사 정보가 주석이 된 명사들을 조사한다. 연구 대상 인물명사와 공기하는 어휘들은 자동검색기(김일환 외, 2010a)를 이용하여 추출하였다.

3. 연구 방법론

기사에서 언급되는 친족명사와 관련되는 명사들의 공기관계를 시각화하여 신문에서 나타나는 사용 양상을 고찰해 보기 위하여 이 논문에서는 사회연결망 분석(social network analysis)을 사용한다.

사회연결망의 기본 전제는 다수의 행위자들 사이의 연결과 관계를 네트워크로 표현하여 한 사회에 존재하는 구성원의 상호 작용 양상을 관찰하는 것이다. 우리가 사용하는 언어도 이와 같은 맥락에서 살펴볼 수 있다. 즉, 하나의 주제를 기술하는 문단 안에서 단어라는 구성원과 t -점수로 나타나는 단어 사이의 연결 관계를 네트워크의 방법론으로 제시한다. 이를 기반으로 친족명사들이 어떤 의미영역을 형성하는지 그리고 지난 10년간 친족명사의 공기어 사용 양상을 네트워크를 통해 시각화하고 네트워크의 구조적 특성을 알아본다.

12) 장편 소설 한편의 어절 수를 5만 어절로 가정할 경우 [물결 21]코퍼스의 크기는 장편 소설 8,000여권에 이르는 분량이다.

이러한 목적을 위해 첫째, 친족명사 각각의 어휘가 공기하는 명사들을 살펴본다. 둘째, 연구 대상이 되는 친족명사의 2000년부터 2009년 동안 공기하는 어휘들을 살펴보고 여러 친족명사와 중첩되는 공기어와 하나의 친족 명사와만 사용되는 배타적인 공기어를 살펴봄으로써 어휘 사용의 변화를 파악한다. 그리고 어휘들이 형성하는 네트워크의 구조 분석을 위해 밀도와 연결성을 고찰한다. 이는 신문 기사에서 사용되는 친족명사와 공기하는 어휘들의 사용양상이 네트워크로 제시될 경우 어떠한 구조적 특성이 도출될 수 있는가를 탐색하게 하고 공기양상을 시각화하여 어휘간의 비교를 용이하게 한다. 이 연구에서는 네트워크의 시각화를 위한 컴퓨터 소프트웨어 프로그램으로 Pajek 1.27(Nooy, Mrvar & Batagelj, 2010)을 사용한다.

3.1. 네트워크 분석

언어 연구에서 논의되고 있는 기존의 네트워크들은 주로 단어들의 의미 관계를 망으로 엮은 것으로 낱말밭(semantic field)이론(김응모, 1989, 2000; 김지형, 1999; 배해수, 1982, 1989), 의미망(semantic network), 자연언어처리를 위해 전산화된 코어넷(황도삼 외, 1999), 그리고 워드넷(Miller, 1995) 등이 대표적인 어휘 네트워크이다. 이 네트워크들은 단어들의 의미관계를 나타낸다고는 하지만 문맥을 배제한 어휘관계를 제시하는 것으로 문맥 내 공기어 간의 관계를 제시하기에는 부족한 면이 있다.

네트워크는 노드(결점, node, vertex)와 링크(link, arc, edge)로 구성이 되며 인터넷, 친족관계, 조직도도 하나의 네트워크라고 볼 수 있다. 이때 노드는 그 환경의 구성요소가 되며 링크는 구성요소들의 관계를 나타내는 선분이다. 네트워크를 표현하는 방식으로는 행렬(matrix)이나 시각화를 위해 그래프를 사용한다(고길곤, 2007). 네트워크로 형상화 되는 사회연결망 이론¹³⁾은 관계의 구조와 패턴을 설명하고 네트워크의 구성 개체와 구조를 설명한다(김용학, 2003). 이 이론은 상호작용하는 단위들 간의 '관계의 유형 및 질서'로 현상을 봄으로써 행위와 구조가 지닌 효과를 설명하려는 시도이다.

네트워크 분석은 분석 대상이 되는 시스템을 참여자와 관계의 함수로 정의하면서 이를 그래프 이론과, 통계확률, 시뮬레이션 등을 이용하여 그 구조를 분석하는 접근 방법이다(Scott, 2000). 이 분석은 전통적인 통계적 방법론과 달리 개별 행위 주체가 독립적이라는 가정 대신에, 개개의 행위자들이 상호 의존적이라는 가정 하에 서로간의 관계에서 도출되는 특

13) Michell (1969)은 일정한 유형들 사이에 형성된 특정한 연계를 전체가 지닌 특성이라 보고, 유형에 포함된 단위들의 행위를 설명하려는 시도가 사회연결망 분석이 지닌 특성이라고 말한다. 사회를 구성하는 요소를 유형으로 표현하고 각 구성요소간의 관계를 연계로 묶어 표현한 사회구조(Social Structure)를 소셜 네트워크(Social Network)라 한다. 현재 대부분의 연구에서 사용되는 사회연결망 표현 방식은 그래프형태를 띠고 있다.

성을 분석하는데 적합한 방법론이다(고길곤, 2007). 따라서 대상어와 공기어가 상호 의존적이라는 사실을 염두에 둔다면, 공기어를 기반으로 한 언어 사용 양상을 분석하기 위해서 네트워크 분석 이론을 언어에 적용하는 것이 효과적인 것이다. 최근 이런 네트워크 분석을 이용하여 정치인들 간의 상호 연결 관계나 자연과학 학술지에서 사용되는 키워드들의 연관도를 파악하는 것에 초점을 둔 연구(co-word analysis, Callon et al., 1991; Bhattacharya & Basu, 1998)가 시도되기도 하였다.

이러한 네트워크 분석의 장점 또는 유용성은 첫째, 어휘의 개별적 속성보다 관계성에 초점을 둔다는 점이다. 이는 기존의 어휘 네트워크 분석이 초점을 두는 개별 어휘의 속성만으로는 단어들의 의미 관계를 설명하는 데에 한계가 있다고 보기 때문이다. 즉 어휘들의 관계에 대한 분석을 통하여 개체의 속성을 밝힐 수 있다는 것이 네트워크 분석의 의의이다. 둘째, 네트워크 안에서 특정 어휘와 의미 부류에 따른 공기 관계를 통하여 의미군집에 대해 설명할 수 있다는 점과 문맥 내에서 함께 나타나는 복잡한 공기어 현상을 분석한다. 이밖에 단어 간의 관계분석을 통하여 숨어있는 구조적인 연결패턴을 발견하고 도식화할 수 있다. 즉 전체적인 수준에서 어휘간 연결 관계를 분석함으로써 공기어와 대상어의 구조적인 특성을 파악할 수 있다. 네트워크 구조는 어휘의 의미를 문맥 내에서 상호작용하는 어휘간의 관계와 질서를 파악할 수 있다는 장점이 있다. 특히 내향연결정도로 대상어와 공기어들, 의미부류, 군집 형성 그리고 유의어의 공기어 사용 양상을 알 수 있다. 그리고 외향연결정도로 대상어 간의 상호 관련성을 제시해 준다는 점은 다른 어휘 관계 분석 방법론에 비교해 네트워크 분석의 큰 장점이라 할 수 있다.

우리가 언어생활에서 사용하는 단어도 네트워크 이론에서는 하나의 구성 성분이 되며 이 성분들은 서로 복잡하게 얽혀 있으면서 의사소통이라는 거대한 질서를 만들어 내고 있다. 네트워크 구조는 어휘의 의미를 문맥 내에서 상호작용하는 어휘간의 관계와 의미 영역을 파악하는데 도움이 될 것이라 기대한다.

3.2. 네트워크 분석 지표

네트워크 구조 및 특성 분석은 연결성과 중심성으로 집약된다. 연결성은 네트워크의 구조와 관련된 정보를 제시하고 중심성은 네트워크의 특성을 알려준다. 이 논문에서 사용하는 자세한 분석지표는 표4에서 제시된다.¹⁴⁾

14) 연결성과 관련된 측정 내용으로 포괄성과 구조적 공백이 더 있으며 중심성과 관련되어 사이중심성과 같은 지표가 있으며 데이터의 성격에 따라 분석지표로 사용된다.

표 4. 네트워크 분석 지표

분석	분석방법	내용	측정내용
구조분석	연결성 (Connection)	연결 정도	degree mean
		밀도	density
특성분석	중심성 (Centrality)	연결 중심성	degree centrality
		전체 중심성	closeness centrality

첫째, 네트워크의 구조를 분석하는 연결성(Connection)은 네트워크를 구성하는 노드 간의 표면적인 연결 상태-특정 관계나 흐름의 정도-를 파악하는 가장 기본적인 분석 지표로 연결의 정도, 밀도, 포괄성, 구조적 공백으로 평가된다.

연결 정도(degree mean)는 네트워크의 결속을 나타내는 대표적인 지표로서 한 노드가 관계하고 있는 다른 노드의 숫자들로 정의된다. 예를 들어 10명 사이의 친구관계를 나타내는 연결망에서 A라는 사람이 가장 많은 친구와 연결되어 있다면 A는 연결 정도가 가장 높다고 말한다.

밀도는 관계가 가능한 총 관계 수(n) 중에서 실제로 맺어진 관계 수의 비율(m), $\frac{m}{n(n-1)/2}$ 으로 정의된다. 네트워크의 밀도가 높다는 것은 다양한 행위자 사이에 연계를 맺고 있다는 것을 의미한다(김용학, 2007: 93). 반대로 밀도가 낮게 나타난다면 이는 주요행위자 간의 연계 및 상호작용이 낮다는 것을 의미한다.

둘째, 중심성(centrality)은 링크 요소의 연결(degree)과 근접도(closeness)를 분석하여 한 노드가 연결망에서 얼마나 중심에 위치하는지에 대한 정도를 측정하는 방법으로 네트워크 구조의 특징을 파악하는 지표이다. 이 지표로 우리는 가장 중심에 있는 노드 및 연결 구조 속의 전체 연계에 대한 중심 연결 관계 정도를 파악할 수 있다.

지역 중심성을 의미하는 연결정도 중심성(Degree centrality) 분석은 연결망을 구성하는 어느 하나의 노드와 이것과 직접적으로 연결된 다른 노드와의 연결 정도를 측정하여, 각각의 점들이 연결망에서 얼마나 중심에 위치하는지를 알아보는 기법이다. 즉, 네트워크 구조에서 노드가 직접적인 관계 또는 흐름을 가질 수 있는 전체 경우의 수에서 실제 노드가 가지고 있는 직접적인 관계 또는 흐름 수의 비율로 특정하다. 이 지표가 높은 성분은 많은 다른 성분과 관계를 맺고 있는 것을 의미하며, 지표가 낮은 성분은 관계구조 속에서 적은 수의 성분과 관계하며 네트워크 주변에 위치하는 것을 의미한다. 전체 중심성(Closeness Centrality)은 연결망의 중앙에 위치하는 정보로 노드간 경로 거리의 합으로 측정된다.

종합하면 네트워크의 분석 지표는 대상어와 공기어의 구조적 특징을 조망하는 연결성(Connection)과 구성 요소인 단어 간의 연결 구조 및 중심 단어와 조직의 관계를 분석하는 중심성(Centrality)이 있으며 이를 통해 네트워크 구조와 그 특성을 분석한다. 이상에서 논

의한 네트워크 분석 지표는 네트워크 시각화 프로그램 Pajek 1.27을 통해 측정된다.

4. 친족명사의 네트워크 분석과 공기어 양상

이 장에서는 연구 대상인 친족명사의 공기어, 네트워크를 통한 시각화, 네트워크 분석 지표를 이용한 네트워크 구조와 특성 파악, 그리고 시각화된 네트워크를 통해 대상어와 관련어로서의 공기어의 사용 양상을 살펴본다.

먼저 친족명사 30개 전체의 네트워크를 제시하여 공기어 양상과 네트워크 구조를 분석한 이후 어휘적으로 관계가 있는 단어들끼리 묶어 네트워크를 보여주고 특성을 분석한다.

공기어 추출은 [물결 21]코퍼스에서 대상어와 함께 한 문단 내에 나오는 어휘들에 대해 t-점수를 부여하여 내림차순으로 정리한 후 상위 50개에 등장하는 어휘만을 보기로 한다. 그 후 대상어와 공기어를 하나의 쌍으로 만들어 매트릭스 형태로 정렬하고 Pajek에 매트릭스 형태로 정리된 데이터를 입력하면 대상어와 공기어의 관계를 시각화된 네트워크로 그 결과물을 얻게 된다. 네트워크 분석에서 활용되는 개념 중에서 이 연구와 관련되는 개념은 크게 두 가지이다. 하나는 연결망의 전체적인 구조 자체가 갖는 속성에 관한 지표이고, 다른 하나는 연결망을 구성하는 각 개체 혹은 노드의 중심성지표이다.

대상어 30개의 친족명사 가운데 ‘아버지’, ‘어머니’, ‘아들’과 한 문단 내에 나타나는 공기어와 t-점수는 표5와 같다.¹⁵⁾

표 5. ‘아버지’, ‘어머니’, ‘아들’의 공기어

대상어	공기어	t-점수	대상어	공기어	t-점수	대상어	공기어	t-점수
아버지	아들	159.9	어머니	아버지	152.7	아들	아버지	159.6
	어머니	152.6		아들	120.0		어머니	119.5
	때	117.9		때	103.9		딸	115.5
	딸	110.9		아이	103.4		아이	93.0
	가족	100.0		딸	94.3		집	90.0
	집	92.3		집	92.9		가족	86.3
	아이	90.2		가족	80.5		남편	84.1
	일	88.0		말	76.2		엄마	83.7
	자신	82.6		일	75.2		전	77.9
말	79.1	사랑	73.6	때	77.1			

15) 공간상의 제약으로 공기어 상위 27개만 제시하였다.

사랑	73.8	자식	69.8	대통령	76.3
사람	73.5	자신	69.1	아내	76.3
뒤	72.4	생각	66.2	부모	71.5
생각	70.8	엄마	66.1	부인	71.2
자식	70.1	마음	64.6	말	71.2
이야기	68.4	사람	62.3	일	70.5
시절	67.6	세상	62.1	자신	70.0
친구	67.3	부모	61.9	회장	68.5
부모	65.0	뒤	59.9	병역	65.1
세상	63.9	눈물	59.1	결혼	65.1
마음	62.2	자녀	59.1	뒤	63.7
자녀	61.4	동생	58.8	친구	61.8
엄마	60.9	이야기	57.8	아빠	61.2
모습	60.2	할머니	57.7	자식	59.7
할아버지	57.8	속	56.6	부부	58.3
속	57.7	시절	55.3	사랑	58.3
날	56.6	친구	55.2	생각	56.2
..

대상어 ‘아버지’의 경우 ‘아들, 어머니, 때, 딸, 가족’ 등의 순서로 빈번하게 함께 출현하며, ‘어머니’는 ‘아버지, 아들, 때, 아이, 딸’, 그리고 ‘아들’은 ‘아버지, 어머니, 딸, 아이, 집’ 등의 어휘와 자주 사용된다.

이런 공기어들과 대상어를 시각화한 네트워크는 아래 그림1과 같다. 그림1에 제시된 네트워크는 친족 명사 30개의 대상어가 각각 공기하는 어휘(노드로 표현)를 고빈도 순으로 50개씩 추출하여 전체 1500개의 어휘쌍을 만들어 친족명사 전체의 네트워크를 구성하고 있다.

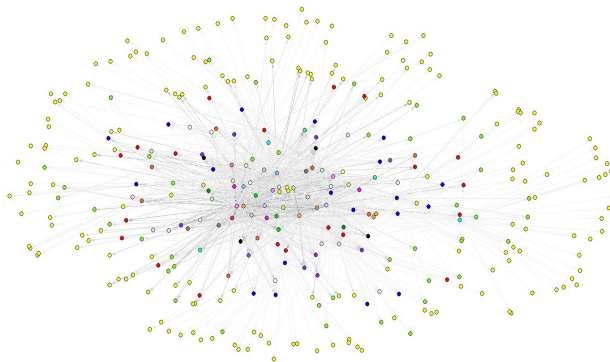


그림 1. 친족명사 30개의 공기어 네트워크

노드의 내향연결은 그 노드가 연결을 받는 링크의 수를 나타낸다(Nooy et. al, 2005). 이를 통해 하나의 노드에 들어오는 정도만 계산하는 내향연결정도(input degree, indegree)로 공기하는 어휘의 양상을 알아볼 수 있다. 그림1에서 한 단어는 평균 8.70개의 어휘와 관계를 맺으며 밀도는 약 0.01이다.

네트워크에서 밀도는 네트워크의 크기(데이터 크기)에 따라 차이를 보이므로 연결정도 중심성(degree centrality)을 통해 각 개별적 노드가 다른 노드와 연결을 많이 맺는 정도를 확인할 수 있다. 기본적으로 많은 연결을 보유한 노드가 연결망 내에서 핵심적 역할을 할 수 있고, 따라서 상대적으로 중요한 위치에 있다는 비교적 단순한 관계적 논리에 의해 계산되는 지표이다. 연결정도 중심성이 높은 노드는 동원할 수 있는 자원이나 정보를 많이 소유할 가능성이 높으므로 조직에서 중심적인 역할을 할 수 있다. 친족명사 30개의 연결정도 중심성은 0.067로 나타난다.

공기어 양상에 대한 지표는 다음 표6과 표7에서 자세히 파악할 수 있다.

표 6. 친족명사 30개의 공기어 연결정도

Input Degree partition
Dimension: 346
The lowest value: 0
The highest value: 30

표6에서 'Dimension: 346'는 네트워크에 나타난 노드(단어)의 수를 뜻하며, 'the lowest value: 0'은 어떤 단어가 다른 단어와 공기된 최소빈도이며 'the highest value: 30'은 최대 30개의 단어와 공기된 단어가 있음을 나타낸다. 친족명사 대상어휘 30개에 대해 각각 50개의 공기하는 어휘를 리스트로 작성하여 상호 연결관계를 나타내는 공기 네트워크 작성 결과 산술적으로 최대 1530개의 노드가 생길 수 있으나 단어 간에 공유하는 상호 공기가 있으므로 이 경우 346개의 노드가 생성되었다.

친족명사와 공기하는 구체적인 어휘는 표7에서 제시된다. 아래 표에서 'cluster'는 공기어와 함께 나타나는 대상어의 숫자로 공기어 '집'의 경우 30개의 친족명사와 함께 출현하고 있다. 세 번째 칸에 나오는 'representative'는 공기어의 예를 뜻한다. 29개의 친족명사와 공기하는 어휘로는 '가족', '딸'이며 28개의 친족명사와 공기하는 어휘는 '아이', '일', '때', 27개의 대상어와 공기하는 어휘는 '아들', '어머니' 등으로 나타난다. [물결 21] 코퍼스의 어휘 빈도 통계를 고려할 때 상위 20위 안에 드는 고빈도 명사인 '딸', '때', '전', '사람', '일' 등을16) 제외하면 공기어는 가족/친족과 관련된 어휘들이다. 연구 대상어 각각 친족명사라는 연관성

16) [물결 21] 코퍼스에서 '딸'(1위), '때'(3위), '사람'(6위), '일'(14위)로 상위 빈도 어휘이다.

이 공기되어지는 어휘에서도 같은 의미영역 내의 단어로 나타나고 있음을 시사한다. 이는 우리가 단어의 정의에 기반하여 개념 또는 직관적으로 설정하던 의미부류를 공기어 양상을 통해 경험적으로도 확인해 볼 수 있도록 한다.

이밖에 문단을 문맥 범위로 하여 추출한 공기어는 연상의 측면에서 우리가 직관적으로 호명하게 되는 단어 이외의 대상을 제시한다. 예를 들어 친족명사 30개와 모두 공기하는 명사는 우리가 쉽게 연상할 수 있는 ‘잡’인 반면, 26개의 대상어와 공기하는 단어는 ‘이야기, 사랑, 결혼’이고 25개의 대상어와 공기하는 어휘는 ‘사람, 마음, 엄마, 남편’이다. 이런 공기어들은 추출된 결과를 통해서 쉽게 이해할 수 있는 단어이지만 실제 우리가 생각하여 언급할 때는 종종 어려움을 겪기도 한다. 따라서 본 연구에서 제시되는 공기어는 의미적으로 연관이 있는 단어뿐만 아니라 우리의 연상 작용과 관련된 단어도 제시한다.

표 7. 30개 친족명사 전체 공기어 양상

Cluster	Freq	Representative
0	1	삼촌
1	190	위원, 사장, 화해, 상봉, 진상, 공개, 판결, 변호사, 유족, 국정원, 판사, 규명, 강화, 대회, 위원회, 진실, 연기, 장인,....
2	41	육아, 학생, 호칭, 가게, 누이, 선물, 조부모, 선생, 국적, 소녀, 사회,
3	17	고부, 직장, 어른, 사위, 동서, 자매, 사촌, 관계, 부모님, 꿈, 소년, 아기,...
4	11	음식, 여동생, 오빠, 시절, 기억, 시간, 소리, 선생님, 고향, 손녀, 몸
5	17	갈등, 시아버지, 사이, 식구, 친정, 시누이, 이름, 할아버지, 사망, 언니, 정, 막내, 행복, 형, 재산, 어린이, 시댁
6	7	혼자, 조카, 나이, 노인, 형제, 눈, 책
7	8	명절, 손자, 애, 돈, 눈물, 영화, 학교, 초등학교
8	6	주부, 손, 회장, 얼굴, 시어머니, 드라마
9	5	생활, 이혼, 삶, 숙, 공부
10	1	가정
11	2	학년, 날
12	1	며느리
13	1	모습
14	4	아빠, 자녀, 부인, 남
15	2	여성, 동생

16	2	뒤, 여자
17	3	남자, 부부, 세상
18	2	자식, 집안
19	1	전
20	3	친구, 아내, 할머니
21	2	부모, 생각
23	2	자신, 말
24	1	아버지
25	4	사람, 마음, 엄마, 남편
26	3	이야기, 사랑, 결혼
27	2	아들, 어머니
28	3	아이, 일, 때
29	2	가족, 딸
30	1	집

표7에서 1개의 대상어와만 공기하는 명사의 종류가 190개인 점과 30개의 대상어가 공유하는 공기어가 1개인 점은 이 네트워크의 밀도와도 관련이 된다. 1개의 대상어와 함께 나타나는 ‘배타적 공기어’가 많을수록 네트워크의 밀도는 약해진다.

연구대상이 되는 친족명사 30개 사이의 연결 관계를 알아보면 다음과 같다. 노드의 외향 연결은 그 노드가 다른 대상어와 연결되는 링크의 수를 나타낸다(Nooy et. al, 2005). 아래 그림에서 외향연결은 한 친족명사가 다른 친족명사의 공기어로 나타나는 정도를 알려준다.

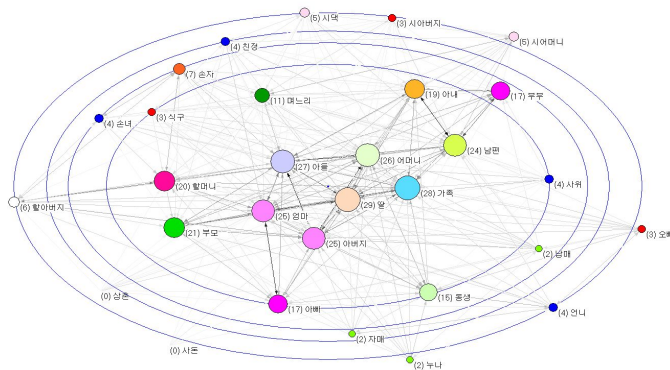


그림 2. 친족명사 30개의 상호 연결 관계

그림2에서 괄호 안에 제시되는 숫자는 각 단어가 다른 단어와 관계되는 빈도이다. 대상어는 그 빈도가 높을수록 전체 네트워크의 중심에 위치한다. 이것은 공기어의 t-점수와도 비례하는 관계이다.

표 8. 친족명사 29개와 ‘딸’의 공기 관계 지수(t-점수)

대상어	t-점수	대상어	t-점수	대상어	t-점수
아들	115.5	할머니	44.8	할아버지	23.4
아버지	110.9	사위	40.3	시어머니	22.6
엄마	109.6	동생	40.0	자매	22.1
어머니	94.3	며느리	36.8	식구	21.4
가족	81.1	손자	27.9	누나	16.9
남편	80.8	친정	27.5	사돈	16.3
아내	74.8	언니	27.3	시댁	14.9
아빠	69.1	손녀	25.9	삼촌	11.4
부모	61.1	남매	25.3	시아버지	10.2
부부	56.6	오빠	23.9		

동심원의 중심부에 있는 어휘들은 ‘딸, 가족, 아들, 어머니, 남편, 엄마, 아버지, 부모’로 나타난다. 친족명사 30개의 외향연결은 아래 표9에서 자세히 알 수 있다.

표 9. 친족명사 30개의 외향연결

외향연결수	대상어	외향연결수	대상어
29	딸	15	동생
28	가족	11	며느리
27	아들	7	손자
26	어머니	6	할아버지
25	엄마, 아버지	5	시댁, 시어머니
24	남편	4	사위, 손녀, 언니, 친정
21	부모	3	식구, 시아버지, 오빠
20	할머니	2	남매, 누나, 자매
19	아내	0	사돈, 삼촌
17	부부, 아빠		

상위 50개의 공기어를 관찰해 볼 때 ‘딸’은 29개의 친족명사의 공기어로 나타나는 반면 ‘사돈’과 ‘삼촌’은 다른 친족명사의 공기어로 나타나지 않는다. 이는 신문 기사에서 다루어지는 가족 관계에서 ‘딸’은 연구 대상이 되는 친족명사들과 하나의 사건이나 주제를 이야기할 때 자주 함께 사용이 된다고 볼 수 있다. 즉, 신문 기사에서 그려지는 우리 사회는 아들보다 딸이 상대적으로 더 많은 가족 구성원의 일과 관련되어 있다고 해석해 볼 수 있다. 이 해석은 종전 우리 사회가 ‘할아버지-아버지-아들’이 중심이 되는 가부장 중심이었다는 것과는 다른 양상을 보여준다. 실제 ‘할아버지’는 다른 친족명사의 상위 공기어로 나타나는 것이 6번에 불과하다. 이런 결과로 볼 때, 2000년부터 2009년 사이 가족 구성원이 언급되는 신문 기사의 중심에 있는 일원들은 딸, 아들, 아빠/아버지, 엄마/어머니가 구성원인 핵가족이 된다고 해석해 볼 수 있다.

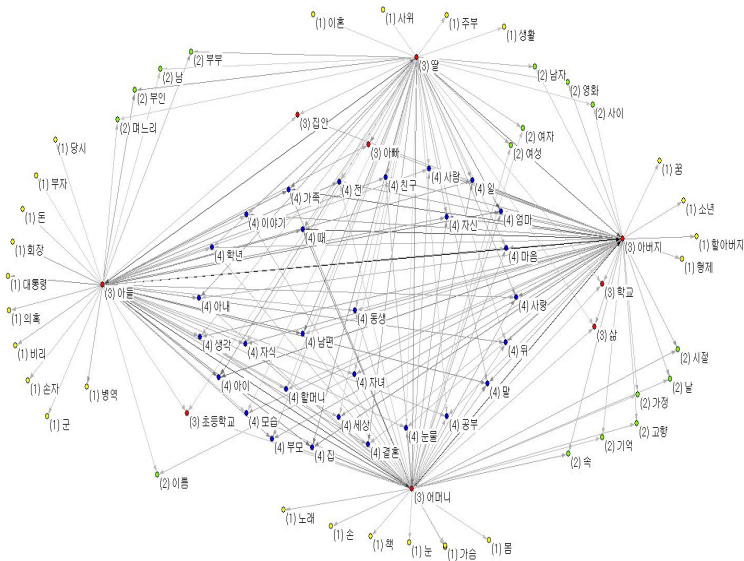
이러한 사실들은 네트워크 지표 가운데 중심성을 이용하면 보다 객관적으로 알 수 있다. 중심성은 네트워크상의 어떤 노드가 얼마만큼 중심 가까이 자리 잡고 있는지를 나타내는 정도로 특정한 노드가 많은 다른 노드들과 연결을 갖게 될 경우 그 노드는 네트워크의 가운데 쪽으로 위치하게 된다. 중심성 지표가 높은 상위 10개의 친족명사와 그 값은 다음과 같다.

표 10. 친족명사 전체 중심성 지표

친족명사	중심성 지표	친족명사	중심성 지표
딸	0.5520000	아버지	0.5118694
가족	0.5340557	부모	0.5051245
아들	0.5332303	엄마	0.5007257
어머니	0.5203620	며느리	0.4949785
남편	0.5156951	부부	0.4872881

표10에서 나타나듯이 ‘딸’은 전체 네트워크상에서 가장 높은 중심성을 갖고 있으며 다른 어휘에 대해 강한 관련성을 내포하고 있다. 네트워크상에서 중심이 된다는 것은 집단 내 영향력이 커지며(Brass, 1984) 조직의 의사결정에 영향을 미치는 효과가 커진다(Friedkin, 1993)는 것을 의미한다. 그림2와 표9와 표10에서 보듯이 최근 10년간 신문에서 기술되는 가족관계는 핵가족의 구성원 사이에 보다 강한 연결성을 갖고 있음을 보여준다.

다음으로 친족 명사 가운데 유의어나 관계어 등 의미에 연관성이 있는 어휘들을 분리하여 살펴보기로 하자. 우선 핵가족의 기본 구성원인 ‘아버지, 어머니, 아들, 딸’로 이루어지는 네트워크를 보면 다음과 같다.

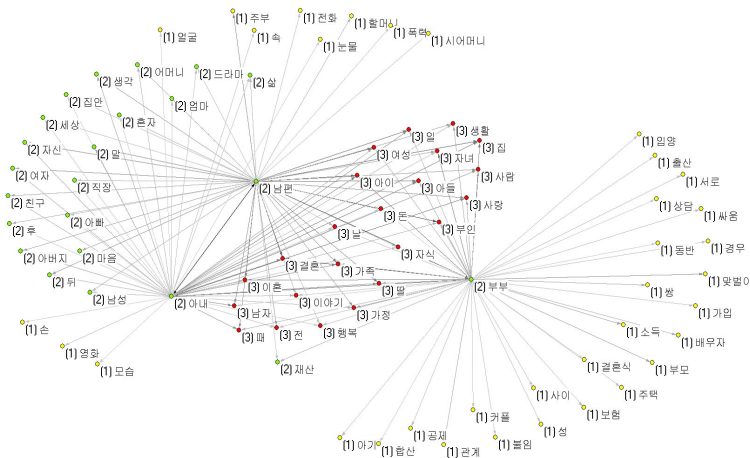


Density = 0.0324570 / Average Degree = 5.0632911
 그림 3. '아버지, 어머니, 아들, 딸'의 공기어 네트워크

그림3에 제시된 네트워크의 밀도는 약 0.03정도이며 하나의 어휘는 평균 5개 정도의 어휘와 연결이 되어있다. 친족명사 30개 전체의 네트워크보다는 밀도가 조금 강하게 나타났는데 이는 공유되어지는 공기어의 비율이 높고 동시에 배타적으로 나타나는 공기어의 비율은 상대적으로 낮기 때문이다.

핵가족 구성원 모두가 공유하는 공기어는 29개이며 '가족, 할머니, 엄마, 자녀, 남편, 아내, 자식, 아이, 동생' 등 실제 가족과 관련된 어휘들이거나 '사랑, 학교, 공부, 결혼, 마음' 등 가정 내에서 관심을 갖고 있는 어휘들이 동일 문단 내에서 사용되고 있다. 이런 공기 양상은 이전에 보여준 친족명사 30개 전체가 보여준 공기어보다 구체적으로 가정생활이라는 의미 영역으로 집중되고 있다. 또한 가정의 관심으로 나타나는 공기어는 우리의 심리적 연상 작용에 대해 실제 사용 맥락을 기반으로 추출된 단어들이다.

이밖에도 일반적인 친족을 나타내는 명사와 그 구성원에 대한 네트워크와 공기 양상을 알아보기 위해 '부부'와 그 구성원인 '남편'과 '아내'의 관계를 보면 다음과 같다.



Density = 0.0210084 / Average Degree = 3.5294118

그림 4. '부부, 남편, 아내'의 공기어 네트워크

'부부'와 그 구성원 친족명사 네트워크의 밀도는 약 0.02로 상호 공기어가 적게 나타나며 이는 평균 연결정도로도 알 수 있다. 그림3의 '아버지, 어머니, 아들, 딸'의 네트워크는 평균 연결정도가 약 5.06로 그림4의 네트워크보다 강한 밀도를 갖고 있다.

'부부, 남편, 아내'의 상호 공기어는 [3]으로 표시되는 23개의 어휘인 '결혼, 자녀, 아들, 행복, 이혼, 사랑, 가정, 일, 때, 남자, 아이, 부인, 돈, 자식, 여성, 날, 딸, 가족, 사람, 이야기, 집, 생활' 등이다. 이 어휘들은 핵가족의 구성원과는 다른 공기어 양상으로 부부의 직접 관심사로 볼 수 있다. 대상어 3개의 상호 공기어, cluster 3에 나타나는 공기어들은 '부부'라는 어휘를 의미관계상 부분어(meronym)인 '남편'과 '아내'의 상호 관계 속에서 다른 친족명사와 구별시키는 역할을 한다.

다음으로 시계열 친족명사의 공기어 양상을 보도록 한다. '시아버지, 시어머니'로 대표되는 시계열 친족명사와 관련 구성원인 '며느리'와 '아들'의 네트워크와 공기어는 다음과 같다.

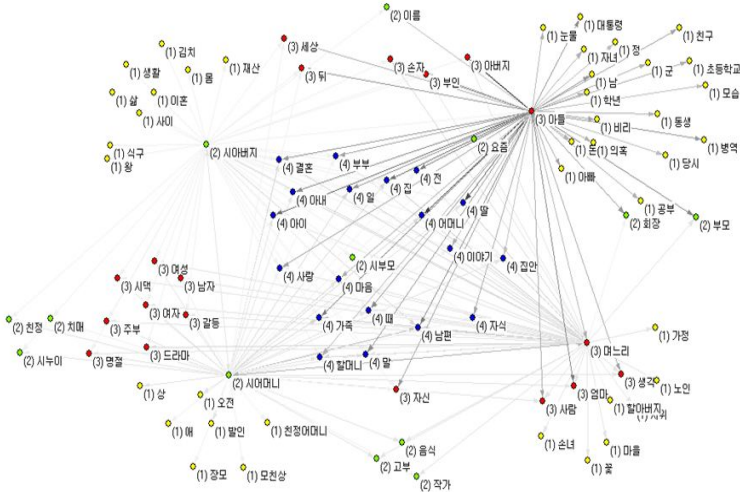
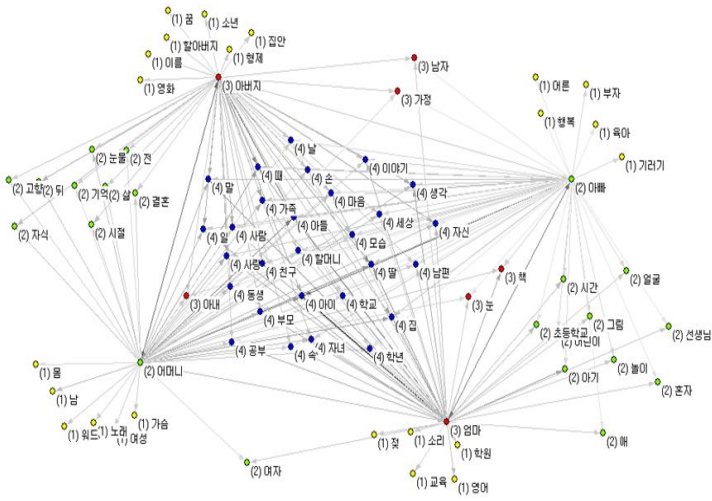


그림 5. '시아버지, 시어머니, 며느리, 아들'의 공기어 네트워크

시계열 어휘와 관련된 상호 공기어로는 '결혼, 부부, 집, 집안, 어머니, 남편, 아내' 등의 어휘가 있다. 그림5의 대상어들은 결혼이라는 관계로 맺어지는 가족관계로서 기타 친족명사에서 공통으로 나타나는 공기어를 배제한다면 '시아버지'와 '며느리'의 상호 공기어가 '시부모'라는 점이 주목된다. 시계열이라는 의미군집에서 '남편'과 '아내'의 상호 공기어는 '부모'이며 '시어머니'와 '며느리'의 관계가 '고부'라는 어휘로 명시되는 점은 우리의 직관이 네트워크 상에서도 확인된다. '시아버지, 시어머니, 엄마'의 상호 공기어로 주목할 만한 단어는 '갈등, 명절, 주부, 시댁' 등이다. 이 공기들은 의미적으로 연관된 즉, 가족과 관련된 단어라기보다 우리의 심리적 연상 작용의 연관어로 볼 수 있다.

지금까지 친족명사 내부에서 선정되는 대상어에 따라 다른 공기 양상을 살펴보았는데 이를 바탕으로 유의어의 공기어 양상을 알아보도록 하겠다. 문맥은 한 어휘의 의미를 분명하게 하는 기능 외에 문맥에서 공기하는 어휘를 살펴봄으로써 유의어를 구별하게 해 준다. 친족명사 가운데 많은 어휘들이 호칭어로도 사용되는 한편 다른 단어로 대치되는 경우가 있다. 그 대표적인 예가 '엄마와 '어머니', 그리고 '아빠'와 '아버지'이다.



Density = 0.0316456 / Average Degree = 5.0000000

그림 6. ‘아빠, 아버지, 엄마, 어머니’의 공기어 네트워크

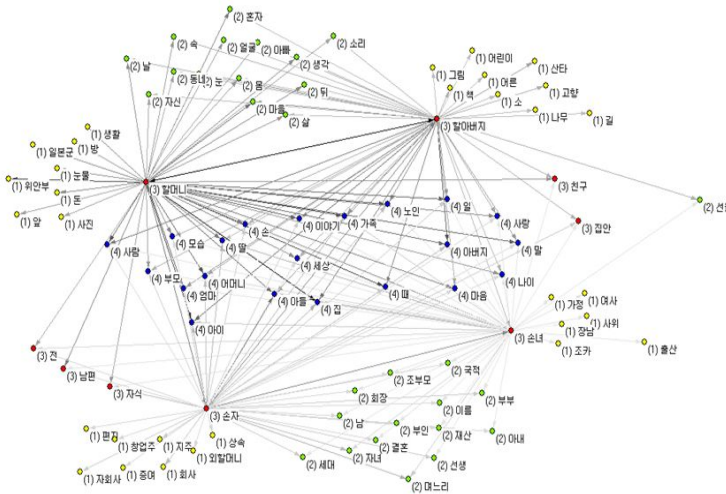
Nida(1973)은 어떤 형태의 언어적 의미를 결정하기 위해서는 대조점이 발견되어야 하며 변별적 차이와 무관한 의미는 없다고 주장한다. ‘아버지’가 지니고 있는 의미는 그 중심의미가 친족명사로서 동일한 의미영역에 나타나는 다른 형태들이 지니는 관계의미와 대조되고 일정한 의미 양상을 공유한다는 점에서 그 의미가 부각된다. 이러한 설명은 그림6에서 제시되는 네트워크와 표11의 공기어 양상으로 다시 확인할 수 있다.

표 11. ‘아버지, 어머니, 아들, 딸’의 공기어 양상

대상어	공기어
아빠	어른, 부자, 행복, 기러기, 육아
아버지	집안, 형제, 할아버지, 이름, 소년, 꿈
엄마	젖, 학원, 영어, 교육, 소리
어머니	몸, 남, 여성, 가슴, 노래
아버지, 어머니	눈물, 고향, 자식, 기억, 시절, 삶, 결혼
아빠, 엄마	초등학교, 아기, 애, 얼굴, 어린이, 그림, 선생님, ...
아버지, 아빠	남자, 가정
어머니, 엄마	여자

일반적으로 ‘아버지’의 유의어로 가장 쉽게 생각하는 단어가 ‘아빠’이며 ‘어머니’에 대해서는 ‘엄마’를 떠올린다. ‘아빠’와 ‘아버지’ 사이에 공유되는 단어는 ‘남자, 가정’이며 ‘엄마’와 ‘어머니’의 상호 공기어는 ‘여자’ 밖에 없다. 오히려 ‘아버지’와 ‘어머니’의 상호 공기어나 ‘엄마, 아빠’의 상호 공기어가 더 많이 나타나는 경향이 있다. 이 점은 실제로 이들 어휘가 함께 사용되는 장르나 문맥의 차이에 기인한다고 볼 수 있다¹⁷⁾.

다음으로 ‘할아버지, 할머니, 손녀, 손자’의 네트워크를 살펴보면 대상어 한 개당 평균 4개의 단어와 연결이 되어 있다.



Density = 0.0249688 / Average Degree = 4.4444444
 그림 7. ‘할머니’, ‘할아버지’, ‘손자’, ‘손녀’의 공기어 네트워크

네 개의 대상어가 상호 공기하는 어휘로는 가족관계 중 아버지나 어머니를 통한 조손관계임을 짐작할 만한 ‘아들, 딸, 노인, 가족, 아버지, 부모, 엄마, 아이, 집, 어머니’ 등이 있다. ‘할머니’와 공기하는 ‘위안부, 일본군’ 등은 다른 대상어와 의미적으로 공유될 수 없는 배타적 공기어의 특징을 잘 나타내 준다. ‘손자’의 경우 경제와 관련된 ‘창업주, 지주, 상속, 증여, 자회사, 회사’ 등의 어휘가 배타적 공기어로 등장하는데 이는 지난 10년간 경제 기사에 ‘손자’와 관련된 일들이 많았던 언어 외적인 현상으로 해석할 수 있다.

17) 강범모(2003)에서 유의어가 장르에 따라 달리 사용됨을 밝혔다.

5. 결론

신문 기사에 나타나는 친족명사와 공기어들의 의미적 관련성을 네트워크로 분석하였다. 같은 주제를 기술하고 있는 하나의 문단 내에서 우리 사회의 구성원인 친족명사 30개와 함께 출현하는 명사들을 추출하고 그 단어들의 상관관계를 t-점수로 계산한 후 도출된 단어 쌍들을 네트워크로 시각화하였다. 공기어 추출의 문맥 범위를 문단으로 설정한 것은 신문 기사가 언급하는 생활 또는 현실 세계에서 어떤 단어가 가리키는 대상과의 연관성을 보여준다. 예를 들어 친족명사 중 26개의 대상어와 공기하는 단어는 ‘이야기, 사랑, 결혼’이고 25개의 대상어와 공기하는 단어는 ‘사람, 마음, 엄마, 남편’이다. 그리고 유의어의 공기어 양상은 현실 세계에서 대상어가 기술되고 있는 상황, 맥락, 그리고 레지스터가 반영됨을 보여 주었다. ‘아빠’와 ‘아버지’가 그리고 ‘엄마’와 ‘어머니’가 유의어이지만 이 단어들이 공유하는 공기어의 양상은 ‘아빠’와 ‘엄마’가 ‘아빠’와 ‘아버지’보다 더 많은 명사와 공기하고 있다. 네트워크를 통해 살펴본 친족명사 30개의 경우 ‘딸 > 가족 > 아들 > 어머니 > 엄마, 아버지 > 남편 >... > 남매, 누나, 자매 > 사돈, 삼촌의 순서대로 연구 대상이 되는 다른 친족명사와 관계를 형성하고 있다.

이 연구의 의의는 다음과 같다. 첫째, 대상어와 공기 관계로 나타나는 단어들은 의미적으로 관련된 단어 외에 연상의 연관성이 드러나는 단어들이 함께 나타난다. 이런 연상의 연관성은 우리가 언어에 대해 의미를 형성해 가는 과정에서 우리를 둘러싸고 있는 사회적 심리적 근거를 제시한다. 즉, 기존의 공기어 기반 연구들이 단어 고유의 의미간의 관계에 중점을 둔 의미적 연관성을 주된 관심으로 둔 반면, 이 연구는 실제 사용 맥락을 중심으로 단어들의 관계를 제시하여 언어 사용자를 둘러싸고 있는 사회적, 문화적, 심리적 요인을 근거로 한 공기어를 보여준다. 이는 하나의 주제나 사건을 기술하는 신문 기사를 대상으로 한 문단 내에서 추출된 공기어는 단순히 우리가 개념적 또는 직관적으로 생각할 수 있는 단어보다 실제 상황이나 레지스터에 경험적, 의미적으로 관련된 다양한 단어들로 확인한다. 본 연구에서 제시된 공기어는 단순한 개념적 의존성이라기보다, 공기적 의존성에 기반하여 사회와 문화의 모습을 반영하는 언어 속에서의 상호 연관성을 보여 준다. 둘째, 시각화된 네트워크를 통해 우리는 여러 개의 대상어와 함께 공유되는 공기어나 특정 대상어와 나타나는 공기어를 쉽게 알 수 있다. 그리고 특정한 형태의 네트워크 안에 있는 단어의 위치로 그 단어가 다른 단어와 어떠한 관계를 갖고 있는지 파악할 수 있다. 즉, 단어들의 상호 관계 속에서 의미부류, 군집 형성 그리고 유의어의 공기어 사용 양상을 파악할 수 있다. 셋째, 네트워크의 여러 분석 지표를 통해 단어와 관련된 우리의 직관을 객관적으로 확인할 수 있었다.

참고문헌

- 강범모, 김홍규, 허명희. (2000). *한국어의 텍스트 장르, 문체, 유형*. 서울: 태학사.
- 강범모. (2003). *언어, 컴퓨터, 코퍼스 언어학: 컴퓨터를 이용한 국어 분석의 기초와 응용*. 서울: 고려대학교 출판부.
- 강범모. (2010). 공기 명사에 기초한 의미/개념 연관성의 네트워크 구성. *한국어의미학* 32, 1-29.
- 고길근. (2007). 정책 네트워크 연구의 유용성과 사회연결망 이론 활용 방법의 고찰. *행정논총*, 45(1), 137-164.
- 김용학. (2007). *사회 연결망 분석*, 서울: 박영사.
- 김응모. (1989). *국어 평행이동 자동사 낱말발*. 서울: 한신문화사.
- 김응모. (2000). *일상언어 자동사 낱말발*. 서울: 한국문화사.
- 김일환, 이도길, 강범모. (2010a). SJ-RIKS Corpus: 세종 형태의미 분석 코퍼스를 넘어서. *민족문화연구*, 52, 373-403.
- 김일환, 이도길, 강범모. (2010b). 공기 관계 네트워크를 이용한 감정명사의 사용 양상 분석. *한국어학*, 49, 119-148.
- 김지형. (1999). 어휘장(낱말발) 연구의 새로운 모색. *한국어의미학*, 4, 281-311.
- 김홍규 외. (2010). [물결 21] 신문 텍스트 기반의 장기간 언어·사회·문화 연구: [물결 21] 사업 제1차 보고서. 서울: 고려대학교 민족문화연구원.
- 박병선. (2005). *한국어 계량적 연구 방법론*. 서울: 도서출판 역락.
- 배해수. (1982). 맛 그림씨의 낱말발. *한글*, 176, 67-91.
- 배해수. (1989). 낱말발 분석 연구의 의의. *한글*, 203, 213-231.
- 안귀남. (2005). 동성마을 친족호칭어의 사회언어학적 연구. *방언학*, 제2호, 1-31.
- 윤영수, 채승병. (2005). *복잡계개론*. 서울: 삼성경제연구소.
- 이성현. (2004). 세종전자사전의 명사 의미부류체계. *세종전자사전 개발계획 학술 발표회 논문집*.
- 최경봉. (1998). *국어 명사의 의미 연구*. 서울: 태학사.
- 최규일. (1986). 한국어의 친족호칭 어휘 연구. *국어교육*, 55, p51-85.
- 한정한, 고석주, 김진해, 이동혁. (2008). *한국어 어휘의미망 구축을 위한 기초 연구*. 서울: 보고서.
- 홍사만. (2008). *국어 의미 분석론*. 서울: 한국문화사.
- 황도삼, 최기선, 김태석. (1999). *자연언어이해: 컴퓨터과학시리즈*. 서울: 홍릉과학출판사.
- Brass, D.J. (1984). Being in the right place: A structural analysis of individual influence in an organization. *Administrative Science*

- Quarterly*, 29, 518-539.
- Bhattacharya, S. & Basu, R. K. (1998). Mapping a research area at the micro level using co-word analysis. *Scientometrics*, 43(3), 359-372.
- Callon, M., Courtial, J. P. & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22(1), 153-203.
- Church, K. W.; Gale, William; Hanks, Patrick; Hindle, Donald (1991). Using statistics in lexical analysis. In L. Erlbaum. *Lexical Acquisition: Using On-line Resources to Build a Lexicon* (pp. 115-164).
- Collins, A. M., & Quillian, M. R. (1972). Experiments on Semantic Memory and Language Comprehension. In L. W. Gregg (Ed.), *Cognition in Learning and Memory* (pp117-138). New York: John Wiley.
- Collins, A. M., & Loftus, E. F. (1975). A Spreading Activation Theory of Semantic Processing. *Psychological Review*, 82, 407-428.
- Freidkin, N. E. (1993). Structural Bases of Interpersonal Influence in Groups: A Longitudinal Case Study. *American Sociological Review*, 58(6), 861-872.
- Ide, N. & Veronis, J. (1998). Introduction to the Special Issue on Word Sense Disambiguation: the State of Art. *Computational Linguistics*, 24(1), 1-39.
- Ke, J. (2007). Complex networks and human language. arXiv:cs/0701135.
- Lyons, J. (1977). *Semantics. Volumes I and II*. Cambridge: Cambridge University Press.
- Miller, G.A. (1995). *WordNet: A Lexical Database for English*. Cambridge: MIT press.
- Newman, Mark E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45, 167-256.
- Newman, Mark E. J. (2004). Analysis of weighted networks. *Physical Review*, E 70, 056131.
- Nida (1975). *Language Structure and Translation: Essays*. Stanford University Press.
- Nooy, Wouter d., A. Mrvar and Vladimir Batagelj. (2005). *Exploratory Social Network Analysis with Pajek*. Cambridge Univ. Press.

- Nooy, W. d., A. Mrvar and V. Batagelj. (2010). vlado.fmf.uni-lj.si/pub/networks/pajek/.
- Ogden, C. K. and Richards, I.A. (1923). *The Meaning of Meaning*. London: Routledge and Kegan Paul.
- Scott, J. P. (2000). *Social Network Analysis: A Handbook*. Thousand Oaks, CA: Sage Publications Ltd.
- Spence, D. P. and Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, 19(5), 317-330.
- Wasserman, S, and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Whorf, B.L. (1956). *Language, thought, and reality: Selected writing of Benjamin Lee Whorf*. J.B. Carroll(ed.). MIT Press.
- Wolfram, S. (2002). *A New Kind of Science*. Champaign, IL: Wolfram Media.
- Veling, A. & P. van der Weerd. (1999). Conceptual grouping in word co-occurrence networks. In *Proceedings of the IJCAI '99* (pp 694-699). Volume 2. San Francisco: Morgan Kaufmann Publishers.

정유진

고려대학교 민족문화연구원 HK연구교수
136-701 서울 성북구 안암동 5가
전화: (02)3290-5281
e-mail: echung2@korea.ac.kr

강범모

고려대학교 언어학과
136-701 서울 성북구 안암동 5-1
전화: (02)3290-2173
e-mail: bmkang@korea.ac.kr

Received on 23 March, 2011

Revised on 6 June, 2011

Accepted on 6 June, 2011