

# 세종말뭉치 문어자료를 활용한 정도부사의 분류와 그 사용양상 - 용언과의 2원 네트워크 분석 -

정성훈\*

---

**Jung, Sung-hoon. The Classification and Use Patterns of Korean Degree Adverbs in Sejong Corpus -Two-mode Network Analysis-**. In this paper I reviewed theoretical views of degree adverbs in modern Korean Language and suggest a proper statistical method for clustering degree adverbs using Korean predicates. So far, the Korean degree adverbs have been classified by the intuitions of researchers, therefore, there have been difficulties that the clusters of degree adverbs are classified differently depending on the intuitions of researchers. In this study, I classify the Korean degree adverbs by applying the two mode network analysis in which I set G-value as weights that is a kind of maximum likelihood statistical significance. The two-mode network analysis is considered to be appropriate for analyzing the relations between Korean adverbs and predicates because its basic assumptions are similar to them of structural linguistics, and for showing the relations between them visually and intuitively. Through this study we could find out new ways to deviate from the traditional classification in degree adverbs.

*Keywords: Sejong Corpus, Degree Adverbs, Network Analysis, Two mode, Dendrogram*

---

## 1. 서론

전통적으로 정도부사는 형용사를 수식 한정하여 형용사의 <정도성>을 나타내는 것으로 기술한다. 그런데 <정도성>이라는 것이 매우 추상적이다. 예를 들어 ‘너무 좋다’, ‘아주 좋다’, ‘매우 좋다’를 생각해 보자. 한국어 화자들은 ‘너무, 아주, 매우’의 <정도성>의 차이를 구분할 수 있을까? (1 가, 나, 다) 중에서 제일 좋은 날씨는 어느 것일까? 연구자의 언어 직관으로 ‘너무, 아주, 매우’의 <정도성>을 구분하는 것은 쉬운 일이 아니다. 게다가 위의 세 정도부사의 관계를 추정하는 것도 쉽지 않다. ‘너무’가 ‘아주’와 더 유사한 의미를 갖는지, ‘매우’와 더 유사한 의미를 갖는지 판단하기 쉽지 않은 것이다.

- (1) 가. (날씨가) 너무 좋다.
- 나. (날씨가) 아주 좋다.
- 다. (날씨가) 매우 좋다.

---

\* 대만국립정치대학교 한국어문학과 초빙 조교수(kobe99@nccu.edu.tw)

본 연구는 말뭉치에 나타난 정도부사(degree adverbs)의 사용빈도를 분석하고, 문어자료에서 공기하는 정도부사와 용언의 관계적 속성을 파악하여 정도부사들을 계량적으로 의미 분류하는 데 그 목적이 있다. 지금까지의 정도부사 연구는 연구자의 직관에 의존한 연구가 대부분이었으며, 각 연구자의 정도부사에 대한 정의가 유사함에도 불구하고 연구자의 언어직관에 따라 정도부사의 유형이 다소 상이하게 나타났다. 본 연구에서는 기존 연구에서 밝힌 연구자들의 정도부사를 살펴보고, 이를 바탕으로 계량적 분석을 시도한다. 이를 위해서 말뭉치에 나타난 정도부사의 유형과 빈도의 분포를 수치화하고 정도부사와 용언의 공기빈도와 계량적 분포가 나타내고 있는 의미를 해석한다. 이는 공기하는 정도부사와 용언은 일정한 의미를 공유한다는 가정에서 출발하여 말뭉치에서 빈도가 높게 공기하는 정도부사와 용언들의 관계는 공기하는 다른 정도부사와 용언에 비해서 그 속성이 강하게 결합되어 있다는 것을 기초로 한다.

본 연구에서는 공기하는 정도부사와 용언의 관계적 속성을 살펴보기 위해서 정도부사와 용언의 2 원 네트워크 분석을 수행한다. 초창기 네트워크 분석은 구조주의와 그래프 이론이 결합한 분야였다. 네트워크 분석의 기본 가정은 언어학에서 가정하는 구조주의와 매우 유사하기 때문에 정도부사를 분석하기에 적합하다. 정도부사와 용언의 2 원 네트워크 분석은 각 정도부사의 의미와 기능을 개별적으로 분석하는 것이 아니라 공기하는 용언과 정도부사의 네트워크를 구축하여 이것을 이용하여 정도부사 간의 관계적 속성으로 분석한다. 이 네트워크 분석은 개별 정도부사가 아니라 정도부사의 전체적인 구조를 고려한다.

## 2. 세종말뭉치와 정도부사

### 2.1. 정도부사 유형에 대한 기존 연구

다음은 기존 연구들에서 다뤘던 정도부사의 유형이다.<sup>1</sup>

#### (1) 정도부사의 유형

<sup>1</sup> 연구자들마다 상이한 정도부사의 유형을 제시하고 있다. 공통으로 제시한 정도부사의 수는 30 여 개 정도이고, 연구자들이 제시한 정도부사의 유형을 모두 포함하면 100 여 개에 이른다.

- ① 서상규(1991): 가장, 거의, 거진, 겨우, 고작, 과히, 굉장히, 그다지, 근, 기껏, 꼭, 꽤, 끔찍하게, 다만, 다소, 단, 대단히, 더, 더욱, 덜, 도무지, 되게, 딱, 딱, 많이, 매우, 몹시, 무려, 무지, 무지하게, 무척, 별로, 보다, 불과, 비교적, 사무, 수없이, 심히, 썩, 아무리, 아주, 약, 약간, 얼마, 얼마간, 여간, 영, 오죽, 완전히, 워낙, 월등히, 유난히, 유독, 일층, 잔뜩, 전혀, 정말 제법, 제일, 조금, 지독하게, 참말, 참으로, 크게, 터무니없이, 특히, 펍, 하도, 한결, 한껏, 한층, 훨씬
- ② 손남익(1995): 가뜩이나, 가우, 가장, 겨우, 고까지로, 고다지, 그지없이, 극히, 근근히, 꽤, 너무, 너무나, 더군다나, 더더구나, 더더군다나, 더더욱, 더욱더, 더욱, 더한층, 덜, 마구, 매, 매때, 매우, 몹시, 무장, 무척, 보다, 불과, 썩, 아주, 영판, 조금, 좀더, 지지리, 펍이나, 훨씬
- ③ 서정수(2005) : 가장, 게다가, 결코, 과히, 굉장히, 그다지, 그리, 극히, 꽤, 너무, 너무나, 대단히, 더, 더구나, 더욱, 더욱이, 덜, 도무지, 도저히, 딱, 마치, 매우, 몹시, 무척, 별로, 사뭇, 상당히, 심히, 아득히, 아스라이, 아주, 워낙, 유난히, 유달리, 자못, 작히, 적이, 전연, 전혀, 절대로, 정히, 제법, 조금도, 좀, 좀처럼, 지극히, 참, 참으로, 최고로, 최대로, 통, 펍, 하도, 한결, 한층, 훨씬, 휘영청
- ④ 남기심 외(2005) : 가장, 거의, 게다가, 그다지, 그리, 극히, 꽤, 대단히, 더, 더구나, 더욱, 더욱이, 덜, 매우, 몹시, 무척, 별로, 사뭇, 상당히, 심히, 아주, 약간, 여간, 유난히, 유달리, 이다지, 이리, 이만, 자못, 저다지, 저리, 저만, 전혀, 제법, 조금, 좀, 좀더, 지극히, 최고로, 최대로, 통, 펍, 하도, 한결, 한층, 훨씬

그런데 위의 연구들에서는 정도부사의 유형이 사뭇 다르다. 최현배(1971)와 허웅(1995)의 정의와는 크게 다르지 않지만, 연구자에 따라 구체적인 정도부사의 유형은 큰 차이를 보인다.<sup>2</sup>

## 2.2. 세종말뭉치 문어자료에 나타나는 정도부사

정도부사는 기본적으로 형용사를 수식 한정하는 빈도가 매우 높다. 따라서 세종말뭉치 문어자료에서 형용사와 공기하는 부사를 모두 추출하였다. 또한 정도부사가 형용사와만 공기하는 것은 아니기 때문에 동사와 공기하는 부사 중에서 기존 연구에서 언급된 부사들을 모두 추출하였다.<sup>3</sup>

이렇게 추출된 부사에는 ‘되게되게’ 나 ‘너무너무’ 등과 같이 중첩으로 구성된 부사들이 있었다. 이들 부사는 사전의 등재여부에 따라 사전에 등재되어 있으면 중첩된 하나의 부사로 간주하고 그렇지 않으면 수식하는 부사와 수식 받는 부사로 나누었다. 즉 말뭉치에서는 하나의 부사로 태그된 ‘되게되게’ 등과 같은 부사는 두개의 정도부사로 간주하고, 이를 수식부사와 피수식부사로 구분하였으나 ‘너무너무’ 등과 같이 사전에 등재된 정도부사는 따로 분리하지 않고 하나의 정도부사로 분석하였다.

한편 세종말뭉치에서는 ‘조금도(조금+도)’ 처럼 형태 분석이 가능한 정도부사도 하나의 부사로 태그되어 있었는데, 본 연구에서는 ‘조금’ 을 기본 정도부사 유형으로 간주하여 ‘조금+도’ 로 분석하였다. 그러나 ‘너무나’ 처럼 하나의 부사로 사전에 등재된 것들은 예외로 하였다. 또한 ‘지나치게(지나치+게)’ 처럼 하나의 품사가 아니라 파생된 부사어는 이 연구대상에서 제외하였다.

본 연구에서는 ‘되게, 정말, 진짜’ 등과 같은 정도부사들을 포함하였다. ‘되게, 정말, 진짜’ 등과 같은 정도부사들은 개별적인 정도부사 연구에서는 논의된 바 있으나 정도부사를 전체적으로 다룬 연구들에서는 그 유형이 제외된 경우가 많았는데, 본 연구에서는 이를 포함하여 분석하였다.

<sup>2</sup> 대부분의 연구자가 정도부사에 대해 유사한 속성을 제시하고 있으나, 연구자의 언어 직관마다 정도부사의 유형이 상이한 점이 특이하다. 위에서 살펴본 연구들에서 연구자들이 제시한 정도부사들을 모두 포함하면 100 여 개가 되는데 이 중 공통된 정도부사의 수는 30 여 개 정도이다.

<sup>3</sup> 본 연구에서는 정도부사와 용언의 관계적 속성에 주목하기 때문에 정도부사가 부사나 명사를 수식 한정하는 경우는 제외하였다.

그리고 본 연구는 말뭉치에 나타나는 정도부사를 분석하기 때문에 정도부사의 변이형을 다루는 문제가 중요하다. 말뭉치에서 ‘조금, 조금, 조금, ..’ 과 같은 변이형은 대표형 ‘조금’ 으로 분류하였으나 ‘좀’ 과 같이 음운 축약이나 음절 축약이 된 변이형들은 다른 유형의 정도부사로 분석하였다.

### (2) 정도부사 유형 선정 원칙

- ㄱ. 중첩부사는 수식부사와 피수식부사로 분류 (사전에 등재된 부사는 예외)
- ㄴ. 형태분석이 가능한 부사는 기본 유형으로 분류 (사전에 등재된 부사는 예외)
- ㄷ. 변이형은 기본적으로 같은 유형으로 분류
- ㄹ. 음운 축약이나 음절 축약 변이형은 다른 유형으로 분류

기존 연구들에서 공통으로 제시된 정도부사의 유형을 최대한 포함하고, 위의 (2)에서 제시된 원칙을 고려하여 본 연구의 대상을 선정하였다. 이러한 점을 모두 고려한 결과, 47 개의 정도부사가 추출되었으며, 그 유형은 다음과 같다.

### (3) 본 연구의 정도부사 유형

가장, 거의, 극히, 꽤, 너무, 너무나, 너무너무, 다소, 대단히, 더, 더욱, 덜, 되게, 디따, 많이, 매우, 몹시, 무척, 별로, 보다, 비교적, 상당히, 심히, 썩, 아주, 엄청, 열라, 완전, 워낙, 유난히, 유달리, 은근히, 전혀, 절대, 정말, 제법, 제일, 조금, 존나, 좀, 지극히, 진짜, 참, 충분히, 퍽, 하도, 훨씬

## 2.3. 정도부사의 빈도와 분포

이 절에서는 47 개의 정도부사가 세종말뭉치의 문어자료에서 어떻게 나타나는지 그 사용빈도와 분포를 살펴본다.

문어자료에서는 실제 47 개의 정도부사 유형 중에서 43 개의 정도부사 유형이 총 84,223 번 나타났다. 문어자료에서 전체 부사가 6,032 개, 총 647,046

번 나타난 것을 고려하면 문어에서는 극소수의 정도부사가 고빈도로 사용되고 있음을 알 수 있다.<sup>4</sup> 문어자료의 정도부사 유형과 빈도는 <표 2-1>과 같다. ‘더, 가장, 좀, 많이, 너무’ 등의 순으로 빈도가 높았고, ‘제일, 되게, 너무너무, 엄청, 심히’ 등의 정도부사는 빈도가 매우 낮았고, 정도부사 ‘완전, 열라, 존나, 디따’ 는 문어자료에서 나타나지 않았다.<sup>5</sup>

표 2-1. 문어자료에서 정도부사 유형과 빈도

빈도순 위	정도부사	빈도
1	더	16,303
2	가장	8,650
3	좀	5,990
4	많이	5,980
5	너무	5,058
6	더욱	4,792
7	거의	3,925
8	아주	3,508
9	정말	3,351
10	전혀	2,769
...	...	...
34	은근히	319
35	씩	260
36	진짜	211
37	비교적	88
38	유달리	66
39	심히	59
40	엄청	42

<sup>4</sup> 세종말뭉치 문어자료에 나타나는 정도부사의 유형은 전체 부사 유형 중 약 0.7%로 그 비율이 매우 작은데 비해서 그 빈도는 전체의 약 13%로 매우 자주 사용됨을 알 수 있다.

<sup>5</sup> 정성훈(2014)은 세종말뭉치 구어자료와 문어자료에서 고빈도로 나타나는 부사를 제시하고 있는데, 공통으로 나타나는 고빈도 정도부사는 ‘좀, 많이, 더, 너무, 거의, 정말, 너무너무’ 등이었고, 공통으로 나타나는 저빈도 정도부사는 ‘심히, 유달리, 씩, 비교적’ 등이었다. 한편 구어자료에서 자주 나타난 정도부사는 ‘되게, 진짜, 제일, 열라, 존나, 디따’ 등이었고, 문어자료에서 자주 나타난 정도부사는 ‘더욱, 무척, 매우, 몹시, 보다’ 등이었다.

41	너무너무	40
42	되게	35
43	제일	3

세종말뭉치 문어자료에서 정도부사 빈도순위와 빈도는 반비례 관계를 보였다. 정성훈(2014)은 전체 부사의 빈도순위와 빈도가 반비례 관계임을 보인 바 있는데, 부사의 하위 분류인 정도부사의 빈도순위와 빈도는 반비례 관계였다.<sup>6</sup> <그림 2-1>에서 정도부사의 빈도순위와 사용빈도가 반비례하는 것을 확인할 수 있다. 그래프의 기울기가 완만한 형태이나, 그래프의 형태는 전체적으로 L자 곡선을 띤다.

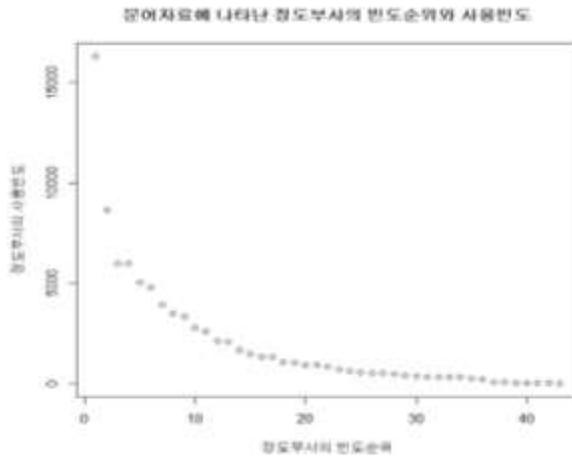


그림 2-1. 문어자료에서 정도부사의 빈도순위와 사용빈도

### 3. 정도부사와 네트워크 분석

3 장에서는 용언과 정도부사의 공기관계를 통해 정도부사 간 관계적 속성을 파악한다. 관계적 속성을 살펴보기 위해서 용언과 정도부사의 2원 네트워크를 구축하고 이를 분석한다. 2원 네트워크 분석(two-mode network analysis)을 통해 정도부사에 대한 개별적 속성뿐만 아니라 관계적 속성을 살펴볼 수 있다. 정도부사의 네트워크 구조를 통해 정도부사의 연결 상태나

<sup>6</sup> Mandelbrot(1953)는 이를 ‘프랙탈 구조(fractal structure)’라고 하였다. 프랙탈 구조는 작은 구조가 전체 구조와 닮은 형태로 되풀이되는 구조이다. 위의 현상이 언어구조에서 일반적인지는 좀더 검토해 보아야겠지만, 한국어 부사에서는 프랙탈 구조를 띠는 것으로 보인다.

연결 구조의 특성 등을 계량적으로 분석하고, 이를 시각적으로 표현한다. 네트워크 분석은 정도부사의 의미정보를 직관적으로 분석하는 데 그치지 않고, 정도부사와 그 의미들을 공간적으로 구조화함으로써 정도부사와 용언의 공기관계 및 정도부사 간 관계적 속성을 정량적으로 파악할 수 있게 한다. 즉 정도부사의 관계를 공간적인 구조로 보여주기 때문에 정도부사의 관계를 시각적으로 해석할 수 있는 가능성을 보여준다.

### 3.1. 2원 네트워크 분석의 기본가정

정도부사와 용언의 2원 네트워크는 공기하는 정도부사와 용언의 의미가 매우 밀접한 관계를 갖는다고 가정한다. 정도부사는 공기하는 용언과 서로 공유하는 의미 유형과 사건구조로 긴밀히 연결되어 있다고 생각하는 것이다 (박소영, 2001). 앞서 언급한 정도부사 ‘너무, 아주, 매우’를 생각해 보자.

- (1) ㄱ. (날씨가) 너무 좋다.
- ㄴ. (날씨가) 아주 좋다.
- ㄷ. (날씨가) 매우 좋다.

(1 ㄱ, ㄴ, ㄷ)에서 ‘너무, 아주, 매우’는 형용사 ‘좋다’와 공기하여 날씨가 좋은 정도를 표현하고 있는데 이들 의미는 매우 유사하기 때문에 세 정도부사의 의미 차이를 구별해 내는 것은 매우 어려운 일이다.

- (2) ㄱ. (철수는 미션 앞으로) 너무 갔다.
- ㄴ. (철수가 고향에) 아주 갔다.
- ㄷ. \*매우 갔다.

그러나 (2 ㄱ, ㄴ, ㄷ)에서 위의 세 정도부사는 동사 ‘가다’와 공기하는 그 의미가 확연히 구분된다. (2 ㄱ)에서 ‘너무’는 철수가 일정 정도를 넘어 미션 앞으로 치우쳐 있는 정도를 의미하고, (2 ㄴ)에서 ‘아주’는 철수가 고향으로 돌아가서 다시는 돌아오지 않을 것을 의미하고, (2 ㄷ)에서 ‘매우’는 동사 ‘가다’와 의미 자질을 공유하지 못하기 때문에 공기하지 못한다. 위의 예에서 단순하게 세 정도부사의 관계를 살펴보면, 정도부사 ‘너무’는

‘매우’ 보다 ‘아주’ 에 가까운 것으로 생각할 수 있다.

정도부사만으로는 직접적인 관계를 파악할 수 없지만, 정도부사와 공기하는 용언을 통해 2 차적으로 정도부사의 관계를 살펴 볼 수 있다. 이러한 정도부사의 관계적 속성은 공기하는 정도부사와 용언의 2원 행렬 자료를 통해 계산된다(정성훈 2014: 74-82). 2원 행렬 자료를 바탕으로 구축된 네트워크를 2원 네트워크(two mode network)라고 하는데, 정도부사와 용언의 2원 네트워크에서는 두 정도부사가 공기하는 용언이 많으면 많을수록 두 정도부사는 그 의미와 기능이 매우 유사하다고 가정에서 출발하는 것이다.

표 3-1. 닫힌 체계의 정도부사와 용언 2원 행렬 자료

	예쁘다	많다	덥다	가깝다	가다	먹다	끝나다	비슷하다
거의	0	0	0	0	1	1	1	1
아주	1	1	1	1	1	1	1	1
매우	1	1	1	1	0	0	0	1
좀	1	1	1	1	1	1	0	1
너무	1	1	1	1	1	1	0	1

<표 3-1>은 닫힌 체계의 정도부사와 용언을 가정하고 2원 행렬 자료를 만든 것이다. 정도부사 ‘거의, 아주, 매우, 좀, 너무’ 와 용언 ‘예쁘다, 많다, 덥다, 가깝다, 가다, 먹다, 끝나다, 비슷하다’ 가 있다고 하자. 정도부사 ‘아주’ 가 형용사 ‘예쁘다’ 와 공기한다면 2원 행렬 자료( $D$ )에서  $D_{아주, 예쁘다} = 1$ 로 표현하고, 정도부사 ‘거의’ 가 형용사 ‘예쁘다’ 와 공기하지 않는다면  $D_{거의, 예쁘다} = 0$ 으로 표현한다(정성훈 2014: 75).

이를 네트워크 구조로 나타내면 <그림 3-1>과 같다.

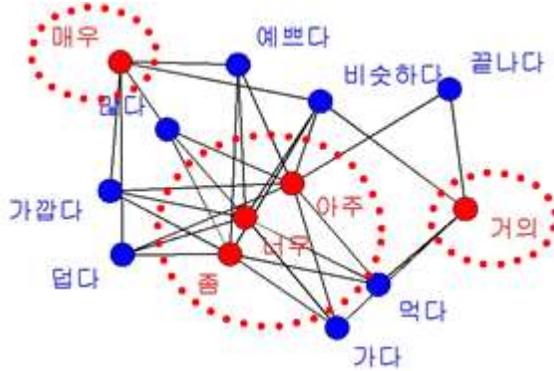


그림 3-1. 정도부사와 용언 2원 네트워크 구조

이와 같은 2원 네트워크 구조는 직관적으로 파악하기 어려운 정도부사의 관계적 속성을 공간적으로 파악할 수 있도록 도와준다. 위의 네트워크 그림에서는 정도부사 ‘아주, 너무, 좀’이 하나의 부류를 구성하고, 정도부사 ‘거의’와 ‘매우’가 각각 한 부류를 이루고 있다. 그러나 위처럼 비교적 간단한 닫힌 체계에서는 정도부사와 용언의 유형이 한정적이기 때문에 시각적으로 쉽게 파악할 수 있으나 정도부사와 용언의 유형이 많아지는 복잡계에서는 2원 네트워크가 매우 복잡한 구조를 띠게 된다. 복잡한 2원 네트워크 구조에서는 정도부사 간 관계적 속성을 시각적으로 한 눈에 파악할 수 없기 때문에 보다 정확하고 계량화된 과정이 필요하게 된다. 이를 위해서 2원 네트워크를 1원 네트워크로 투사(project)한다. 투사 과정은 일반적으로 전치행렬을 이용하는데 이를 통해 2원 행렬 자료를 1원 행렬 자료로 치환할 수 있다.<sup>7</sup> <표 3-2>는 정도부사와 용언의 2원 행렬 자료를 정도부사의 1원 행렬 자료로 치환한 것이다. 치환된 행렬 자료의 값은 두 정도부사가 함께 공기하는 용언의 수를 나타낸다. 즉 1원 행렬의 값이 높으면 두 정도부사는 의미나 기능적으로 매우 가까운 관계로 볼 수 있다. 위의 네트워크 구조에서는 정도부사 ‘아주, 좀, 너무’가 가까운 관계에 있고, 정도부사 ‘거의, 매우’는 상대적으로 먼 관계에 있다.

표 3-2. 정도부사의 1원 행렬 자료

<sup>7</sup> 투사과정에 대한 자세한 설명은 정성훈(2014) 3장 참조.

	거의	아주	매우	좀	너무
거의	-	4	1	3	3
아주		-	5	7	7
매우			-	5	5
좀				-	7
너무					-

2원 행렬 자료의 방법과 마찬가지로, 1원 행렬 자료도 네트워크를 구조화할 수 있는데, 이는 <그림 3-3>과 같다. 2원 네트워크 구조와 비교해 보면 정도부사 ‘매우’는 정도부사 ‘아주, 너무, 좀’과 상대적으로 가까워진 반면, 정도부사 ‘거의’는 떨어져 있는 모습을 보인다.

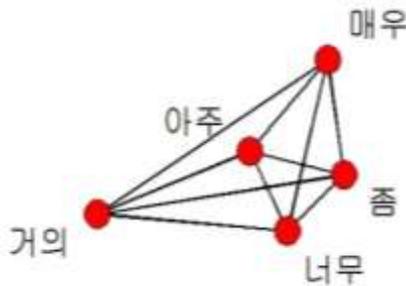


그림 3-3. 정도부사의 1원 네트워크 구조

#### 4. 세종말뭉치 문어자료의 정도부사 네트워크

3장의 정도부사 네트워크가 연구자의 직관으로 구축한 네트워크의 예였다면, 4장에서는 말뭉치를 바탕으로 정보부사 네트워크를 구축하고 분석한다. 연구자의 직관으로는 닫힌 정도부사 체계를 구축할 수밖에 없는데 말뭉치를 이용한다면 보다 열린 정도부사 체계를 구축할 수 있기 때문에 실제 정도부사 구조에 근접한 네트워크를 구축할 수 있다.<sup>8</sup>

말뭉치에 기초한 2원 네트워크에서는 공기하는 정도부사와 용언의 관계에

<sup>8</sup> 네트워크 분석의 장점 중 하나는 계량적 연구 방법임에도 질적 방법과 양적 방법을 동시에 적용할 수 있다는 점이다(정성훈 2014: 73). 닫힌 체계의 네트워크처럼, 연구자의 언어 직관을 바탕으로(질적 방법)으로 언어 직관을 검증할 수 있고, 열린 체계의 네트워크처럼, 말뭉치를 바탕으로(양적 방법)으로 수치를 계량적으로 검증할 수도 있기 때문이다.

가중치(weight)를 부여한다. 예를 들어, 정도부사 ‘매우’와 용언 ‘답다’의 공기 빈도가 정도부사 ‘아주’와 용언 ‘답다’의 공기 빈도보다 높다면 ‘매우’는 ‘아주’보다 용언 ‘답다’와 더 긴밀한 관계에 있다고 보고 가중치를 부여할 수 있다. 이 연구에서는 정도부사와 용언의 공기관계를 추정하는 G-검정(G-test)에서 추출된 G 값을 가중치로 부여한다.<sup>9</sup> 이번 장에서는 정도부사와 용언의 공기관계를 가중치로 하는 정도부사 네트워크를 구축한다. 정도부사 네트워크의 기본 지표를 측정하고, 정도부사 간 관계적 속성을 밝히고 이를 통해 정도부사의 분류를 시도한다.

#### 4.1. 세종말뭉치와 가중치 측정

3장에서 살펴본 정도부사 네트워크는 가중치가 없는 ‘2진 네트워크’ 구조였다. 2진 네트워크 구조란, 노드 간 링크가 0과 1로만 측정되는 네트워크를 말한다. <표 3-1>처럼, 2진 네트워크에서는 정도부사 ‘아주’가 형용사 ‘예쁘다’와 공기하면 ‘아주-예쁘다’의 링크가 1로 표현되고, 정도부사 ‘거의’가 형용사 ‘예쁘다’와 공기하지 않을 때 0으로 표현된다. 그러나 언어 직관을 설명하고자 할 때는 2진 네트워크 구조가 유용하지만 실제 언어자료를 설명하려 한다면 2진 네트워크보다는 가중치 네트워크가 더 유용할 것으로 보인다. 가중치 네트워크는 링크를 1로만 표현하는 것이 아니라 그 정도에 따라 값을 차등적으로 부여한다. 그렇다면 정도부사와 용언의 링크는 어떻게 가중치를 부여할까?

여러 가지 방법이 있겠지만 이 연구에서는 G 값을 가중치로 부여한다. 정도부사와 용언 간에는 의미와 기능에서 매우 밀접한 관계가 있어서 자주 공기하는 관계가 있는 반면, 정도부사와 용언이 우연하게 몇 번 사용된 관계도 있다. 자주 공기하는 단어들은 ‘유의미한 공기관계’를 형성하는데 정도부사와 용언의 빈도를 근거로 각 정도부사와 용언에 대해 G-검정을 수행하고 그 값을 링크의 가중치로 부여한다.

<sup>9</sup> G-검정(G-test)은 우도비(likelihood-ratio) 또는 최대 우도(maximum likelihood)를 이용한 유의성 검정의 하나이다. G-검정(G-test)은 카이제곱 검정의 대안으로 알려져 있다. G-검정은 카이제곱 검정과 계산 방법이 유사하지만, 실험 디자인(experimental designs)도 다르고, 귀무가설(null hypotheses)도 다르기 때문에 서로 다른 통계 검정(statistical test)으로 간주된다.

G-검정은 우도비 검정(likelihood ratio test) 중의 하나로, 카이제곱 검정에 비해 더 정교한 통계적 장치이다. G-검정은 적합도 검정과 독립성 검정으로 나눌 수 있는데, 적합도 검정은 이론적인 기대 빈도와 명목변수의 관찰 빈도를 비교하여 실제 관찰 빈도가 기대 빈도와 유의미하게(significantly) 차이가 나는지를 알아보는 것이고, 독립성 검정은 두 개 이상의 명목변수들의 관찰 빈도를 비교하여 변수들의 관찰 빈도가 유의미하게(significantly) 차이가 나는지를 알아보는 것이다. (1)은 G 값을 계산하는 일반식이다.

(1) G 값의 일반식

$$G = 2 \sum_i [O_i \times \ln(\frac{O_i}{E_i})], \text{ (단, } \ln \text{은 자연로그이다)}$$

$O_i$ 는 각 정도부사와 용언이 공기하는 실제 관찰 빈도이고,  $E_i$ 는 각 정도부사와 용언이 공기할 것으로 계산할 수 있는 기대 빈도이다. G-검정은 일반적으로 두 개의 명목변수들을 비교하기 때문에 빈도 분할표로 자료를 정리한다. 여기서 계산된 G 값이 네트워크 가중치가 된다.<sup>10</sup>

#### 4.2. 문어자료의 정도부사 네트워크

세종말뭉치 문어자료에서는 47 개의 정도부사 중에서 42 개의 정도부사가 용언과 유의미한 공기관계를 이루었다.

(3) 문어자료의 네트워크 정도부사

가장, 거의, 극히, 꽤, 너무, 너무나, 너무너무, 다소, 대단히, 더, 더욱, 덜, 되게, 많이, 매우, 몹시, 무척, 별로, 보다, 비교적, 상당히, 심히, 썩, 아주, 엄청, 워낙, 유난히, 유달리, 은근히, 전혀,

<sup>10</sup> G-검정에서 자유도가 1 일 때, p=0.001 의 G 값은 10.83 인데, 이 연구에서는 G 값이 10.83 이상일 경우에만 정도부사와 용언은 유의미하게 공기하는 것이라고 간주한다. 따라서 G 값이 10.83 이상인 정도부사와 용언의 의미와 기능이 매우 밀접하게 연결되어 있다고 생각하고, 10.83 이하인 정도부사와 용언의 링크 가중치는 0 으로 조정하였다.



네트워크라고 볼 수 있는데, 네트워크가 비교적 서로 긴밀히 연결되어 있는 네트워크로 판단된다. 네트워크의 군집화 계수는 약 0.974 였다. 이 네트워크에서는 정도부사와 이웃 정도부사의 이웃과 서로 연결될 확률이 매우 높다. 즉 정도부사 네트워크는 기본적으로 ‘정도성’이라는 의미를 공유하는 하나의 큰 집단으로 분류될 수 있을 것이다.

한편 문어자료의 정도부사 네트워크의 Opsahl et al.(2010) 중심성은 다음과 같다. 앞서 살펴보았듯이, Opsahl et al.(2010) 중심성은 단순한 빈도나 연결강도의 영향력이 아니라 연결강도의 변이까지 고려한 각 정도부사의 중심성이기 때문에 연결강도의 순위나 빈도순위와 또 다르게 나타난다.

표 4-1. 세종말뭉치 문어자료에서 정도부사 네트워크의 Opsahl et al.(2010) 중심성

	정도부사	링크의 수	Opsahl et al.(2010) 중심성
1	가장	41	132,612.9
2	몹시	39	119,699.9
3	더욱	41	105,077.2
4	다소	40	100,356.2
5	매우	41	85,971.0
6	너무	41	83,429.5
7	참	41	74,575.7
8	무척	41	71,994.6
9	극히	36	70,228.5
10	더	41	68,690.0
...	...	...	...

세종말뭉치 문어자료의 정도부사 네트워크의 근접 중심성은 다음과 같다. 근접 중심성이 높은 정도부사는 네트워크에서 새로운 정보가 들어오거나 변화가 생겼을 때 다른 정도부사들에게 가장 빠르게 정보를 전달하거나 전달 받을 가능성이 높다는 것을 의미한다(정성훈 2014:64-65). 그러나 대체로 정도부사의 근접 중심성은 크게 다르지 않았다.

표 4-2. 세종말뭉치 문어자료에서 정도부사 네트워크의 근접 중심성

	정도부사	근접 중심성	표준 근접 중심성
1	몹시	0.0317	0.0008
2	무척	0.0310	0.0008
3	매우	0.0297	0.0007

4	가장	0.0293	0.0007
5	너무	0.0291	0.0007
6	극히	0.0287	0.0007
7	참	0.0282	0.0007
8	아주	0.0269	0.0007
9	더욱	0.0266	0.0006
10	너무나	0.0265	0.0006
...	...	...	...

문어자료의 정도부사 네트워크의 매개 중심성은 다음과 같다. 네트워크 내에서 매개 중심성이 높은 정도부사는 다른 정보부사 간의 의미 상호작용(interaction)을 촉진시키거나 억제시킬 수 있는 힘도 가진 것으로 평가할 수 있다(Freeman, 1979). 정도부사 네트워크에서 매개 중심성이 높은 정도부사들이 존재한다면 그 정도부사를 근간으로 네트워크가 잘 조직되어 있다는 것을 의미한다.

표 4-3. 세종말뭉치 문어자료에서 정도부사 네트워크의 매개 중심성

	정도부사	매개 중심성
1	가장	433
2	너무	311
3	더	266
4	참	262
5	무척	162
6	더욱	141
7	몹시	112
8	별로	102
9	참	95
10	정말	79
...	...	...

다음은 문어자료의 정도부사 네트워크의 나무구조 그림이다.<sup>11</sup>

<sup>11</sup> 나무구조 그림(dendrogram)은 네트워크 군집 분석을 이용한다. 자세한 내용은 정성훈(2014) 3장 참조.



없는 것으로 판단된다.

이렇듯 네트워크 분석은 언어 자료를 공간적으로 제시하여 쉽게 언어 정보를 파악하게 해 주고, 연구자의 언어 직관과 비교할 수 있는 틀을 제공할 수 있다.

## 5. 결론

우리는 지금까지 세종말뭉치 문어자료에 나타난 정도부사의 사용빈도를 분석하고, 정도부사와 공기하는 용언과의 관계적 속성을 고려한 2원 네트워크 분석을 통해 정도부사들을 분류하고자 하였다. 세종말뭉치 문어자료에서 용언과 공기하는 정도부사는 모두 42 유형이 있었고, 총 84,223 번 출현하였다. ‘더, 가장, 좀, 많이, 너무, ...’ 순으로 빈도가 높았는데, 용언과의 관계적 속성을 알아보기 위해서 정도부사에 대한 2원 네트워크 분석을 수행하였다. 이 연구에서는 정도부사와 용언의 ‘유의미한’ 공기관계에 주목하고, 정도부사의 구조를 부사와 용언의 2원 구조를 통해 살펴보았다. 정도부사는 용언과 2원 네트워크를 형성하였고, 이는 방향성과 가중치가 존재하는 네트워크였다. 우리는 이를 다시 1원 정도부사 네트워크로 투사하여 살펴보았다. 이러한 네트워크 분석은 한국어 정도부사의 의미와 기능의 관계적 속성을 밝히는 새로운 방법론으로, 그 동안 언어 직관에만 의존해 왔던 정도부사 분류에 새로운 관점을 제공할 수 있을 것이다. 또한 네트워크 분석을 위한 사전 작업으로 수행된 정도부사와 용언의 공기관계 기술은 말뭉치에서 실제 언어자료를 관찰하고, 부사와 용언의 공기관계를 기술할 수 있기 때문에 이론적 관점에서 부사와 용언의 문법현상을 설명하는 기초 자료를 제공할 수도 있을 것이다.

## 참고문헌

- 고경태(2009) 「언어 관계를 통한 한국어 성분 부사 선정에 대한 시론」, 『한국어 의미학』 28, 한국어 의미학회, 29-48.
- 권재일(2012) 『한국어 문법론』 대학사.
- 김경훈(1996) 『현대국어 부사어 연구』 서울대학교 박사학위논문.
- 김용학(2007) 『사회 연결망 분석(개정판)』 박영사.
- 남기심 외(2006) 『왜 다시 품사론인가』 커뮤니케이션북스.
- 남기심, 고영근(1993) 『표준국어문법론(개정)』 탑출판사.
- 바라바시, A.(2002) 『링크: 21 세기를 지배하는 네트워크 과학(번역본)』, 강병남·김기훈 역(2002), 동아시아.
- 박병선(2003) 『국어 공기관계의 계량언어학적 연구』, 고려대학교 박사학위논문.
- 박소영(2001) 「정도부사 ‘아주’, ‘거의’, ‘매우’의 의미와 분포」, 『담화와 인지』 8-1, 담화인지언어학회, 85-108.
- 배진영(2012) 「구어와 문어 사용역에 따른 정도부사의 분포와 사용 양상에 대한 연구」, 『국제어문』 54, 국제어문학회, 95-140.
- 서상규(1991) 「정도 부사에 대한 국어학사적 조명과 그 분류에 대해」, 『언어문학』 23, 연세대학교 국어국문학과, 219-48.
- 서정수(1996) 『현대국어문법론』 한양대학교 출판원.
- 서정수(2005) 『한국어의 부사』 서울대학교 출판부.
- 손남익(1995) 『국어 부사 연구』 박이정.
- 신지연(2002) 「정도부사의 범주화 기준에 대하여」, 『어문학』 78, 한국어 문학회, 71-87.
- 신효필(2009) 『언어학과 통계모델』 서울대학교 출판부
- 임규홍(2003) 「국어 정도 부사의 화용화」 『언어과학연구』 24, 언어과학회, 283-302.
- 임유종(1999) 『한국어부사연구』 한국문화사.
- 임유종(2010) 「정도부사의 결합 관계와 한국어 교육」 『한국언어문화』 42, 한국언어문화학회, 371-393.
- 정성훈(2014) 『현대 한국어 부사에 대한 계량언어학적 연구-확률 통계 모형과 네트워크를 이용한 분석』 서울대학교 박사학위논문.
- 최현배(1971, 1985) 『우리말본』 정음문화사.
- 한영균, 고은아(2011) 「유의적 정도부사의 빈도, 분포, 결합관계의 분석과 그 활용 - 학습자 사전의 용법 기술의 관점에서-」 『한국어 의미학』 35, 한국어 의미학회, 335-394.
- 허웅(1995) 『20세기 우리말의 형태론』 샘문화사.
- 허웅(2000) 『20세기 우리말의 형태론(개정판)』 샘문화사.
- 홍사만(2002) 「국어 정도 부사의 하위 분류」 『어문논총』 36, 경북어문학회, 31-74.

- B. Mandelbrot(1953) 「An informational theory of the statistical structure of languages」 In Jackson, W. (ed.) 『*Communication theory*』 Butterworth, London, 486-502.
- D. Watts(1999) 『*Small worlds: the dynamics of networks between order and randomness*』 Princeton, N.J.: Princeton University Press.
- D. Watts & S. Strogatz(1998) 「Collective dynamics of “small-world” networks」 『*Nature*』 393, 440-442.
- H. Baayen(2001) 『*Word frequency distributions*』 Kluwer Academic Publishers, Dordrecht.
- H. Baayen(2008) 『*Analyzing Linguistic Data: A Practical Introduction to Statistics using R*』 Cambridge.
- L. Freeman(1978) 「Centrality in social networks: Conceptual clarification」 『*Social Networks*』 1, 215-239.
- L. Kaufman & J. Rousseeuw(2009) 『*Finding groups in data: an introduction to cluster analysis*』 Vol. 344, John Wiley & Sons.
- R. Albert, Jeong, H., & A. Barabasi(1999) 「Diameter of the World Wide Web」, 『*Nature*』 401, 130-131.
- R. Burt(1978) 「A structural theory of interlocking corporate directorates」, 『*Social Networks*』 1-4, 415-435.
- R. Burt(1992) 『*Structural holes : the social structure of competition*』 Cambridge, Mass.: Harvard University Press.
- S. Borgatti & M. Everett(1997) 「Network analysis of 2-mode data」, 『*Social networks*』 19, 243-269.
- S. Wasserman & K. Faust(1994) 『*Social Network Analysis: Methods and Applications*』 New York, NY: Cambridge University Press.
- T. Opsahl(2009) 『*Structure and Evolution of Weighted Networks*』 University of London, London, UK
- T. Opsahl & P. Panzarasa(2009) 「Clustering in weighted networks」. 『*Social Networks*』 31-2, 155-163.
- T. Opsahl, F. Agneessens & J. Skvoretz(2010) 「Node centrality in weighted networks: Generalizing degree and shortest paths」, 『*Social Networks*』 32, 245-251.