creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

**저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.**

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

문학석사학위논문

# 한국어의 저지현상에 대한 계량언어학적 연구

2021년 8월

서울대학교 대학원

국어국문학과 국어학전공

조 성 원

# 한국어의 저지현상에 대한 계량언어학적 연구

지도교수 이 진 호

이 논문을 문학석사 학위논문으로 제출함.

2021년 4월

서울대학교 대학원
국어국문학과 국어학 전공
조 성 원

조성원의 문학석사 학위논문을 인준함.

2021년 7월

위 원 장 _____

부위원장 _____

위 원 _____

# 국문초록

이 논문에서는 한국어의 접사에 의한 파생에서의 저지(blocking)가 일어나는 양상을 계량언어학적 방법론을 도입하여 포착하고 그러한 모습으로 나타난 이유에 대해 탐구하고자 했다. 저지란 어휘부에 존재하는 어떤 형태가 다른 동의적인 경쟁 형태의 출현을 방해하는 현상이다. 선행 연구에서는 서구의 연구자와 국내의 연구자를 가리지 않고 저지를 형태론적 현상의 변방에 두어 홀대한 경향이 있다. 고전적인 연구에서 저지현상은 언어적 제약으로 설명할 수 없는 단어의 출현 제약을 이름 짓는 역할에 머물렀고, 코퍼스와 빈도를 통한 최근의 연구에서도 기술이 비교적 간단한 생산성에 관한 논의가 대부분이었다. 특히 계량언어학적 측면에서 저지를 독립적으로 다룬 국내의 연구는 전무하다고 보아도 무방하다.

그러나 최근 품질이 높은 한국어 말뭉치가 수차례 구축되고 자료를 처리할 통계적인 방법론과 소프트웨어도 다양하게 개발되면서 앞선 연구에서 타협한 문제에도 도전할 수 있는 여건이 마련되었다. 파생과 같은 형태론적 현상은 빈도와 깊은 관련을 맺고 일어나기 때문에 이는 그동안 소홀했던 주제인 저지현상을 보다 나은 계량언어학적 방법으로 연구하기에 절호의 기회라고 여겨졌다. 이와 같은 인식 아래 본 논문에서는 크게 세 가지 목적을 정하여 논의를 전개하였다. 첫째, 직관적인 분석에 그쳤거나 기술적인 한계가 있어 부족한 점이 있었던 기존의 계량적 연구 방법을 종합하고 수정한다. 둘째, 접사별 파생어들의 최신화된 빈도 분포 자료를 수집하고 분석하여 필요한 지표를 구하고 언어학적 의미를 탐구한다. 셋째, 정리한 자료를 바탕으로 한국어 저지현상의 공시적인 모습과 기존에 알려지지 않은 발생 양상을 파악한다.

2장에서는 코퍼스에 나타난 언어의 상태를 논의의 목적에 알맞게 가공할 계량적 방법을 고안하였다. 선행 연구에 따르면 어떤 파생어가 어휘부에서 인출될 때는 단어 전체가 통째로 저장·인출되거나 접사+어근으로 분석(parsing)되어 새로 조합하는 두 가지의 경로 중 하나를 통한다. 이 현상은 단어의 빈도와 밀접하게 관계되어 있다. 빈도가 높은 파생어는 대체로 어휘부에 그대로 저장되는 경향이 강하지만 빈도가 낮은 파생어는 필요에 따라 조합해서 쓰는

경향이 강한 것이다. 생산성에 관한 계량적 연구에서는 이들 두 부류의 파생어 집합이 만나는 지점인 분석 임계점, 혹은 저장 임계점을 찾을 방법을 찾기 위해 부단히 노력해 왔다. 접사의 생산성을 계산하기 위해서는 분석 경로를 따르지 않은 파생어를 분석 대상에서 제외할 수 있어야 하기 때문이다. 이를 위해서는 낮은 빈도에서는 주로 분석 경로를 거치다가 빈도가 높아짐에 따라 저장되는 경향이 더 강해지기 시작하는 임계점을 특정하고 그 빈도를 구할 수 있어야 한다. 2장에서 기존 연구를 종합하여 고안한 빈도 분석 방법은 바로 이 지점을 찾기 위한 것이다.

3장에서는 경쟁 관계에 있다고 생각되는 접사들과 그렇지 않은 접사들을 묶어 파생어의 빈도 분포를 분석하고 각 접사의 생산성과 저장 임계 빈도를 추정하였다. 파생어의 빈도 정보는 4장에서 저지의 실현 양상을 확인하는 데에 필수적인 기초 자료가 된다. 기존의 연구 결과에서 주장한 바와 같이 저지는 어휘부에 미리 저장된 파생어가 다른 파생어의 출현을 막는 것이므로 저장 강도를 간접적으로 나타내는 지표인 빈도를 통해 분석할 수 있기 때문이다.

그런데 4장에서 무엇보다 새롭게 주장하고자 한 점은 신어 파생에서도 경쟁하는 접사들 사이에 저지가 일어난다는 것이다. 여러 접사의 생산성과 임계 빈도를 구해 보면 경쟁 접사의 그룹 내에서 생산성의 전반적인 격차가 규칙적으로 나타난다는 사실도 확인할 수 있다. 빈도와 접사의 의미에 따라 파생어들을 분류해보면, 이는 경쟁하는 접사들 사이의 의미별 분업에 의한 것으로 해석할 수 있다. 이러한 신어 파생에서의 분업은 접사의 전형적인 용법과는 관계없이 일어나는 것으로 나타났다. 그렇다면 의미적으로는 파생할 수 있으며 실제로 종종 파생이 일어나는 영역에서도 단지 경쟁하는 접사가 담당했다는 이유만으로 파생이 억제되고, 결과적으로 생산성의 격차가 발생한 것이므로 이를 신어 파생에서의 접사 간 저지로 볼 수 있다.

저지현상은 어휘부의 저장과 관련된 현상이므로 빈도를 통한 계량언어학적 연구에 적합하지만, 생산성과 같이 접사 내적인 빈도 구조를 살피는 것으로 충분한 개념들과 달리 여러 접사를 함께 비교해야 하므로 더 고차원적인 분석 방법이 필요했다. 즉, 파생어들을 빈도와 함께 단순히 나열하고 그 목록을 직접 해석하는 기존의 방식으로는 저지를 충분히 이해할 수 없다. 언어 자료가 갖는 통계적 특성을 적극적으로 활용함과 동시에 관련된 언어학적 개념을 도

입했을 때에야 비로소 인지할 수 있는 거시적인 구조가 언어 내에 분명 존재하는 것이다. 이런 측면에서 볼 때, 형태론뿐 아니라 언어학의 다양한 분야에서 계량적 방법론이 새로운 시각을 제시할 수 있는 여지는 무궁무진하다고 할 수 있다.

주요어 : 저지, 파생, 경쟁, 생산성, 분석(parsing), 빈도
학 번 : 2019-26018

# 차 례


1. 서론 ····· 1
1.1. 연구 목적 ····· 1
1.2. 연구사 검토 ····· 3
1.3. 용어 정리 ····· 6
1.4. 논의 순서 ····· 8

2. 연구 방법론 ····· 10
2.1. 연구 대상과 분석 도구 ····· 10
2.2. 핵심 지표의 도출 ····· 12
2.2.1. 생산성 ····· 13
2.2.2. 단어의 저장과 빈도 ····· 18
2.2.2.1. 기반 이론 ····· 19
2.2.2.2. 빈도 분석의 실제 ····· 26

3. 파생 접사의 빈도 특성 ····· 33
3.1. 도입 ····· 33
3.2. 생산적 접사 ····· 34
3.2.1. ‘-가(家)’, ‘-자(者)’, ‘-인(人)’ ····· 34
3.2.2. ‘-거리-’, ‘-대-’ ····· 40
3.2.3. ‘-이’, ‘-음’, ‘-기’ ····· 45
3.2.4. ‘-스럽-’, ‘-하-’, ‘-롭-’, ‘-지-’, ‘-적(的)’ ····· 51
3.2.5. ‘-꾼’, ‘-쟁이’ ····· 63
3.2.6. ‘비(非)-’ ····· 67
3.2.7. 경쟁하지 않는 접사 ····· 69
3.3. 비생산적 접사와 신생 접사 ····· 83
3.3.1. ‘-이-’ ····· 83

3.3.2. ‘불(不)-’ ····· 85
3.3.3. 기타 접사들의 파생어와 빈도 ····· 86

4. 한국어의 저지현상 ····· 88
4.1. 도입 ····· 88
4.2. 유형 저지 ····· 88
4.2.1. 저장 임계 빈도와 생산성의 상관관계 검증 ····· 89
4.2.2. 접사 간 저지 양상 ····· 94
4.3. 항목 저지 ····· 106
4.3.1. 일반적 항목 저지 ····· 108
4.3.2. 예외적 항목 저지 ····· 112
4.3.2.1. ‘-거리-’, ‘-대-’ 파생어 간 저지 ····· 112
4.3.2.2. ‘-스럽-’, ‘-적(的)’, ‘-하-’ 파생어 간 저지 ····· 115

5. 결론 ····· 117

참고문헌 ····· 119
Abstract ····· 125

# 1. 서론

## 1.1. 연구 목적

언어는 자연의 다른 모든 부분과 마찬가지로 끊임없이 변화하는 비정형적 현상이다. 그러나 변화의 방향과 그 안에서 변화를 겪는 요소들에는 정형화된 규칙이 있다. 우리는 규칙 덕분에 말을 즉석에서 만들어낼 수 있고, 처음 접한 문장이나 단어의 뜻을 이해할 수 있다. 따라서 규칙은 언어의 핵심이다. 규칙이야말로 무의미한 소리나 단어의 나열로부터 언어를 구분하는 본질적인 요소이다. 따라서 대개의 언어학 연구는 기본적으로 규칙이 언어에 내재한다는 점을 전제로 하여 언어의 여러 분야를 탐구해 왔다.

그런데 계량언어학적 연구는 언어 현상에 이들과 반대 방향으로 접근한다. 언어에는 규칙이 있지만, 동시에 끊임없이 변화하며 구성 요소마다의 모습이 제각각인 현상이기도 한 것이다. 계량언어학적 연구에서는 변화하는 언어의 한 시점의 모습을 사진 찍듯이 포착해서 그것을 분석하며 거꾸로 규칙을 찾아나간다. 그리고 이 과정에서 대규모 코퍼스를 필수적으로 사용하게 된다. 이와 같은 이유로 계량언어학적 연구는 데이터 처리 및 관련 통계 기술이 충분히 발달한 최근에 들어서야 본격적으로 가능하게 되었다.

본 연구는 계량언어학적 방법을 사용하여 한국어의 접사 파생 과정의 특징을 밝혀보려는 한 시도이다. 접사에 의한 파생어 형성은 계량언어학적 방법론을 통해 가장 활발히 연구되어 온 주제이다. 통사적 구성은 어떤 형태에 뒤따를 수 있는 형태가 거의 무한하지만 접사에 의한 파생은 한정된 수의 접사에 의해 이루어지므로 분석의 기준을 정하기가 훨씬 용이하다. 또, 생산성이나 저지 현상과 같이 전통적인 방법으로는 설명이 쉽지 않았던 불규칙적 현상들이 관련되어 있어 계량적 방법을 통해 연구하는 것이 적절하다.

파생 형태론 연구에 계량적인 방법을 본격적으로 처음 도입한 연구는 Baayen(1989)이다.[1] 그는 코퍼스를 활용한 빈도 분석을 통해 생산성을 수치

1) Baayen(1989)는 형태론 연구에 단순 통계를 넘어선 계량언어학적 방법론을 처음으로 도입하였으므로 매우 중요한 의미를 갖는다.

화하였다. 이 연구가 나온 이후 통계학, 심리언어학적 측면에서 형태론적인 분석을 도울 수 있는 다양한 방법과 도구들이 계속 고안되었으며 현재까지도 관련된 연구가 활발하게 이루어지고 있다. 국내에서도 이와 같은 방법론이 소개되면서 파생어의 생산성과 관련된 여러 연구들이 이러한 연구 방법을 도입하였다.

생산성이 계량언어학적으로 대대적인 주목을 받은 데 비해 접사 간의 경쟁이나 저지현상에 대해서는 많은 연구가 이루어지지 않았다. 머릿속 어휘부에서 어떤 의미가 인출될 때 의미가 비슷한 접사 사이에는 선택받기 위한 경주가 일어나며 어떤 접사가 다른 접사의 파생을 막기도 하는데 이를 경쟁(competition)과 저지현상(blocking)이라고 한다.[2] 이런 현상은 한 접사에만 집중할 수 있는 파생과 생산성에 비해 양상이 복잡하고 비정형적이어서 규칙을 찾기가 훨씬 어렵다. 계량적인 언어 연구들이 경쟁과 저지보다는 생산성을 주로 다룬 것은 아마도 이런 이유에서였을 것이다.

특히 국내에서는 이광호(2009)가 저지를 다룬 계량적 연구로는 유일하다. 이처럼 선진적인 연구 결과임에도 불구하고, 이 연구 역시 생산성에 대한 탐구가 주된 목적이었기 때문에 저지현상에 대해서는 다양한 각도에서 충분히 심도 있게 다루지는 못한 것으로 보인다. 즉, 저지를 설명하기 위해 직관에 의존한 기존의 분석 방식에서 벗어나 코퍼스를 활용했지만, 여전히 코퍼스를 분석하는 데 유용한 통계적 도구를 충분히 활용하지 못하고, 파생어를 그 빈도와 함께 나열한 목록만을 참고한 단순한 분석에 머물렀다. 따라서 그에 따라 도출된 결과도 부정확한 면이 있다.[3]

본 연구의 목적은 계량언어학적 형태론 연구가 가진 위와 같은 한계를 보완하고, 한국어의 저지현상과 접사 간 경쟁의 본질을 밝히는 데 있다. 이를 위해 다음과 같은 구체적인 목표를 설정하였다. 첫째, 논의 전반에 걸쳐, 생산성에 편향된 기존의 국내외의 계량언어학적 형태론 연구 기조에서 벗어나 저지현상의 탐구에 집중한다. 둘째, 산발적으로 다양한 영역에서 개발된 연구 방법을 종합하여 파생어 전체가 어휘부에 저장되는 지점을 도출할 방법을 고안한다.

---

2) 이렇게 보면 저지현상은 경쟁으로 일어날 수 있는 결과 중 한 가지 유형에 해당하는 것이다.

3) 이는 저지현상만이 아니라 생산성 분석에서도 전반적으로 발견된다.

셋째, 도출된 빈도와 생산성이 동의적 접사 간의 경쟁 속에서 어떻게 나타나는지를 밝힌다.

본 연구에서는 생산성을 계산하되 선행 연구와는 달리 저지현상을 탐구하기 위한 지표로만 사용한다. 저지에 관한 논의에서 생산성을 함께 다룰 수밖에 없는 이유는 두 개념이 직접적으로 연관되어 있기 때문이다. 저지현상은 어휘부 저장과 긴밀히 연관된 현상이기 때문에 빈도를 통해 분석하는데, 생산성도 어휘부 저장과 깊은 관련을 맺고 발생하며 빈도를 통해 산출된다. 또 접사 간의 경쟁은 개별 파생어의 출현을 저지하는 데 그치는 것이 아니라 접사의 생산성에도 체계적으로 영향을 미친다. 이러한 생산에서의 경쟁 양상을 포착하기 위해서는 각 접사의 생산성 측정이 먼저 이루어져야 하는 것이다.

## 1.2. 연구사 검토

오늘날 통용되는 접사 파생에서의 저지현상의 정의는 '동의 관계에 있는 한 접사의 파생어가 높은 빈도에서 발생하는 저지력에 의해 다른 접사의 파생어 출현을 방해하는 현상'이다. 이 정의의 각 부분, 즉, 저지가 상대방의 출현을 완전히 막는 현상이 아니라는 점, 저지의 주체가 접사가 아닌 각 파생어라는 점, 저지를 유발하는 요인이 빈도에 있다는 점은 모두 일련의 연구들 속에서 수정을 거쳐 완성된 것이다.

처음으로 저지(blocking)[4]이라는 용어를 사용하고 그 개념을 널리 알린 것은 Aronoff(1976)이다. 후에 Kiparsky(1982:16)가 단어 형성 규칙에 적용되는 원칙의 하나로 제시한 동의어 회피 원칙이 추가되면서 저지의 생성문법적 정의가 완성되었다. 동의어 회피 원칙이란 어휘 규칙의 출력형은 기존의 어휘 항목과 동의 관계에 있을 수 없다는 원칙이다. 그는 여기서 직접적으로 저지를 거론하지는 않았지만 간접적으로 피저지어와 저지어의 의미가 동일해야 한다는 저지현상에서의 의미 원칙을 추가한 셈이었다. 생성문법적 이론에 따르면 저지현상이란 '파생에 아무런 제약이 없더라도, 한 파생 접사가 파생한 같은 의미의 파생어가 이미 존재한다면 다른 접사의 파생이 저지되는 현상'이다.

4) 논의에 따라 morphological/statistical preemption과 같은 용어를 사용하기도 한다.

이와 같은 의미에서의 저지현상은 국내의 연구에서도 일찍이 다룬 바 있다. 송철의(1992)는 국어의 명사 파생 접미사 '-이, -음, -기'가 다음과 같이 배타적으로 등장하므로 저지현상의 예가 될 수 있다고 하였다.

| 어근 | -이 | -음 | -기 |
|---|---|---|---|
| 걷-(步) | * | 걸음 | * |
| 묻-(問) | * | 물음 | * |
| 울- | * | 울음 | * |
| 죽- | * | 죽음 | * |
| 풀- | 풀이 | * | * |
| 몰- | 몰이 | * | * |
| 벌- | 벌이 | * | * |
| 달리- | * | * | 달리기 |
| 던지- | * | * | 던지기 |
| 쓰-(書) | * | * | 쓰기 |

표 1.1. 송철의(1992)에서 예시한 '-이, -음, -기' 간의 저지현상

그러나 저지현상을 피저지어의 완전한 출현 배제로 정의하면 그 예외가 너무 많아지게 된다. 의미가 완전히 같지 않은데도 저지현상이 일어나거나, 의미가 같은데도 저지가 일어나지 않고 동일한 어근을 공유하며 공존하는 파생어는 어렵지 않게 찾을 수 있다. '-이, -음, -기'의 경우를 살펴보자. 동사 '널뛰-'는 일반적으로 '-기'와 결합한 '널뛰기'로만 파생되며 또 그럴 것으로 예상되지만 어떤 수치가 심하게 오르내리는 현상을 일컬을 때에는 충분히 '-음'과 결합하여 '널뜀'으로도 나타날 수 있다. 이는 두 단어의 의미가 완전히 같아서 항상 바꿔 쓸 수 있다는 뜻이 아니다. 의미영역은 다르지만 특정한 맥락에서는 의미의 차이가 거의 느껴지지 않고, 전형적으로 어울리던 접사도 저지력을 일정 수준 잃게 된다는 의미이다. Plag(2003:79)도 'liver[간]'가 동작주 명사 'live-er['사는 자(者)', live-er]'를 저지한다는 기존의 주장해 대해, 어떤 파생어의 사용에 있어서는 어근과 접사의 전형적인 결합 가능성보다는 맥락이 더 결정적이라는 것을 지적하였다.[5)]

저지현상은 기존에 주장된 바와 같은 철저한 저지가 아니라 빈도에 기반한

점진적(gradual)인 출현 방해이다. 저지의 이러한 특성은 파생어 형성 연구에 코퍼스나 대규모 사전을 통한 계량적 방법이 도입되면서 밝혀졌다. Rainer(1988)에 따르면 저지어가 피저지어의 출현을 막는 힘인 저지력은 빈도에 비례한다.[6] 따라서 저지어의 빈도가 충분히 높을 경우에만 완전한 저지가 일어나고, 빈도가 낮아지면 경쟁 항목들이 쉽게 공존한다.[7] 즉, 저지는 항상 저지어가 피저지어의 출현을 완전히 막는 식으로 일어나지 않는다. 이와 같은 발견을 바탕으로 그는 저지를 '복합어의 형성이 어휘부에 이미 존재하는 동의어에 의해 방해받는(frustrated) 현상'으로 정의했다(Rainer1988:161).[8] Plag(2003)은 이상의 논의를 종합하여 저지현상이 일어날 수 있는 조건을 다음의 세 가지로 요약하였다.

① 동의성 조건
② (피저지어의) 생산성 조건
③ 빈도 조건

조건 ①, ③은 앞서 설명한 내용이다. 즉, 저지어는 피저지어와 의미가 같고 빈도가 충분히 높아야만 경쟁어의 출현을 완전히 막을 수 있다는 조건이다. 조건 ②는 피저지어가 생산적인 접사에 의해 파생된 단어여야 한다는 조건이다. 이 조건은 반대로 어떤 단어쌍이 저지현상의 예외인지를 판별할 때에도 적용할 수 있다. 예를 들어, 국어의 동사 '가물-'은 '가물'이라는 명사 파생형

5) "…provided the appropriate context, the putative oddness of the agent noun *liver* disappears."
6) 정확히 표현하자면 빈도에 비례하는 어휘부에서의 저장 강도에 비례하는 것이다. 그러나 머릿속의 저장 강도는 관찰 불가능하므로, 실증적 연구에서는 저지력이 빈도의 함수라고 표현하더라도 큰 문제가 없다.
7) 두 단어가 공존하는 경우 저지어와 피저지어는 대체로 한 쪽이 압도적인 빈도로 사용되기 때문에 직관적으로 쉽게 구분이 가능하다. 간혹 경쟁하는 형태의 빈도가 비슷한 경우에는 두 파생어에 결합한 접사들의 전체적인 분포를 참고하거나 과거의 자료를 참고하는 것이 일반적이다. 구체적인 내용은 3장에서 살펴보고 4장에서 정리할 것인바, 파생어 계열체 내에서의 완전한 저지는 대부분의 접사에서 잘 유지되며, 그것이 심각하게 침해되는 것은 대체로 접사의 세력이 매우 약해지고 있기 때문인 것으로 보인다.
8) Plank(1981)에서 이미 '화자가 기존의 대체 형태[저지어가 될 수 있는 형태]를 활성화시키는 데 실패하면 저지가 일어나지 않는다'라고 지적한 바 있다. 다만 그는 빈도를 직접적으로 언급하거나 빈도와 저지를 연결짓지는 못했다.

이 이미 존재하고 있으면서도 '가뭄'의 파생을 허용하지만, '가물'이 생산성 조건을 위배하므로 저지의 예외로 볼 수 없다.[9] 영 파생은 공시적으로 새로운 단어를 만든다고 보기 어렵지만 '-음' 명사 파생은 생산적이기 때문이다.

저지의 생성문법적인 정의와 후대의 정의 사이에는 앞서 설명한 것 말고도 저지의 주체에 관한 차이가 더 있다. 생성문법적 이론에서는 접사가 접사를 저지한다고 정의하였으나 Rainer(1988)에서는 각 파생어가 경쟁하는 파생어를 저지한다고 정의하게 된 것이다. 이러한 차이는 저지가 유형 저지와 항목 저지라는 서로 다른 두 종류의 저지로 구분된다는 발견에서 비롯되었다.[10] 현대에 널리 받아들여지는 저지의 개념과 성격은 이러한 논의를 거쳐 1980년대 말경에 확립되고 그대로 사용되어 온 것이다.

## 1.3. 용어 정리

본격적인 논의를 시작하기 전에 정리해야 할 용어가 몇 가지 있다. 저지를 다루는 연구와 생산성을 다루는 계량적 파생형태론 연구 사이에서는 통용되므로 본 연구에서도 그대로 사용하였으나 그 밖의 영역에서는 의미가 다르거나 생소한 용어들이 그것이다. 생소한 용어는 그 정의와 본 논의에서의 중요성을 짚고, 일반적인 연구와 용법이 다른 용어는 그 개념을 명확히 할 필요가 있다. 본 연구에서 유형 저지와 항목 저지는 전자에 해당하고, 생산성은 후자에 해당한다. 먼저 유형 저지와 항목 저지에 대해 살펴보고 본 연구에서 사용할 생산성의 정의를 소개하기로 한다.

전술하였듯 Rainer(1988)는 저지의 정의를 더 정교하게 다듬는 과정에서 유형 저지(type blocking)와 항목 저지(token blocking)를 구분하게 되었다. 두 가지의 저지 중 실존하는 현상은 항목 저지뿐이라는 점에서 당시에는 이 구분이 매우 중요하고 필수적이었으나 같은 이유에서 곧 그 구분이 무의미하게 되었다.

파생어 형성에서의 유형 저지는 '생산적인 한 접사의 파생이 다른 생산적인 접사에 의해 저지되는 현상'으로 정의할 수 있다.[11] Aronoff(1976:53)에서

9) 이는 송철의(1992)에서 저지현상의 예외로 소개한 바 있다.
10) 유형 저지와 항목 저지의 정의는 1.3에 정리하였다.

예로 든 '−ness'와 그 경쟁 접사 '−ity' 사이의 저지가 대표적인 유형 저지에 해당한다. 그는 명사형 X가 존재하는 Xous에 대해서는 두 접사가 다음과 같이 배타적으로 결합한다고 주장하였다.

| Xous | -ity | -ness |
|---|---|---|
| glorious | *gloriousity | gloriousness |
| furious | *furiousity | furiousness |
| gracious | *graciosity | graciousness |
| spacious | *spaciousity | spaciousness |
| fallacious | *fallacity | fallaciousness |
| acrimonious | *acrimoniosity | acrimoniousness |
| laborious | *laboriosity | laboriousness |
| bilious | *biliosity | biliousness |

표 1.2. Aronoff(1976)가 예시한 유형 저지

그러나 이러한 주장은 얼마 지나지 않아 반박되었다. 두 접사의 의미가 일관되게 다른 데다, 실제 언어 사용에서 'destructiveness - destructivity', 'ovalness - ovality'와 같이 두 접사의 파생형이 동시에 나타나는 반례를 쉽게 찾을 수 있었기 때문이다. 이후로도 Aronoff(1976)에서 주장한 것과 같은 유형 저지 관계는 전혀 확인된 바 없다. Rainer(1988)은 바로 이런 점을 지적하며 저지는 항목 저지로만 정의해야 한다고 주장한 것이다.

항목 저지는 이미 존재하는 단어가 다른 단어의 출현을 저지하는 것이다. 항목 저지의 예로는 'thief−*stealer', 'went−*goed' 등이 있다, 접사 파생에 있어 항목 저지가 유형 저지와 다른 점은 어떤 접사 파생어 전체가 다른 접사 파생어 전체를 일관되게 저지하지 않고, 일부 파생어에 대해서만 저지가 이루어진다는 데 있다. 앞서 소개한 '−이, −음, −기' 역시 항목 저지 관계에 있

11) 유형(type, pattern)은 원래 파생 접사뿐만이 아니라 굴절, 합성 등의 다양한 어휘항목을 생성하는 과정을 아우르는 용어다. 그러나 접사에 의한 파생을 다루는 연구에서는 편의상 접사를 유형과 같은 의미로 사용하는 경우가 많다. 가령 '유형 저지'를 접사 간의 저지라는 뜻으로 사용하는 식이다. 이러한 표현은 특히 저지나 경쟁과 관련된 연구에서 자주 사용하였다. 본고에서는 전통을 중시하여 유형 저지라는 용어는 그대로 사용하되, '유형'의 개념이 필요할 때에는 혼동을 방지하기 위해 '접사'를 사용하였다.

다. 이는 한 접사가 다른 접사의 파생을 전부 막는 것이 아니라 어근에 따라 차단 여부가 다르게 나타난다는 점을 통해 알 수 있다. 예를 들어, '-이, -음, -기' 사이의 저지에서 '-이' 파생어 전체가 '-음, -기' 파생어 전체를 저지하지 못하므로 '벌음'이나 '벌기'를 저지하는 것은 접사 '-이'가 아니라 '벌이'라는 항목이다.

그런데 유형 저지는 선행 연구에서 주장한 것과는 다른 형태로 실제로 나타나며 이를 관찰할 방법도 있다. 이 유형 저지는 동의적인 관계에 있는 경쟁 접사 중 무표적인 접사가 나머지 접사의 신어 생산을 방해하는 형태로 나타난다. 이것은 기존에 논의된 바 없는 형태의 저지이다. 일반적으로는 쓸모가 없어진 유형 저지라는 용어와 개념이 본 연구에서 중요한 이유가 여기에 있다.

다음은 생산성이라는 용어를 살펴보겠다. 널리 사용되는 정의에 따르면 생산성이란 특정 시점에 어떤 접사가 파생어를 만들어내는 능력, 또는 결합하는 어근의 다양성이다. 그러나 계량적 연구에서는 새로운 단어를 파생하는 능력으로 생산성을 정의하고 코퍼스에서 해당 접사의 신어를 마주칠 확률로 그 크기를 추정한다. 본 연구는 후자의 정의를 따를 것이나, 기존의 생산성 개념이 필요할 경우 '결합력'이라는 용어를 사용하고자 한다. '유형 빈도(type frequency)' 역시 계량적 형태론 연구에서 비슷한 의미로 사용하지만, 이는 어떤 접사에 대해 주어진 자료 내에서 관측된 결합력을 의미하므로 별개의 개념이다.[12)]

## 1.4. 논의 순서

논의는 다음과 같은 순서로 구성하였다. 저지현상의 기본이 되는 개념을 설명하고 이러한 개념이 정립된 이론적 배경을 소개하는 1장 서론에 이어, 2장에서는 저지에 대해 분석하기 위해 사용할 방법론을 소개한다. 본 논의에서 다룰 유형 저지를 포착하기 위해서는 언어학적인 배경 이론뿐 아니라 통계적으로 시각화하는 기술이 특별히 필요하므로 이를 집중적으로 설명하고자 한다. 3장에서는 2장에서 소개한 분석 도구를 바탕으로 각 접사의 빈도 구조를

12) 유형 빈도란 주어진 코퍼스에서 어떤 접사가 파생한 단어의 가짓수를 의미한다. 항목 빈도(token frequency)는 접사의 각 파생어가 주어진 코퍼스에서 출현한 횟수를 의미한다.

살펴본다. 4장에서는 여기서 수집 및 정리한 자료를 바탕으로 한국어의 저지 현상을 탐구할 것이다.

# 2. 연구 방법론

## 2.1. 연구 대상과 참고 자료

본 논의에서 연구 대상은 파생 접사에 의한 파생과 관련된 저지현상에 한정하기로 한다. 저지가 파생어 사이에만 일어나는 현상은 아니지만, 영 파생이나 활용[13], 합성과 같은 과정은 파생 접사에 의한 파생처럼 계량하기가 쉽지 않기 때문이다.

조사 대상 접사를 선정하기 위해 우선 표준국어대사전에 수록된 접사 목록 533개를 후보로 정하였다. 그리고 동의·유의 관계에 있어 경쟁 관계에 있는 접사들을 위주로 분석할 접사를 선정하였다.

이광호(2009)의 조사에서 이미 생산성이 전혀 없는 것으로 나타난 접사는 조사하지 않았다. 사동, 피동 접미사나 대부분의 접두사가 여기에 해당된다. 이광호(2009)는 접사의 분석 자체가 목적이었으므로 이런 접사들을 다루는 것이 유의미했지만, 본 논의에서는 저장과 생산성 지표를 모두 계산할 수 있는 접사가 필요하다. 따라서 비생산적 접사는 아예 제외하거나 간단히 현대 코퍼스에 나타난 파생어 목록 및 빈도 정보만 제시하였다. 생산성과 상관없이 최대한 조사 대상에 포함하고자 했던 것은 경쟁 관계에 있는 접사들이다. '-거리-', '-대-'와 경쟁하는 '-이-'가 여기에 해당된다. '뒤, 드-, 들-, 휘-'처럼 조금만 의미적 유사성이 의심되어도 일단 조사하였다. 이와 별도로 신생 접사도 조사하였다. '갓-', '꿀-', '-짱' 등이 여기에 해당된다. 신생 접사는 사전에 등재되어 있지 않으므로 코퍼스에서 파생어 출현이 확인된 것들 중 일부를 목록에 포함시켰다.

다음은 위와 같은 선정 과정을 거쳐 본 연구에서 최종적으로 분석 대상에 포

---

13) 한국어의 활용형 간의 저지에 대해서는 연구된 바가 없다. 하지만 서구권에서 굴절형(inflectional form) 사이의 저지현상은 1970년대 이후로 줄곧 활발히 연구되었으며 저지현상의 연구에 있어 핵심적인 주제이다. 영어 동사의 불규칙 굴절형이 규칙적인 굴절형의 출현을 저지하는 것(예를 들어 brought-*bringed)이 여기에 해당한다.

함한 접사의 목록이다.

| | |
|---|---|
| 접두사 (18개) | 불(不)-, 비(非)-, 생(生)-, 무(無)-, 반(反)-, 맨-, 몰(沒)-, 뒤-, 짓-, 치-, 휘-, 드-, 들-, 풋-, 처-, 왕-, 갓(god)-, 꿀- |
| 접미사 (26개) | -가(家), -자(者), -인(人), -수(手), -이(명사 파생 접미사), -음, -기, -받, -되(피동사)-, -스럽-, -하(형용사 파생 접미사)-, -롭-, -지-, -적(的), -답-, -끼리, -꾼, -성(性), -화(化), -거리-, -대-, -이-(동사 파생 접미사), -당하-, -쟁이, -장이, -짱 |

표 2.1. 조사 대상 접사 목록

본 연구에서 현대 한국어 조사에 사용한 코퍼스는 국립국어원에서 2020년에 배포한 국립국어원 형태 분석 말뭉치(버전 1.0)이다. 이 코퍼스는 문어 200만 어절, 구어 100만 어절로 총 300만 어절 규모이다. 여기에 더해, 과거에 이미 출현한 항목을 판별하기 위해 세종 코퍼스를 참고하였다. 구체적으로는 세종 코퍼스의 현대 문어 원시 말뭉치와 현대 구어 원시 말뭉치를 활용했다. 현대 문어 원시 말뭉치는 1934년부터 2004년 사이에 출판된 텍스트를 모은 것으로 총 9,071,054어절 규모이다. 현대 구어 원시 말뭉치는 2001년에서 2005년 사이에 구축된 것으로 총 805,646어절 규모이다.

본 연구에서 사용한 코퍼스가 이광호(2009)에서 사용한 세종계획 형태 분석 코퍼스에 비해 갖는 가장 큰 장점은 구어 자료가 추가되었다는 점이다. 연구에 활용한 코퍼스에 구어 자료가 포함되지 않은 것은 이광호(2009)의 가장 큰 약점이었다.[14] 구어와 문어는 단어를 생산하는 경향이 다르므로 구어 자료가 포함되지 않으면 생산성 분석에 왜곡이 생길 수밖에 없다. 생산성과 다른

14) 이러한 약점은 연구자 스스로 인정한 바 있다. “구어 코퍼스가 연구 대상에 포함되지 못한 점은 자료적 측면에서 본 연구가 가지는 가장 큰 약점이라 할 수 있다.”(이광호 2009:294)
한편, 문어에서 신어가 발견되는 빈도가 구어에서보다 높다는 이유를 들어 문어 코퍼스만 사용하는 것을 옹호하는 주장도 있었다(Baayen and Renouf 1996:73). 문어의 생산성이 모든 접사에 걸쳐 더 높다는 점도 믿기 어렵지만, 이를 사실로 가정한다 해도 문어 코퍼스만을 사용하여 구한 P(생산성 추정치)는 실제 생산성보다 체계적으로 과다추정될 것이므로 역시 수긍하기 어렵다.

변수의 상관관계를 검증하고자 하는 본 연구의 특성을 고려하면 자료 차원에서의 체계적인 왜곡은 더욱 치명적일 수 있다.

그런데 국립국어원 형태 분석 말뭉치는 과거와는 다른 형태로 만들어져서 다루는 데에 기존의 한국어 코퍼스 분석 프로그램을 활용할 수 없었다. 따라서 본 연구에서는 프로그래밍 언어인 Python(버전 3.9.1.)을 사용하여 데이터를 직접 가공하였다. 이 과정에서 추가적인 패키지는 설치하지 않고 기본 모듈만 사용하였다.

끝으로, 각 접사의 파생어별 빈도 자료는 프로그램을 통해 1차 가공된 그대로를 사용하지 않고 최대한 직관에 따라 2차 수정을 거친 것이다. 특히 신어나 저빈도어로 나타난 단어는 본 논의에서 중요한 의미를 갖는 생산성이나 임계 빈도에 큰 영향을 미치므로 전부 일련의 수정 과정을 거쳤다. 따라서, 여러 차례에 걸친 검토 작업에도 불구하고 앞으로 제시할 자료에는 여느 통계 자료와 마찬가지로 조사자에 의한 오류가 약간은 존재할 수 있음을 밝혀 둔다.

## 2.2. 핵심 지표의 도출

코퍼스를 통해 수집한 접사별 파생어 빈도 분포 자료는 유형 저지나 항목 저지의 모습을 살펴보는 데에 두루 활용할 중요한 것이다. 그러나 본 논의에서 존재를 입증하고자 하는 유형 저지는 빈도의 단순한 나열만으로는 쉬이 인지하기 어렵다. 따라서 이를 시각화하기 위한 일련의 과정이 필요하다. 본 논의에서는 이를 위해 각 접사의 생산성 추정치 P와 저장 임계 빈도 추정치 θ라는 구체적인 값을 지표로 사용할 것이다.

두 값은 계량적 형태론 연구에 있어 단순히 시각화에 필요한 도구로서의 지표 이상의 의미가 있다. 사실, 이 분야에서 지금까지 있었던 대부분의 논의가 생산성과 단어의 빈도상 저장 지점을 수치화하는 데에 관련되었다고 해도 과언이 아니다. 본 논의에서도 이러한 핵심 지표들의 산출 방식을 보다 정교하게 다듬고자 고심하였다. 2.2에서는 이러한 내용을 다룰 것이다. 구체적으로는 각 개념의 정의와 필요성 및 실제로 값을 구하는 방법에 대해 논의하고자 한다.

### 2.2.1. 생산성

생산성은 저지와 아주 긴밀하게 연관된 개념이다. 두 개념 모두 파생어가 빈도와 관련을 맺고 출현하는 현상을 의미하기 때문이다. 저지에서 경쟁 파생어의 출현을 막는 힘은 그 파생어의 빈도에 비례한다.[15] 생산성은 기본적으로 신어의 출현 빈도를 의미하므로, 그 자체가 빈도적 개념이다. 생산성과 저지는 정의상의 연관성보다 좀 더 직접적으로 관련될 수 있다. 접사는 경쟁 접사의 신어 파생을 방해하기도 하는데, 이는 한 접사가 직접적으로 경쟁 접사의 신어 파생에 영향을 미치므로 일종의 저지현상으로 볼 수 있기 때문이다. 저지현상이 접사 간의 생산성 격차라는 형태로 드러나기도 한다는 것이다. 즉, 각 접사의 상태를 파악하거나 접사 간의 양적인 관계를 밝히는 데 있어 생산성을 구하는 과정은 필수적이라고 할 수 있다.

생산성에 대한 언어학 분야의 기존의 정의에서 벗어나 생산성을 수치화할 방법을 처음으로 제시한 것은 Baayen(1989)이다. 그는 생산성을 확률적으로 정의하고 추정치를 계산할 수 있다고 주장하였다. 그에 따르면 생산성은 '그 접사의 파생어를 마주쳤을 때 그것이 신어일 확률'로 정의할 수 있다.[16] 그렇다면 생산성은 '코퍼스에서 그 접사의 파생어를 마주쳤을 때 그것이 신어일 확률'이라는 통계량을 통해 추정할 수 있을 것이다. 코퍼스는 실제 사용된 언어라는 모집단에서 수집한 표본이기 때문이다.

이러한 논리 아래 그가 제시한 공식을 살펴보기 전에 생산성의 새로운 정의가 기존의 것과 달라서 혼동될 수 있는 부분을 짚고 넘어가고자 한다. 기존의 논의에서 사용하던 정의는 접사의 신어 생산력을 확률적으로 바라보지 않을 뿐만 아니라 어근과의 전반적인 결합 능력까지 포괄하기 때문이다. 다시 말해, 기존의 논의에서는 어떤 접사가 새로운 단어를 얼마나 만들어내는 편인가 하

15) 이에 대해서는 1.2에서 정리하였다.

16) 이 정의는 수학적인 것이기 때문에 이질적으로 느껴질 수 있다. 이는 수학의 극한이나 경제학의 '한계' 개념과 유사한 것이다. 지금까지 접한 단어들(어휘부)의 묶음이 주어졌을 때 추가적으로 마주치게 되는 단어 하나는 접해보지 못한 단어(신어)이거나 그렇지 않거나의 두 가지 경우 중 하나일 수밖에 없다. 따라서 생산성은 근본적으로 이 추가적인 하나의 단어가 신어일 확률만을 따지면 되는 것이다.

는 경향성(혹은 확률)은 전혀 염두에 두지 못한 반면, 신어와는 직접적인 관계가 없는 접사의 '세력(혹은 파생어의 총빈도 및 유형 빈도)'과 비슷한 의미로 생산성을 사용해왔다는 것이다. 이러한 차이는 접사의 총 빈도에 따라 극적으로 드러날 수 있다. 가령 총 빈도가 10회인 접사와 100회인 접사의 신어가 1회씩 관찰되었다면 새로운 정의에 따랐을 때에는 전자의 생산성이 10배 높지만, 기존의 정의에서는 후자가 오히려 10배 높다고 판단하였을 것이다. 접사를 많이 사용하는 것과 새롭게 사용하는 것은 별개이므로 용어 역시 분리하면 혼동이 적을 것이다. 새로운 용어를 마음대로 만들어 쓰는 것 역시 혼란을 초래할 수 있으므로 본 논의에서는 기존에 존재하였으며 계량언어학계에서 약간 다른 의미로 사용해온 생산성이라는 용어를 그대로 사용하되, 접사의 활발함을 일컫고자 할 때에는 '세력'과 같은 비유적인 표현을 사용하겠다.

이어서 Baayen(1989)의 생산성 산출 방법을 살펴보자. 일련의 수학적 도출 과정에 따라 그가 제시한 식은 다음과 같다. P는 실제 생산성의 산출된 추정치, $n_1$은 단발어의 수, N은 접사가 파생한 단어들의 총 빈도를 의미한다.

$$P = \frac{n_1}{N}$$

그런데 이러한 공식은 주로 수학적 논리에 의해 도출되었기 때문에 언어학적으로 바라보았을 때에는 구조적인 오차를 유발할 수 있는 취약점을 몇 가지 발견할 수 있다. 이광호(2009)에서는 이러한 취약점을 다음과 같이 네 가지로 정리하였다.

① $n_1$에 신어만이 아니라 원래 저빈도인 단어(몇몇 사람의 어휘부에만 존재하거나 극히 제한된 문맥에서만 사용되는 단어)가 포함된다.
② $n_1$에 통계적 우연(크기나 하위 텍스트 구성 등 코퍼스의 특성에 따른 우연)에 따라 나타난 단어가 포함된다.
③ 빈도가 1보다 큰 단어 중에도 신어가 관찰된다.
④ N에 어기와 접사가 분석적으로 인식·처리되는 단어만이 아니라 단일체로

인식되는 고빈도 단어까지 포함된다.

이 중 네 번째는 위의 공식에서 분모 부분에 해당하는 오류로 Baayen(1993)에서 이미 인식하고 수정하였던 것이다. N을 이렇게 수정해야 하는 이유는 한 접사가 파생한 단어가 머릿속 어휘부에서 모두 동일하게 처리되지 않는다는 선행 연구의 입장을 따랐기 때문이다. 빈도가 높은 파생어는 어기와 접사 사이의 결합 과정이 충분히 인식되지 않고 단일체로서 입력·출력되는 경향이 높다.[17] 그런데 생산성은 파생 과정을 거친 단어 중에 신어가 등장할 확률을 따지는 개념이므로 고빈도로 사용되어 단어 전체가 저장되었다가 인출된 경우는 제외해야 한다. 그렇지 않다면 분모 N이 너무 커져서 생산성이 과소추정된다. Baayen(1993)은 이를 수정하기 위해 '저장 임계 빈도'라는 개념을 도입하고 전체 접사에 대해 동일한 임의의 값으로 설정하였다. 이 빈도 이상으로 출현한 파생어는 파생 과정을 거쳤다고 볼 수 없다고 간주하여 분모인 총 빈도에서 제외하였다.[18]

③은 단발어가 아닌 2회 이상 출현어에서도 무시할 수 없는 수준의 신어가 관찰되므로 분자에 단발어를 사용하면 생산성이 과소추정될 수밖에 없다는 지적이다. 이 문제점은 Baayen and Renouf(1996)에서 일부 수정되었다. 이 연구의 조사 결과 다음과 같이 각 접사의 5회 출현어까지도 신어가 꽤 높은 빈도로 관찰되었기 때문이다. 신어가 단발어로만 나타나지 않는 데에는 여러 이유가 을 수 있으나, 주된 이유는 새로 만들어낸 말이더라도 같은 텍스트 내에서는 여러 번 사용하기가 쉽기 때문이다.[19] 이광호(2009)에서는 이와 같은 연구를 근거로 하여 신어를 판별할 빈도 기준을 최대 2회로 정하여 공식의 분자 부분을 수정하였다.

---

17) 이처럼 파생어가 통째로 저장 및 출력되거나 접사+어근의 분석 과정을 거쳐 새로 출력되는 두 가지 경로 중 하나를 거친다는 주장은 흔히 '이중 경로 이론(dual route theory)'라고 불리는 것이다. 이에 대해서는 2.2.2에서 살펴보기로 한다.

18) 저장 임계 빈도에 관한 이론은 2.3.1에서 살핀다. 한편, 이처럼 Baayen(1993)에서 기존의 공식의 약점에 대해 인정하고 수정한 부분은 이광호(2009)에서도 재차 지적된 바 있다.

19) 실제 텍스트를 살펴보면 둘 이상의 화자가 있는 경우에는 상대방의 말을 받는 경우, 한 화자의 발화인 경우에는 자신의 어휘를 자가복제하는 경우가 많아 신어의 빈도가 이처럼 부풀려지기 쉽다.

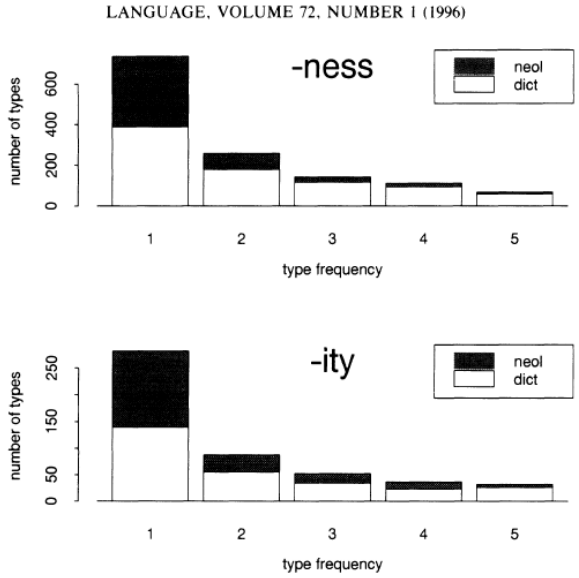



FIGURE 2. Numbers of types in *-ness* and *-ity* occurring in WIII dictionary (dict) and number of neologisms (neol) with respect to this dictionary for token frequencies 1–5.

그래프 2.1. Baayen and Renouf(1996)에서 제시한 '-ness'와 '-ity'의 파생어 빈도별 신어 분포. 가로축은 파생어의 빈도를, 세로 축에서 검정 색으로 칠해진 부분은 n(가로 축의 숫자)회 출현한 파생어 중에 신어가 총 몇 종류 있었는지를 나타낸다. 흰색 부분은 신어와의 대조를 위해 기존 어휘부인 사전의 빈도별 단어 분포를 제시한 것으로 보인다.

①, ② 역시 Baayen(1989)의 공식에서 분자에 해당하는 부분의 약점을 지적한 것이다. 단발어에 의한 추정치도 코퍼스의 크기가 수천만 어절 이상으로 매우 커지면 실제 신어를 이용한 값과 거의 차이가 없어지지만, 몇백만 어절 단위의 중·소규모 코퍼스를 이용한 연구에서 단발어로 신어를 대체하면 생산성이 과대 추정된다. 포함하지 말아야 할 단어들을 신어로 처리하게 되기 때문이다. 따라서 440만 어절 규모의 코퍼스를 활용한 이광호(2009)에서도 산출 공식의 분자 부분을 최대한 신어에 가깝게 정제하는 작업이 필요했다.[20]

20) 이광호(2009)에서는 Baayen(1989)의 비판하며 단발어 자체가 신어를 측정하기에 적합하지 않은 지표라고 하였는데 이는 조금 부정확한 지적이다. Baayen의 초기 논의에서 사용된 단발어의 수 $n_1$은 N이 무한히 커질 때 실제 신어의 수 n에 접근하는 추정치로, 수학적 극한의 측면에서 다룬 것이지 신어를 그대로 대체하는 것은 아니다. 실제로 Baayen은 여러 후

이상의 수정과정을 종합한 이광호(2009)의 최종적인 생산성 산출 공식은 다음과 같다.

$$P = \frac{\text{N(2회 이하로 출현한 신어의 빈도의 합)}}{\text{N(어기-접사의 결합과정이 인정되는 파생어)}}$$

분모에서 '어기-접사의 결합과정이 인정되는 파생어'는 Baayen(1993)이 제안한 '저장 임계 빈도 이하의 파생어'와 같은 의미이다. 이광호(2009)는 임의로 정한 빈도 50을 저장 임계 빈도로 사용하였다. 임의로 임계점을 정한 것은 이 공식의 가장 큰 약점이다.

본 연구에서는 앞선 연구에서 제안된 생산성 산출 공식의 한계를 극복하기 위해 공식의 일부를 재수정하였다. 가장 중요한 수정 사항은 접사마다 언어학적 근거가 있는 저장 임계 빈도 추정치를 각각 도출하고 이를 통해 P의 분모를 계산했다는 점이다. 각 접사의 올바른 저장 임계 빈도를 사용하면 생산성의 추정이 정확해짐은 물론이고 더 나아가 임계 빈도와 생산성이라는 두 지표를 두 변수로 삼아 상관관계를 확인할 수도 있게 된다. 이 관계를 알아내면 모든 접사의 전체적인 추세에 부합하는 잠재 생산성을 구할 수 있다. 즉, 어떤 세타(θ, 저장 임계 빈도의 추정치) 수준에서 일반적인 P의 크기가 대략 어느 정도가 될 것인지 예측이 가능해지며, 따라서 좀 더 객관적인 기준에서 생산성의 높고 낮음을 판단할 수 있게 된다. 또, 본 연구에서 정의한 유형 저지를 포착하려면 두 변수 사이의 상관관계를 반드시 먼저 파악해야만 한다.

다음으로, 신어를 확인할 때 빈도 상한선을 두지 않았다. 신어의 대부분이 2회 이하 출현어에 포함된다는 선행 연구 결과가 있었지만, 나머지 소수의 신어를 완전히 포기하면 생산성 추정치가 왜곡된다. 신어는 생산성 추정 공식의 분자에 해당하는데, 코퍼스의 크기가 작아지면 분모도 작아지기 때문이다. 이는 형태분석 말뭉치의 규모가 작은 현대 한국어 연구의 여건에서 특히 중요하다. 규모가 작은 말뭉치에서는 1-2회 출현어의 빈도까지도 절대적으로 별로

---

속 연구에서 단발어를 신어 대신 사용하기 위해서는 충분한 크기의 코퍼스 사용이 전제되어야 한다고 강조한 바 있다.

크지 않으므로 3회 이상의 상대적인 고빈도 구간에서 간혹 출현하는 신어가 생산성 추정치를 크게 바꿀 수 있는 것이다.

본 연구에서는 신어 확인을 위해 5회 출현어까지는 전부 기존의 코퍼스와 사전에 출현하였는지 확인하고, 그 이상의 고빈도 단어에 대해서는 신어로 의심되는 단어를 확인하였다. 어떤 신어가 여러 번 출현했다면 같은 텍스트에서 동일한 화자가 반복해서 발화했을 가능성이 있다. 이런 식으로 추가된 빈도는 분석에 그대로 반영하기 어렵다. 따라서 본 연구에서는 고빈도로 출현한 모든 신어를 말뭉치 파일에서 직접 찾아 확인하고, 동일한 텍스트에서 반복 출현한 신어는 1회 출현으로 처리하였다.[21]

본 연구에서는 이상의 재수정 과정을 거친 다음과 같은 생산성 산출 공식을 분석에 사용하였다. θ는 연구 대상 코퍼스를 통해 구한 저장 임계 빈도 추정치를 의미한다.

$$P = \frac{N(\text{신어의 빈도의 합})}{F_{\theta}(\text{빈도가 } \theta \text{ 이하인 파생어들의 빈도의 합})}$$

### 2.2.2. 단어의 저장과 빈도

전술하였듯 저지나 생산성과 같은 파생에 관한 형태론적 현상과 빈도는 밀접하게 연관되어 있다. 따라서 저지나 생산성을 제대로 이해하기 위해서는 파생어의 빈도 자료를 정확하게 분석할 수 있어야 한다. 그런데 복잡한 숫자로 이루어진 빈도 자료를 직관에 의해서만 살핀다면 정확한 분석이 이루어지기 어렵다. 따라서 빈도 분석을 위한 일종의 도구를 고안할 필요가 있다.

생산성 산출에 필요한 빈도 분석 이론은 심리언어학적·계량언어학적 이론을

21) 이 작업은 대체로 신어나 저빈도 구어에 대해서만 행하였으므로 전반적으로 약간의 왜곡이 있을 수 있다. 그러나 조사 대상 코퍼스의 비교적 짧은 편인 텍스트에서 같은 단어가 여러 번 중복 출현하는 일은 매우 드물다. 그럼에도 신어를 이처럼 텍스트 단위로 정제한 것은 분자에 해당하는 부분의 크기가 매우 작아서 약간의 차이에도 큰 차이를 낳기 때문이다. 그에 비해 분모는 본 논의에서 대부분 세자리 이상으로 큰 값으로 나타나기 때문에 작은 오차가 결과에 별다른 영향을 주지는 않는다. 세력이 약하여 총빈도가 40, 123과 같이 매우 작은 경우도 있었는데, 이런 접사들은 모든 파생어에 대해 텍스트 중복 출현 검증을 하여 오차를 줄이려고 노력하였다.

기반으로 한 선행 연구에 기대고 있다. 이 연구들에 따르면 어떤 접사에 의한 파생어들은 단어가 형성될 때 심리언어학적으로 두 가지 다른 경로 중 하나를 따른다. 이 두 부류는 코퍼스에 나타난 파생어를 빈도순으로 나열했을 때 고빈도와 저빈도 구간이 확연히 나뉘는 것으로 확인할 수 있다. 여기서 두 부류로 나뉘는 경계가 되는 빈도를 분석 임계 빈도(parsing threshold)라고 하며, 반대로 저장 임계 빈도라고 부를 수 있다.

본 연구에서는 모든 접사에 걸쳐 연구자가 임의로 정한 임계 빈도를 사용하던 선행 연구의 방식에서 벗어나, 각 접사의 파생어 빈도 구조를 통해 언어학적인 근거를 가진 값을 도출하고자 하였다. 빈도 구조를 파악하기 위한 빈도 분석 방법의 핵심은 크게 세 가지로 요약할 수 있다. 그것은 이중 경로 이론(dual route theory)과 상대 빈도, 마지막으로 파생어의 빈도 순위 자료를 도입했다는 점이다. 순위-빈도 분석 도구는 이 이론들을 차례로 도입하여 단점을 보완하고 장점을 조합한 것이다. 2.2.2.1에서는 위 요소들이 본 논의의 방법론에 도입된 과정을 차례로 살펴보고자 한다. 그러면 2.2.2.2에서 소개할 빈도 분석 도구의 최종적인 모습을 쉽게 이해할 수 있을 것이다.

#### 2.2.2.1. 기반 이론

이중 경로 이론은 본 논의에서 구성한 방법론의 가장 기초가 된다. 저장 임계 빈도가 논리적으로 중요한 만큼, 그 값을 구하기에 앞서 이러한 지점이 존재하며 빈도를 통해 확인할 수 있다는 점이 확실히 논증되어야 할 것이다. 임계점의 존재를 일찍이 주장한 것이 바로 소위 이중 경로 이론이다. 이중 경로 이론이란 언어 사용자가 어떤 어휘를 사용할 때 분석(parsing)과 단어 전체 인출(whole word access) 중 하나로부터 마치 갈림길에서 경로를 정하듯 선택한다는 심리언어학적 이론이다.[22]

---

22) 물론 실제 언어 사용 양상을 보면 파생어의 성격이 어떤 빈도를 기점으로 완전히 극적으로 달라지지는 않는다. 이 빈도는 실질적인 임계점보다는, 코퍼스를 통해 수집된 여러 언어 사용자들의 전반적인 사용 행태를 살펴보았을 때 두 부류의 파생어가 만나서 겹쳐지는 중간 지점에 가깝다. 본 연구에서는 선행 연구의 용어를 받아들여 '저장 임계 빈도'를 그대로 사용하지만 '임계'의 개념까지 그대로 수용한 것은 아니다. 즉 이 임계점은 '언어 사용자 전체를 놓고 봤을 때 저장과 분석의 경향이 더 강한 집단을 나눌 수 있는 지점'이지, '모든 언어

이를 빈도와 접목하여 설명하는 연구도 뒤이어 이루어졌다. Anshen & Aronoff(1988)에 따르면 한 접사(혹은 유형)의 파생어들은 저빈도와 고빈도 항목에서 서로 다른 빈도 분포를 보이며, 언어 사용자들이 파생어를 어근-접사로 분해하여 인식하는 능력도 두 부류에서 유의미한 차이를 보인다.[23] 다시 말해 인지적으로 접사-어근이 분석되는 경향이 더 뚜렷한 파생어들은 그렇지 않은 단어들과 인지언어학적 실험을 통해서만이 아니라 빈도 분포의 계량적인 분석을 통해서도 거의 정확하게 구분할 수 있었다는 것이다.

이것은 어휘부에 대한 기존의 생성문법적 시각을 심리언어학적 입장에서 전면적으로 수정하는 주장이었다. 즉, 이중 경로 이론에 따르면 파생어는 기존의 이론대로 파생됨과 동시에 무조건 어휘부에 통째로 저장되고 인출되는 것이 아니라, 빈도가 높은 항목만 저장되고 낮은 빈도의 파생어는 필요할 때마다 어근과 접사의 결합 과정을 거쳐 만들어진다.

이중 경로 이론은 이처럼 심리언어학적인 접근을 통해 제기되었지만, 관점을 달리하면 특정 빈도를 기점으로 단어가 저장되기 시작한다는 계량언어학적 측면에서 해석할 수도 있다. 저장 임계 빈도는 바로 이러한 시각에서 고안된 개념이다. 실제 분석에서는 저장 임계 빈도의 추정치를 의미하는 세타(θ)를 사용하는데, 세타는 분석 경로를 통하지 않고 통째로 인출된 항목을 생산성 산출에서 제외하는 기준이 된다(Baayen 1993).

그런데 Baayen(1993)이 세타를 생산성 산출에 도입한 것은 타당한 선택으로 보이지만, 전술하였듯 자의적으로 정한 세타를 모든 접사에 걸쳐 획일적으로 적용한 데에는 동의하기 어렵다. 그는 코퍼스의 크기만을 고려하여 '100만 어절당 몇 회'와 같이 설정했을 뿐 접사의 파생어 집합의 크기는 고려하지 않았다. 세타를 접사의 빈도를 고려하지 않고 동일하게, 절대 빈도로 설정하면 납득하기 어려운 결과를 얻게 된다. θ = 10을 설정했다고 가정해 보자. 이 기준은 코퍼스에서 10번 등장한 접사와 1000번 등장한 접사에 동일하게 적용

---

사용자가 기계적으로 저장과 분석을 전환하는 지점'은 아니다. 사실 후자의 정의는 계량적 연구에서 추구하는 탐구 방식에 따른 것이라고 보기 어려우며 실험을 통한 논증이 가능한 연역적 연구의 영역에 가까운 것으로 보인다.

23) Baayen and Lieber (1991), Baayen and Neijt(1997), Hay and Baayen(2002a) 등은 Anshen and Aronoff(1988) 이후 이중 경로 이론의 형태론적 타당성을 논증하거나 이중 경로 이론을 형태론적 분석에 활용한 주요 연구 업적이다.

되므로, 전자는 무조건 분석 경로를 통해 만들어지고 후자는 대부분 통째로 인출된다고 주장해야만 할 것이다. 이는 직관적으로도 틀렸지만 코퍼스에서 확인할 수 있는 통계적 사실과도 합치하지 않는다.

따라서 다음으로는 방법론에 상대 빈도를 도입하게 되었다. 절대 빈도를 사용한 언어 자료 분석에는 이처럼 문제가 많은데다, 특히 신어 생산이나 단어 분석(parsing)과 같이 빈도와 연관이 있는 형태론적 현상은 상대 빈도에 따라 행동하는 것으로 알려져 있기 때문이다. Malicka-Kleparska(1985)는 어근에 비해 파생어의 빈도가 상대적으로 낮을수록 저지력이 약해진다고 보고하였다. Hay(2001)은 (어근의 빈도/파생어의 빈도)[24]와 같은 상대 빈도를 사용했을 때 단순히 'n(회)'이라는 형태로만 표시 가능한 절대 빈도를 사용했을 때보다 파생어 내부의 투명성을 잘 분석할 수 있음을 보였다.

상대 빈도를 사용하면 모든 접사에 동일한 세타를 적용하지 않고 각 접사의 규모에 알맞는 세타를 설정할 수 있다는 장점이 있다. 가령 임계 빈도를 θ = 10이 아니라 θ = (파생어의 빈도/접사의 빈도) = 0.01과 같은 식으로 정한다면 앞서 지적한 문제는 사라지게 된다.[25]

그러나 이런 방식으로 상대 빈도를 사용할 경우 통계적인 분석에 있어 또 다른 심각한 문제가 발생하게 된다. 빈도 분석을 위해서는 접사가 파생하는 파생어 항목마다 다른 값을 갖는 변수가 필요한데, 같은 접사의 파생어끼리는 '접사' 변수의 값이 모두 같기 때문이다. 그러므로 어떤 접사 파생어들의 빈도 분포를 파악하기 위해 평면상에 각 파생어를 (접사의 빈도, 파생어의 빈도)와 같이 표시할 경우, 이는 일직선상에 나열된 점들로 나타나게 된다. 즉, (1000, 1), (1000, 2)… 따위의 모습을 띨 것이다. 이런 경우 두 변수가 서로의 값에 따라 변화하면서 나타나는 빈도 분포를 관찰할 수 없다. 억지로 분석을 시도하더라도 결국 파생어들의 빈도를 단순히 나열하고 직관에 따라 해석하는 기존의 방식과 다를 바 없다. 따라서 접사 빈도와 파생어의 빈도를 변수로 하는 상대 빈도 분석은 무의미하다.

Hay and Baayen (2002)에서는 이러한 한계를 인식해서인지 어근과 파생어

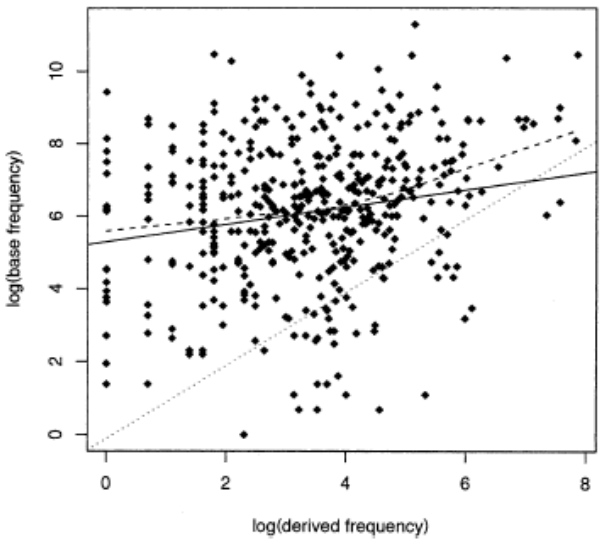
---

24) (어근의 빈도, 파생어의 빈도)와 같이 좌표로 표시하더라도 여전히 상대 빈도라고 볼 수 있다.

25) 물론 여전히 기준을 임의로 정했다는 문제는 남는다.

의 상대 빈도를 활용한 '분석 기준선(parsing line)'을 고안하였다. 이 연구에서는 분석에 상대 빈도를 활용하였으므로 접사마다 별도의 빈도 분석을 수행할 수는 있었다. 그러나 접사와 파생어가 아닌 어근과 파생어 간의 상대 빈도를 통해 임계 지점을 찾으려는 부자연스러운 시도를 하는 바람에 검증되지 않은 수학적 도구를 과도하게 사용하면서도 설명력이 낮은 설명을 하는 데 그치고 말았다. 이 논의는 너무 복잡하므로 그 기반이 된 Hay(2001)의 빈도 분석 방식을 그래프 2.2에 소개한다. 상대 빈도를 사용한 덕분에 각 접사가 두 변수 (파생어의 빈도, 어근의 빈도)에 따라 어떻게 분포하는지는 관찰할 수 있으나 여기에 어떤 수학적인 공식에 의해 도출된 곡선을 그려서 분석-저장을 설명한다는 것은 사실상 불가능해 보인다. 게다가 어근의 빈도와 파생어의 빈도 사이의 관계는 통계적으로 유의(significant)하지만 상관계수는 매우 낮다. 즉, 어근과 파생어의 상대 빈도를 통해 파생어의 저장을 설정하기에는 변수의 설명력도 부족한 것이다.[26]

26) 그래프 2.2에서 확인할 수 있듯 두 변수 사이에 그다지 뚜렷한 상관관계를 찾기 어려움을 알 수 있다.

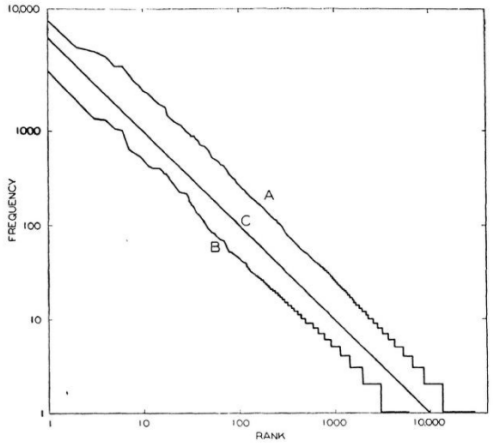



그래프 2.2. Hay(2001)에서 제시한 515가지 영어 접두파생어의 어근-파생어 분포. 피어슨 상관계수와 스피어만 상관계수 등의 값이 유의하기는 하지만(모두 $p<0.001$) 각각 0.2, 0.21으로 매우 작다. 즉 두 축의 변수는 서로 그다지 연관되지 않은 것이다.

이처럼 파생어에 대한 계량적 연구에 상대 빈도를 사용하면 개별적인 분석이 가능해지지만, 이를 통해 저장 임계점을 찾기는 어렵다. (파생어의 빈도, 접사의 빈도)를 변수로 삼으면 접사의 빈도가 모든 파생어에서 동일한 값을 갖기 때문에 아무런 분석을 할 수 없고, (파생어의 빈도, 어근의 빈도)를 변수로 삼으면 임계점을 찾기가 매우 어려운데다 두 변수 사이에 통계적인 관계도 미미하다. 따라서 우리의 목적을 달성하기 위해서는 파생어마다 다른 값을 띠면서도 파생어나 접사에서 벗어나지는 않는 제3의 변수를 찾아야만 한다. 또한 그 변수를 활용하는 완전히 새로운 빈도 분석 방법도 필요할 것이다.

파생어의 빈도 순위 자료는 이런 맥락에서 분석에 도입하게 된 것이다. 이때 빈도 순위란 같은 접사에 의한 파생어들을 빈도순으로 나열하였을 때의 순위를 말한다. 어떤 단어 집합(코퍼스) 내 단어들의 빈도-빈도 순위 간에 일정한 분포를 찾을 수 있다는 사실은 Zipf(1949)가 발견한 것이다. 그에 따르면

어떤 자연어 코퍼스에 등장하는 단어의 빈도는 순위의 역수에 비례한다.[27] 1위의 빈도가 2위의 빈도에 비해 대략 두 배, 3위에 비해서는 대략 세 배...라는 뜻이다. 따라서 같은 코퍼스에 속하는 각 단어의 빈도와 순위에 대해 로그를 취하고 순위-빈도 평면에 표시하면 그래프 2.3과 같이 직선에 가까운 형태로 나타나게 된다. 이러한 실증적 법칙을 지프의 법칙이라고 한다. 지프의 법칙은 코퍼스를 가리지 않고 성립한다. 따라서 이와 같은 분포는 언어가 갖는 '자연스러운' 모습이라고 할 수 있다.



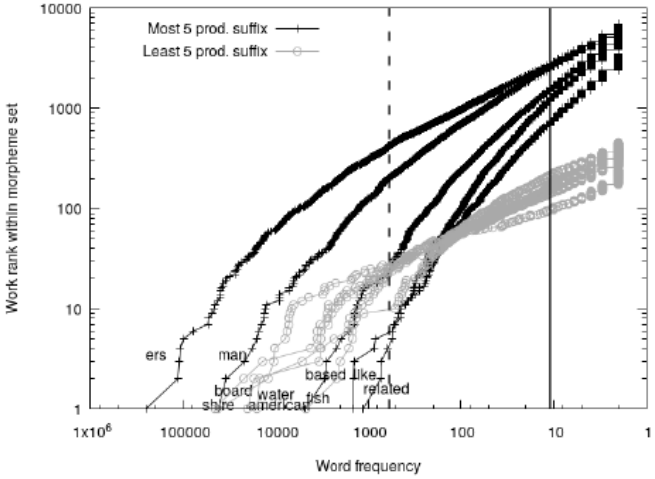

그래프 2.3. 단어의 순위-빈도 분포(Zipf 1949). A, B, C는 각각 서로 다른 코퍼스에 속한 단어들의 분포를 나타낸 것이다.

단어의 빈도가 순위와 일정한 관계를 맺는다면 순위-빈도 분석은 파생어와 접사의 분석에 사용하기에 더할 나위 없이 적합한 도구가 될 수 있다. 각 파생어 항목의 순위는 2.1.2에서 지적한 문제점을 하나도 갖지 않으면서도 파생어 항목마다 다른 값을 갖는, 훌륭한 '제3의 변수'이기 때문이다.

그런데 지프의 법칙을 파생어에 바로 적용하기는 어렵다. 각 접사의 파생어 집합이 일반적인 코퍼스를 구성하는 단어의 집합과 같은 분포를 가지지는 않

27) 여기서의 코퍼스는 어떤 종류의 자연어 집합도 포함할 수 있는 개념으로, 분야와 텍스트 크기를 조절하여 인위적으로 구축한 현대의 코퍼스와 다르다. Zipf(1949)에서 활용한 코퍼스는 소설 『율리시스』였다.

을 것이기 때문이다. 접사의 세력이나 생산성에 따라 그 접사가 파생한 어휘의 빈도 분포는 크게 달라진다.[28] 그런데 지프의 법칙을 '같은 어휘 집합에서는 순위-빈도 분포가 일정하게 나타난다'는 측면에서 이해한다면 파생어의 빈도 분석과 세타의 도출에 활용할 수 있다. Pierrehumbert & Granell(2018)에서는 몇몇 영어 접사의 파생어에 대해 순위-빈도 분석을 수행하여 이에 대한 가능성을 보여주었다.




그래프 2.4. 일부 접사의 순위-빈도 분포(Pierrehumbert & Granell 2018). 선으로 이어진 각 계열은 접사 하나의 파생어 집합이다. 진한 선은 조사 대상으로 선정한 접사 중 가장 생산적인 접사 5개, 연한 선은 가장 비생산적인 접사 5개의 분포이다.

순위-빈도 그래프를 보면 오른쪽 끝에서 시작하여 저빈도 구간에서 유지되던 패턴이 특정 빈도에 다다르면 깨지는 것을 볼 수 있다. Pierrehumbert & Granell(2018)은 대략 절대 빈도 600 수준에서 패턴의 변화가 관찰된다고 결론지었다. 앞서 소개한 표현에 따르면 $\theta$ = 600이라는 결론이다. 그러나 이는

28) 3장에 제시한 여러 접사의 분포를 통해 이를 확인할 수 있다. 활발히 사용되는 '-하-', '-적(的)' 등의 접사는 거의 모든 구간이 일정한 분포를 보이지만 세력이 미약한 '-꾼', '-지-' 등의 분포는 불규칙적이다.

모든 접사의 계열이 절대적으로 같은 빈도 수준에서 저장될 것이라는 가정이 전제된 결론이므로 언어학적으로 타당하다고 볼 수 없다. 실제로 위의 그래프에서도 접사마다 저빈도에서의 패턴으로부터 벗어나는 지점이 다르게 나타나고 있는데도 가정에 결론을 억지로 맞춘 셈이다.[29] 다시 말해, 이 연구는 분석에 상대 빈도를 사용하지 않아서 잘못된 결론에 이르게 되었다. 세타의 측정에 상대 빈도를 사용하면 절대 빈도로 표시된 값을 표준화시켜 서로 비교할 수 있게 해주는 장점이 있다. 이것은 앞서 생산성을 계산할 때 분모에 접사(접사의 파생어)가 출현한 총빈도를 사용한 것과 평행하다.

순위-빈도 분석에 상대 빈도를 사용하면 비로소 접사 별 세타를 구할 수 있게 된다. 순위-빈도 패턴이 저빈도의 패턴에서 처음으로 크게 벗어나는 지점의 빈도가 바로 세타라고 할 수 있다. 접사마다 고유한 세타를 구하면 생산성 추정치도 올바르게 구할 수 있다. 또 두 값을 모두 구해야 여러 접사에 걸쳐 세타와 추정 생산성과의 상관관계를 계산하고 잠재 생산성 개념을 분석에 활용할 수 있다.

한편, 이와 같은 빈도 분석 방법은 이제까지 언어학 연구에서 채택한 방식에 비해 훨씬 입체적인 정보를 제공해 준다. 일종의 순위-빈도 분석을 시도한 선행 연구가 있었으나 파생어의 순위별 빈도를 표로 만들어서 직관에 따라 저빈도어와 고빈도어로 양분하는 정도에 그쳤다. 대규모 수치 자료를 직관에 따라 해석하면 저장 임계 빈도를 도출하는 정교한 작업을 할 수 없음은 물론이고, 로그값이 아닌 자연수에 기반한 판단을 하게 되므로 부정확한 결론에 이르게 된다.

#### 2.2.2.2. 빈도 분석의 실제

앞서 소개한 이론들은 모두 논리에 약점을 갖고 있지만 저장 임계 빈도에 대해 저마다 타당한 주장을 하고 있으며 서로 약점을 보완하는 관계에 있다. 이들을 하나의 방법론으로 통합하면 보다 신뢰할 수 있는 저장 임계 빈도를 도

29) 로그 좌표를 사용하고 있으므로 이탈 지점의 차이가 실제로는 상당히 크다. 빈도 분석에 로그 좌표를 사용하는 것은 일반적인 좌표에 비해 언어 현상에 대한 설명력이 탁월하기 때문이다. 지프의 법칙도 로그 좌표의 설명력을 뒷받침하는 한 예가 될 수 있다.

출할 수 있을 것이다.

본 연구에서는 파생어의 빈도 분석에 있어 기본적으로 Pierrehumbert & Granell(2018)의 순위-빈도 분석 방법을 따르되, 분석에 파생어의 상대 빈도를 사용하고 각 접사마다 세타를 따로 구할 것이다. 빈도 분포를 시각화한 뒤에 단절점을 찾는 단계에서는 그래프를 보고 직접 판단할 수밖에 없다. 따라서 이때의 기준을 세우지 않으면 분석이 혼란스러워질 것이다. 본 연구에서 분포의 단절점을 판별한 기준들은 다음과 같이 정리할 수 있다.

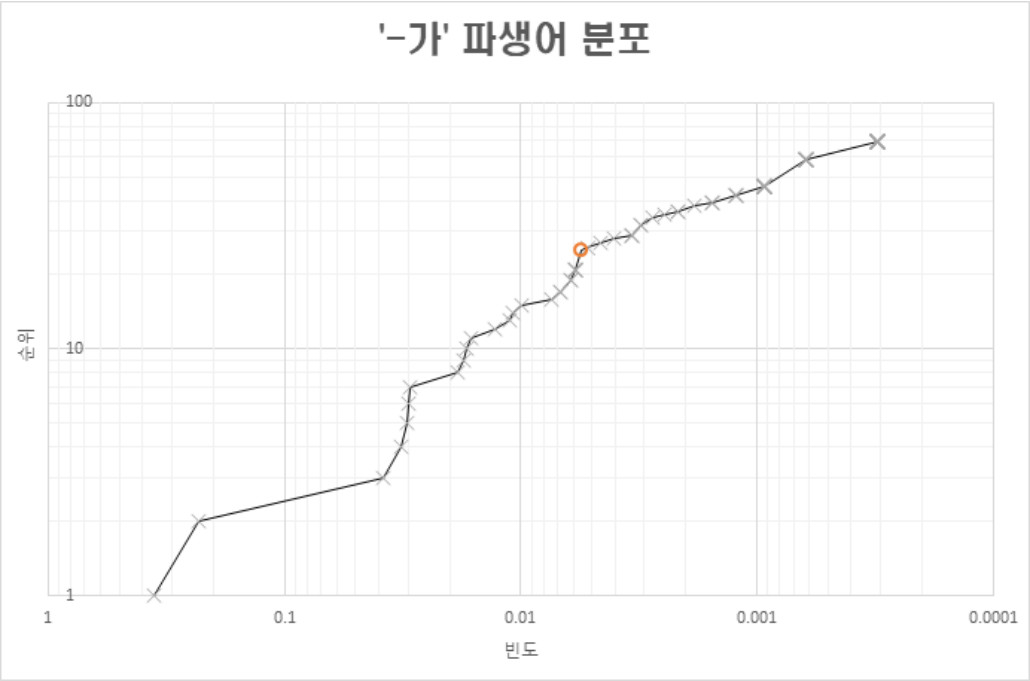
① 고빈도 구간은 로그를 취하더라도 불규칙적으로 나타나는 경우가 대부분이므로 최대한 저빈도에서 단절점을 찾는다.

② 단순히 그래프가 꺾이는 점을 찾는 것이 아니라 패턴의 변화하는 지점을 찾는다.

②-1. 그 지점까지 유지되던 패턴이 명확하다면 그것에서 급격히 이탈하는 지점을 찾는다.

②-2. 저빈도 구간의 패턴이 어지럽거나, 그 지점에서 충분히 이탈한 것인지 헷갈릴 때에는 전체적인 분포 패턴을 고려한다.

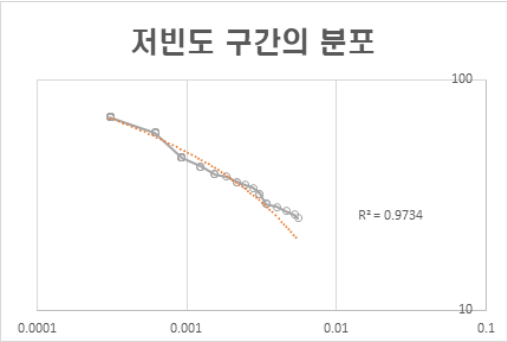
①에서 고빈도 구간이 불규칙적이라고 한 것은 대부분의 접사 파생어가 고빈도의 몇 단어에 크게 편중되어 있기 때문이다. 일반적인 자연어 코퍼스였다면 지프의 법칙에 따라 순위에 따라 일정한 빈도 변화를 보였겠지만 각 접사의 파생어 집합은 그렇지 않다. 그래프 2.5의 명사 파생 접사 '-가(家)'의 순위-빈도 분포는 이러한 분포의 전형적인 모습을 보여준다.

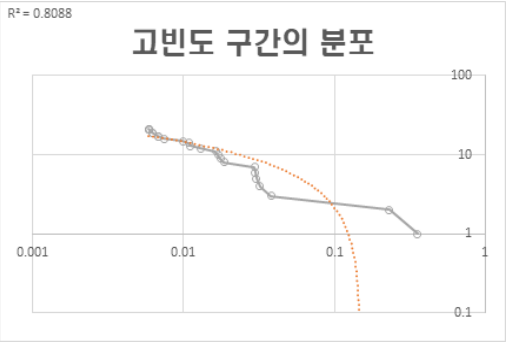



그래프 2.5. '-가' 파생어 분포

'-가'의 파생어 분포는 오른쪽 끝에서 시작하는 가장 낮은 빈도의 구간에서 대체로 직선에 가까운 모습을 보여주다가 ○ 기호로 나타낸 지점에서 급격히 꺾이고, 그 이후로는 매우 불규칙적으로 변화한다. 이러한 차이는 그래프 2.6, 2.7과 같이 로그 $R^2$을 구하면 어느 정도 객관적인 수치로 확인할 수 있다. $R^2$이 1에 가까울수록 내가 제시한 식(로그 방정식)에 관측치가 잘 부합하여 규칙적이라는 뜻이다.[30]

30) 관측치의 규칙성은 흔히 거듭제곱 $R^2$을 사용하지만 여기서는 로그 좌표를 사용하고 있어서 거듭제곱으로는 규칙적/불규칙적 분포 간에 값의 차이가 확연하지 않은데다 후술할 '-하-'와 같이 저빈도 구간의 규칙성이 비선형적으로 나타나는 경우가 많으므로 로그를 사용하는 편이 바람직하다.



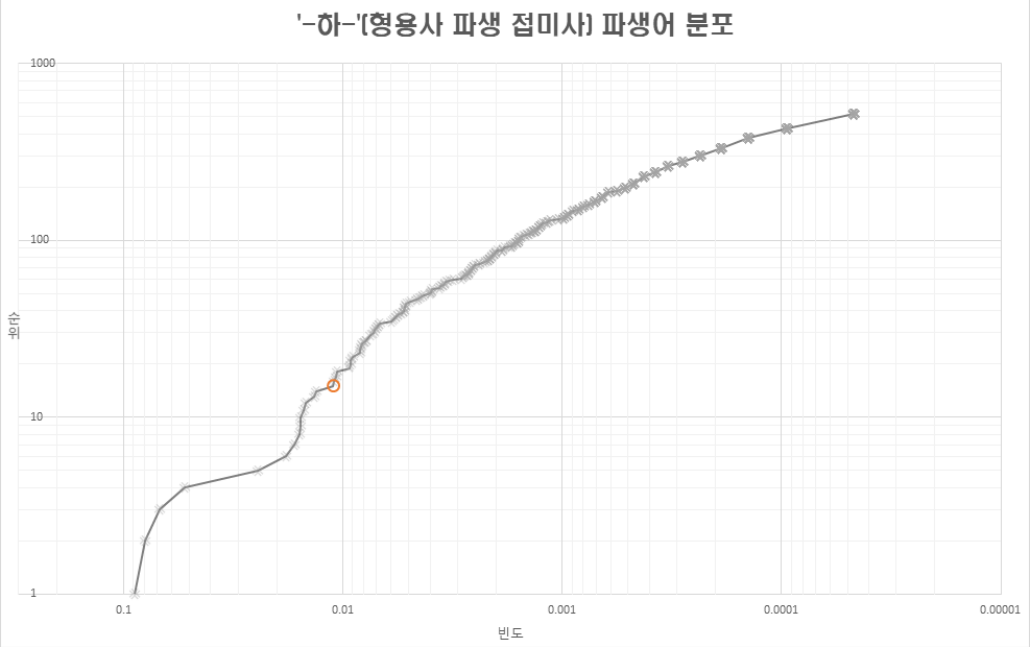

그래프 2.6. '-가'의 저빈도 구간 분포　　그래프 2.7. '-가'의 고빈도 구간 분포

저빈도 구간의 $R^2$값은 0.973이고 고빈도 구간에서는 0.801이므로 두 구간의 규칙성이 크게 차이난다는 것을 객관적으로도 확인할 수 있다.[31] 그러므로 그래프 2.5에 표시한 지점을 저장 임계점이라고 할 수 있다. 이 점은 '안무가'에 해당한다. '안무가'는 코퍼스에서 총 18회 등장하였고, 접사의 총 빈도가 3245이므로 세타는 약 0.00555(=18/3245)가 된다.[32]

그런데 저빈도의 규칙적 패턴이 반드시 직선에 가까워야 하는 것은 아니다. '-하-', '-적(的)'과 같이 매우 활발하게 사용되는 접사는 저빈도 구간의 파생어 분포가 아래쪽을 기준으로 오목한 형태를 띠는 특징이 있다. 다음의 그래프 2.8은 '-하-'의 파생어 분포이다.

31) 그러나 $R^2$값은 어디까지나 앞서 제시한 기준을 보조하는 지표로만 활용하여야 한다. 즉, 저장 임계점이 저빈도와 고빈도 구간의 $R^2$값 차이가 극대화되는 지점과 동일하지는 않다. '-가'의 경우에도 순위 7 지점(왼쪽에서 7번째 점)에서 나누면 두 구간의 $R^2$값 격차가 근소한 격차로 최대가 되지만, 이는 가장 낮은 구간에서 찾아볼 수 있던 직선에 가까운 패턴이 명백히 깨진 뒤이므로 저장 임계점이라고 할 수 없다.

32) 따라서 생산성 추정치 P의 분모는 (18회 이하로 출현한 '-가' 파생어의 총 빈도)가 된다. 한편, 여기서는 임계점을 찾는 원칙을 보이기 위해 $R^2$값까지 제시한 것으로, 3장에서는 이를 생략하고 결과만을 바로 제시하였다.




그래프 2.8. '-하-' 파생어 분포

저빈도 구간에서 곡선에 가깝게 이어지던 분포 패턴이 ○ 부근에서 크게 변화하는 것을 알 수 있다. 관측치가 매우 촘촘하게 분포하고 있으므로 직전이나 그 직후의 변곡점을 선택할 수도 있을 것으로 보인다. 이는 연구자가 판단할 몫이지만, 만일 연구자의 자유로운 선택에 의해 최종적인 결과가 좌우된다면 큰 허점이 될 수 있다. 그러나 이 정도의 작은 차이가 결과에 미치는 영향은 오차로 볼 수 있을 만큼 미미하다.[33] 언어학적인 측면에서도 이 지점이 이웃하는 파생어 중 정확히 어디인지를 따지는 것은 큰 의미가 없다. 전술하였듯 저장 임계 빈도는 그 용어와는 달리 실제로 그 지점을 기점으로 전혀 저장되지 않던 파생어가 완전히 저장되기 시작하는 지점을 의미하지 않으며, 주로 저장 혹은 분석에 의존하는 두 이질적인 집단이 만나는 접경에 가깝기 때문이다. 이러한 접경은 마치 회색지대처럼 정확한 선을 긋기 어려운 형태로 나타날 수 있는 것이다.

33) 4.2에서 논의의 타당함을 따지기 위해 계산할 최종적인 지표인 Kendall's tau 값은 약 0.02, P-값은 약 0.001정도 달라지는 데 그친다. 이러한 변화는 결론에 아무런 영향을 주지 못한다.

그래프에 표시한 저장 임계점은 '불안하'에 해당한다. 이 단어는 코퍼스에서 198회 등장하였고, '-하-' 파생어의 총 빈도는 21234이므로 세타는 0.00933(=198/21234)가 된다.

그래프 2.8. '-롭-' 파생어 분포

기준 ②-2는 '-롭-'과 같은 분포를 분석할 때 필요하다. 초반의 분포에서 처음으로 크게 벗어난 지점을 찾아야 하는데, 그 지점을 ○가 아니라 △쯤이라고 주장할 수도 있는 것이다. '-하-'에서와 달리 두 지점 사이의 빈도 차이는 무시할 수 없을 만큼 크다. △에서의 이탈이 충분히 큰 것인지는 어떻게 판단할 수 있는가. 이 때 우리는 한 접사의 파생어를 크게 둘로 나누고 있음을 기억해야 한다. △보다는 ○ 지점에서 이탈의 강도도 훨씬 강하고, 전체적인 분포를 고려했을 때 자연스럽게 이등분할 수 있다.[34]

34) 결국에는 자료를 해석하는 사람의 직관이 개입될 수밖에 없는 것이다. 이것을 모형화할 알

이렇게 찾은 저장 임계점은 '향기롭'에 해당한다. 이 단어는 코퍼스에서 8회 등장하였고, '-롭-' 파생어의 총 빈도가 621이므로 산출된 저장 임계 빈도는 0.01288(=8/621)이다.

고리즘을 만들어내는 것은 매우 어려울 뿐만 아니라 불필요해 보이기도 한다. 그래프 2.4와 같이 거의 모든 통계 분석 연구에서도 마지막에 자료를 해석하는 것은 결국 직관의 영역에서 이루어진다.

# 3. 파생 접사의 빈도 특성

## 3.1. 도입

3장에서는 본 연구에서 선정한 접사의 파생력이나 경쟁 양상과 같은 공시적인 상태를 확인하기 위해 파생어의 빈도 특성에 대해 살펴본다. 빈도 특성은 순위 분포 그래프를 통해 확인할 수 있는 분포 형태나 생산성 추정치(P), 저장 임계 빈도 추정치(θ)와 같은 접사 고유의 값들을 포함하는 개념이다.

가공되지 않은 자료인 코퍼스로부터 언어학적인 의미를 가진 빈도 특성을 구하기 위해서는 2장에서 설명한 방법론을 일련의 과정에 따라 적용해야 한다. 우선 프로그램을 통해 코퍼스를 가공하여 파생어와 빈도의 목록을 구한다. 다음으로 목록 상의 모든 파생어 항목에 대해 신어인지 확인하는 작업을 거치면 모든 전처리가 끝나고 순위-빈도 분포를 구할 수 있다.

빈도 특성은 순위-빈도 분포에서 모두 드러나므로, 3장에서는 이해를 돕기 위해 분포도를 각 접사마다 제시할 것이다. 순위-빈도 분포를 통해 접사의 양적 구조를 분석하고 신어 목록과 θ(저장 임계 빈도의 추정치), P(생산성의 추정치)를 구하여 제시할 것이다. 이때, 이광호(2009)와 겹치는 접사에 대해서는 필요할 경우 통시적 변화도 살펴보고자 한다. 접사의 공시적인 모습은 통시적인 세력 변화까지를 고려할 때 제대로 파악할 수 있기 때문이다. 이렇게 파악한 접사의 모습은 4장에서 두 가지 저지에 관해 논의할 기초 자료가 된다. 접사별 파생어들의 빈도는 항목 저지의 양상을 파악하기 위해 반드시 필요하고, 각 접사의 P와 θ를 정리한 자료는 신어 파생에서 유형 저지가 일어난다는 사실을 밝히기 위해 필수적이다.

한편, 접사에 대한 언어학적 분석은 의미적인 면에 집중하였다. 접사의 의미적 특성은 계량적인 특성과 직접적으로 연관되어 나타나기 때문에 의미적 분석을 더하면 자료 해석이 한층 다각화된다는 장점이 있다.[35] 분석 결과를 제시할 때는 파생어와 그 빈도의 전체 목록은 전부 나열하지 않고 신어에 대해서만 간략하게 보였다. 순위-빈도 그래프에서도 2장에서 그랬던 것처럼 모든 파생어가 나타내는 점을 표시

35) 의미와 자료의 상관관계는 접사별 분석에서도 어느 정도 드러나지만 접사 간 분석에서는 확연하게 나타난다. 자료의 접사 간 분석은 4장에서 진행할 것이다.

하지 않고 저장 임계점에 해당하는 파생어만 표시하여 시인성을 높였다.

본격적인 분석에 들어가기 전에, 3장에서 제시하는 P와 θ는 실제 생산성과 저장 임계 빈도가 아님을 짚어둔다. 계산한 값을 실제 값으로 볼 수 없는 이유는 다음과 같다. 첫째, 코퍼스는 어디까지나 실제 사용된 말의 총체에서 극히 일부만을 표본으로 가져온 것이기 때문이다. 둘째, 본 연구에서 사용한 분석 방법은 물론 어떤 분석 방법을 사용하더라도 자료의 분석 과정에서 왜곡이 발생할 수밖에 없기 때문이다. 추정치가 실제 값을 성공적으로 예측한다는 검증이 이루어지기 전까지는 두 개념을 분리해야 한다.[36] 따라서 자료를 통해 구한 각 접사의 생산성과 저장 임계 빈도는 '산출된 생산성', '산출된 저장 임계 빈도' 혹은 P, θ와 같은 기호로 나타내기로 한다.

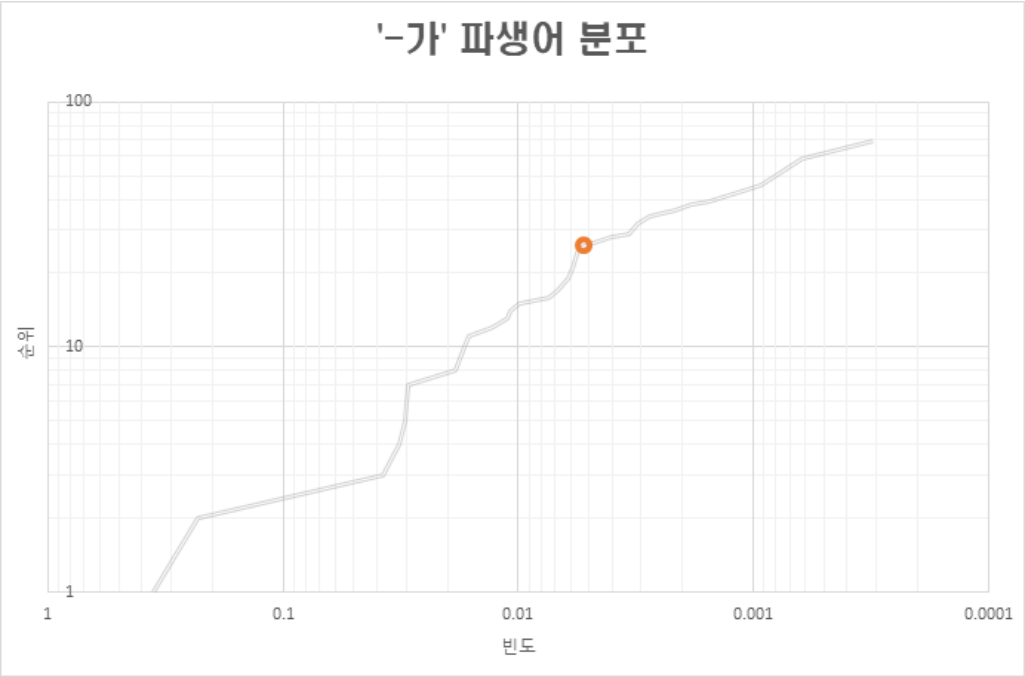
## 3.2. 생산적 접사

### 3.2.1. '－가(家)', '－자(者)', '－인(人)'[37]

#### 3.2.1.1. '－가(家)'

36) 이러한 검증은 4장에서 진행하였다.

37) '-사(師)', '-사(士)'도 '사람'을 의미하는 대표적인 한자어 접미사지만 '-사(師), -사(士)'는 분석 대상에 포함하지 않았다. 기존의 용법에 맞지 않는 새로운 단어가 나타났을 때 그것이 둘 중 어떤 접사에 의한 파생어인지 판단하기가 쉽지 않기 때문이다. 가령, 조사 대상 코퍼스에서 사람을 의미하며 '-사' 형태로 나타난 파생명사 중 신어는 '전략가'의 의미로 쓰인 '전략사' 하나였는데, 화자가 이 단어를 만들어 낼 때 둘 중 어떤 접사를 사용하였는지는 알 수 없다.




| 파생어 | 빈도 |
|---|---|
| 탐서가 | 1 |
| 인권가 | 1 |
| 야망가 | 1 |
| 교습가 | 1 |
| 반역가 | 1 |
| 총빈도: 3244<br>신어: 5 | |

표 3.1.. '-가(家)'의 신어와 빈도

'-가(家)'는 일부 명사와 결합하여 '그것을 직업으로 하는 사람', '그것에 능한 사람', '그것이나 그 특성을 많이 가진 사람'의 뜻을 더하는 접사이다. 순위-빈도 그래프에 표시된 세타 파생어[38]는 '안무가'이며 빈도 18이다. 세타 이하인 파생어의 총빈도는 270이다. 산출된 생산성과 세타(저장 임계 빈도)는 다음과 같다.

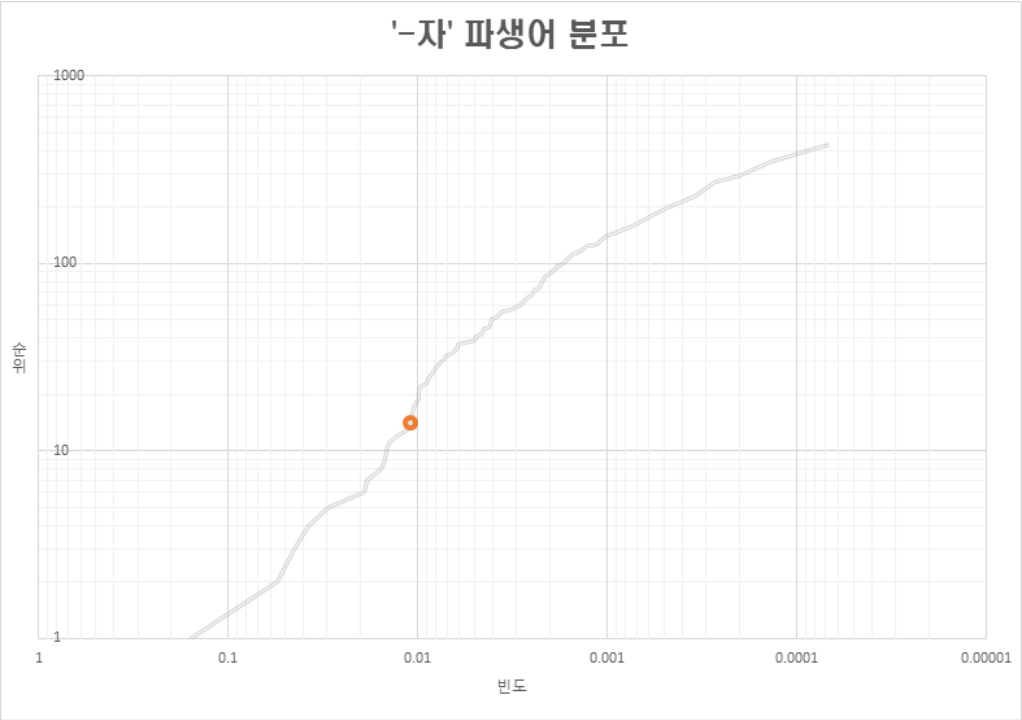
38) 저장 임계점에 있는 파생어를 편의상 '세타 파생어'라고 하였다.

P=0.01852
θ=0.00555

‘-가(家)’의 파생어 중에는 형태적으로 동일하지만 분석을 다르게 해야 할 것으로 보이는 것들이 있다. ‘-가(家)’의 파생어로 나타나는 ‘X화가’가 그 예가 될 수 있다. 이 유형으로는 ‘벽화가, 풍속화가, 한국화가, 동양화가, 문인화가, 서양화가, 만화가’ 등이 관찰되었다. 이들은 의미적인 관점에서 [[X+화][-가]]로 분석하는 것이 타당할 듯하지만, 단어에 따라서는 [[X][화가]]라고 분석해야 하는지 고민하게 된다. 본 연구에서는 이런 경우 ‘X화’의 어휘부 내에서의 위치가 확고하여 사전에 표제어로 등록되어 있다면 ‘X화가’ 전체를 별개의 ‘-가’ 파생어로 보았다. 즉, ‘-가’ 앞의 어근에 자립성이 있는 경우에만 별개로 처리하였다. 만약 그렇지 않다면 일률적으로 ‘화가’ 항목에 포함된 것으로 보고 빈도를 더했다. 이러한 기준에 따르면 ‘벽화가, 풍속화가, 동양화가, 문인화가, 서양화가, 만화가’는 별도의 파생어로 취급되고 ‘한국화가’는 ‘화가’에 통합된다. 이상의 파생어 처리 기준에 따라 ‘시민운동가’와 ‘독립운동가’도 별도로 분류하였다.

한편, ‘-가’와 함께 사람을 의미하며 유의 관계에 있는 접미 한자어로 ‘-자(者)’가 있다. 두 접사는 의미가 대체로 비슷하지만 구별된다(노명희 2005:253). 두 접사가 분포에서 차이를 보이는 것은 이러한 의미 차이에 기인하는 것으로 보인다. ‘-가’는 파생어가 ‘~을 직업으로 하는 사람’, ‘~을 직업처럼 자주 하는 사람’의 의미를 띠는 어근과 배타적으로 결합한다. ‘-가’와 ‘-자’의 배타적 결합은 항목 저지로 분류할 수 있다. 이에 대해서는 4장에서 다룰 것이다.

#### 3.2.1.2. ‘-자(者)’




| 파생어 | 빈도 | 파생어 | 빈도 |
|---|---|---|---|
| 전문자 | 1 | 직위자 | 1 |
| 출사자 | 1 | 초연자 | 1 |
| 최고점자 | 1 | 피감자 | 1 |
| 패용자 | 1 | 행진자 | 1 |
| 휴무자 | 1 | 회생자 | 1 |
| 위험자 | 1 | 산화자 | 1 |
| 개통자 | 1 | 협론자 | 1 |
| 격리자 | 1 | 감면자 | 2 |
| 결함자 | 1 | 면세자 | 2 |
| 내사자 | 1 | 시상자 | 2 |
| 단식자 | 1 | 신규자 | 2 |
| 면책자 | 1 | 영주자 | 2 |
| 모의자 | 1 | 종결자 | 2 |
| 상환자 | 1 | 추정자 | 2 |
| 시구자 | 1 | 파병자 | 2 |
| 약정자 | 1 | 포획자 | 2 |
| 운송자 | 1 | 휴가자 | 2 |
| 운행자 | 1 | 주력자 | 2 |
| 전형자 | 1 | 재범자 | 2 |

| 제한자 | 1 | 채용자 | 3 |
|---|---|---|---|
| 총빈도: 14936<br>신어: 54 | | | |

표 3.2.1. '-자'의 신어와 빈도

'-자(者)'는 '-가(家)'와 마찬가지로 일부 명사와 결합하여 '~하는 사람', '~의 특성을 가진 사람'의 뜻을 더하는 접미사이다. 순위-빈도 그래프에 표시된 파생어는 공동 14위인 '탈북자', '당사자'이며 빈도 163이다. 세타 이하인 파생어의 총빈도는 8262이다. 산출된 생산성과 세타는 다음과 같다.

P=0.00654
θ=0.01091

'-자(者)' 파생어는 어근의 음절 수에 따라 어근이 자립적인 정도가 다른 것으로 보인다. 노명희(2005:251)은 1음절 어근은 2음절 어근과 달리 비자립적인 어근으로만 나타나지만 2음절(이상)인 어근은 그렇지 않다고 하였다. 그렇다면 1음절 어근이 2음절 어근에 비해 접사로부터 분리하기 어려울 것으로 예상해볼 수 있다. 다음은 본 연구에서 국립국어원 형태 분석 말뭉치에 나타난 '-자'의 1음절 어근 파생어와 그 빈도이다.

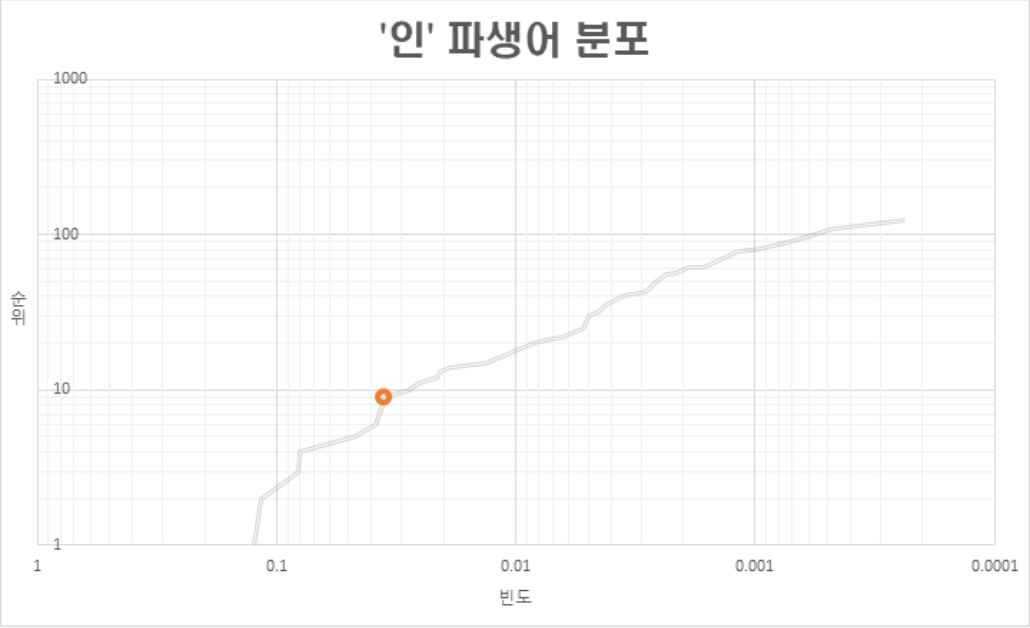
| 파생어 | 빈도 | 파생어 | 빈도 |
|---|---|---|---|
| 기자 | 1540 | 신자 | 35 |
| 환자 | 819 | 필자 | 33 |
| 타자 | 299 | 패자 | 14 |
| 부자 | 274 | 현자 | 7 |
| 학자 | 214 | 망자 | 7 |
| 주자 | 210 | 화자 | 6 |
| 저자 | 167 | 병자 | 5 |
| 약자 | 93 | 목자 | 2 |
| 승자 | 50 | 성자 | 2 |
| 업자 | 49 | 임자 | 1 |
| 강자 | 39 | | |

표 3.2.2. '-자'의 1음절 어근 파생어와 빈도

위의 표를 보면 실제로 1음절 어근은 세타 빈도 이상으로 출현하는 비율이 매

우 높고, 세타 빈도 아래의 단어들도 전체적으로 빈도가 아주 높은 것을 알 수 있다.[39] 2장에서 선행 연구를 통해 설명하였듯 파생어의 빈도가 높을수록 접사와 어근의 결합 강도는 강해진다. 이로써 선행 연구에서 어근의 음절 수와 자립도에 대해 직관을 통해 해석했던 것이 계량적 방법을 통해 어느 정도 입증된 셈이다.

#### 3.2.1.3. '-인(人)'




| 파생어 | 빈도 | 파생어 | 빈도 |
|---|---|---|---|
| 공무인 | 1 | 중계인 | 1 |
| 한수원인 | 1 | 합창인 | 1 |
| 감례인 | 1 | 미납인 | 1 |

39) '-자' 파생어 전체의 유형 빈도는 623이고 이 중 세타 빈도를 초과한 파생어는 13 항목이다. 유형 빈도란 어떤 접사(또는 어휘항목 형성 과정)가 파생한 항목의 수를 의미한다. '-자'의 유형 빈도가 623이라는 것은 코퍼스에서 '-자' 파생어가 623가지 관찰되었다는 뜻이다.

| | | | |
|---|---|---|---|
| 뮤지컬인 | 1 | 검도인 | 1 |
| 오지인 | 1 | 해국인(害國-) | 1 |
| 성균인 | 1 | 수협인 | 1 |
| 계량인 | 1 | CJ인 | 1 |
| 측량인 | 1 | KAIST인 | 1 |
| 하역인 | 1 | 한센인 | 2 |
| 도외인 | 1 | 총빈도: 1853, 신어: 20 | |

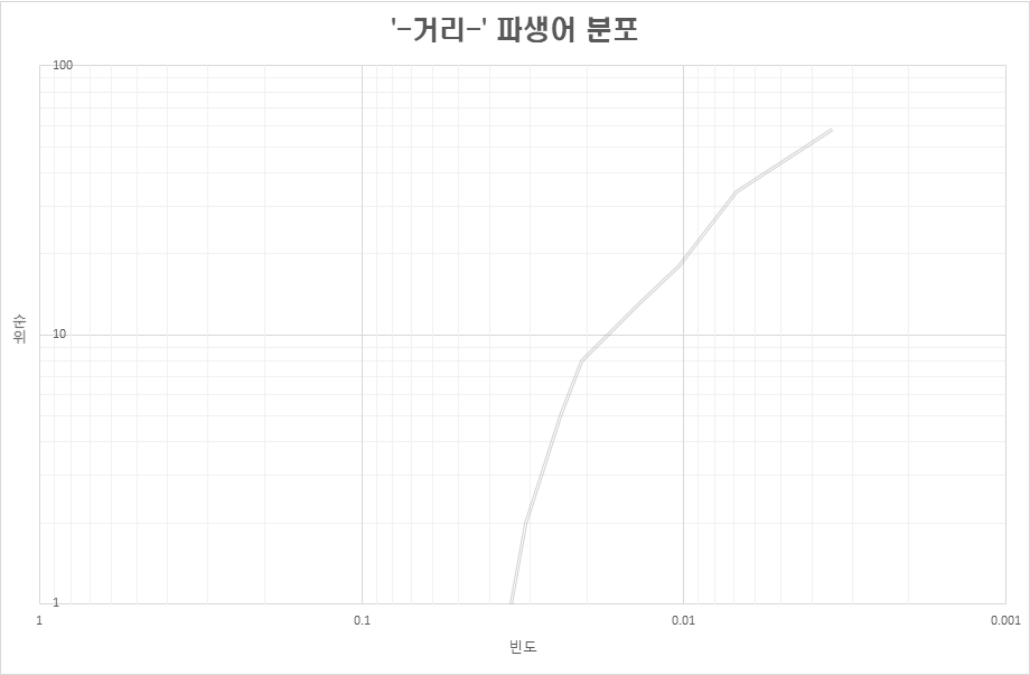
표 3.3. '-인'의 신어와 빈도

'−인'은 '−자', '−가'에서 파생하는 '~하는·한 사람'이나 '그러한 (특징이 있는) 사람', '그것을 직업으로 하거나 직업처럼 하는 사람'과 같은 의미 외에도 '그러한 집단에 속한 사람'이라는 특수한 의미를 더할 수 있는 명사 파생 접미사이다. '−인'의 신어는 대부분 이 특수 의미로 파생되었다. 순위−빈도 그래프에 표시된 파생어는 순위 9, 빈도 149인 '일반인'이다. 세타 이하인 파생어의 총빈도는 1853이며 산출된 생산성과 세타는 다음과 같다.

P=0.010793
θ=0.035375

### 3.2.2. '−거리−', '−대−'

#### 3.2.2.1. '−거리−'




| 파생어 | 빈도 |
|---|---|
| 훌쩍훌쩍거리 | 1 |
| 폭신거리 | 1 |
| 띡띡거리 | 1 |
| 화끈화끈거리 | 1 |
| 하늘하늘거리 | 1 |
| 뻐덕거리 | 1 |
| 뭉그럭거리 | 1 |
| 까실거리 | 1 |
| 버석버석거리 | 1 |
| 틱틱거리 | 3 |
| 총빈도: 290<br>신어: 12 | |

표 3.4.1. '-거리-'의 신어와 빈도

'-거리-'는 동작성, 상태성 어근 및 의성·의태어 어근과 결합하여 '어근이 의미하는 동작이 반복됨'의 의미를 더하는 접미사이다. 선행 연구에서 '-거리-'는 '-대-'와 함께 가장 생산적인 동사 파생 접미사로 분류되어왔는데, 이광호(2009)에서 이미 이것이 사실임을 코퍼스를 통해 확인한 바 있다. 본 연구에

서도 '-거리-'는 상당히 높은 생산성을 유지하고 있는 것으로 나타났다. 산출된 생산성은 다음과 같다.

P=0.04138[40]

'-거리-'의 분포에서는 저빈도 구간을 분리해낼 만한 분포 패턴의 단절을 관찰할 수 없다. 따라서 저장 임계 빈도를 구할 수 없다. 유형 빈도는 149로 적은 편이 아니지만 최고 빈도인 파생어도 10회 출현하는 데 그칠 만큼 유달리 낮고 고른 빈도 분포를 보여준다. 이는 접사가 어근으로부터 분리되어 인식되는 경향이 매우 큰 접사에 기대할 수 있는 빈도 특성이다. 접사가 어근과 분리된다면 그 원인은 접사에 있거나 어근에 있을 것이다.

먼저 어근을 살펴보자. '-거리-'와 결합하는 어근은 소리 그대로를 옮긴 '아작', '삐걱'과 같은 의성어나 행동을 묘사한 '기웃', '까딱' 등의 의태어가 대부분이고, 필요에 따라 반복 합성하여 '삐걱삐걱', '기웃기웃'과 같이 언제든지 의미 변화 없이 따로 떼어 사용할 수 있으므로 독립성이 높다고 할 수 있다. '-거리-'가 높은 생산성을 띠는 것도 이처럼 거의 무한히 만들어낼 수 있는 의성어나 의태어와 자유롭게 결합하기 때문이라고 설명할 수 있다.

그러나 결합 가능한 어근을 '-거리-'와 많이 공유하는 '-이-'에서는 이러한 빈도 특성을 찾아볼 수 없으므로 어근에 의한 설명은 설득력이 없다. 그렇다면 '-거리-'에서 나타나는 접사와 어근의 분리 경향성의 원인은 접사에 있다고 보아야 한다. 이와 관련해 일찍이 '-거리-'가 어근 사이에 다른 요소의 개입을 허용한다는 지적이 제기된 바 있다(이병근 1986:400 재인용). 가령, '출랑거리다'의 어근과 접사 사이에는 조사 '도, 만'이나 접미사 '-들'을 삽입할 수 있다.[41]

한편, '-거리-'가 항상 이와 같은 빈도 특성을 가졌던 것은 아니다. 이광호(2009)에서 조사한 결과에 따르면 저빈도 구간과 대비되는 압도적인 고빈도

40) 저장 임계 빈도를 구할 수 없으므로 (신어/총빈도)로 계산하였다. 저장 임계 빈도를 구할 수 없는 다른 접사들의 P도 이렇게 계산한 것이다.

41) 이와 같은 어근분리현상을 보이는 국어의 파생 접미사는 동사나 형용사 파생 접미사에 국한된다. 이러한 접사로는 '-거리-' 외에 '-하-', '-스럽-', '-대-'가 있다(송철의 1992:194). 본 연구에서 분석한 바에 따르면 '-당하-'도 여기에 포함될 수 있을 듯하다.

구간이 나타난다. 해당 연구와 본 연구의 조사 결과를 비교해 보면 '-거리-'의 통시적 변화의 속도가 빠르고 확연하다는 것을 알 수 있다. 파생어 전체를 나열한 뒤 비교하면 너무 번잡스러워지므로 다음과 같이 전체적인 분포에 대한 이해를 돕는 수치를 통해 비교하였다. '1위 빈도'는 해당 접사의 항목들 중 최고빈도를 보인 파생어의 빈도를 의미한다.

| | 이광호(2009) | 본 연구 |
|---|---|---|
| 유형 빈도 | 538 | 148 |
| 총빈도 | 3994 | 290 |
| 1위 빈도 | 398 | 10 |

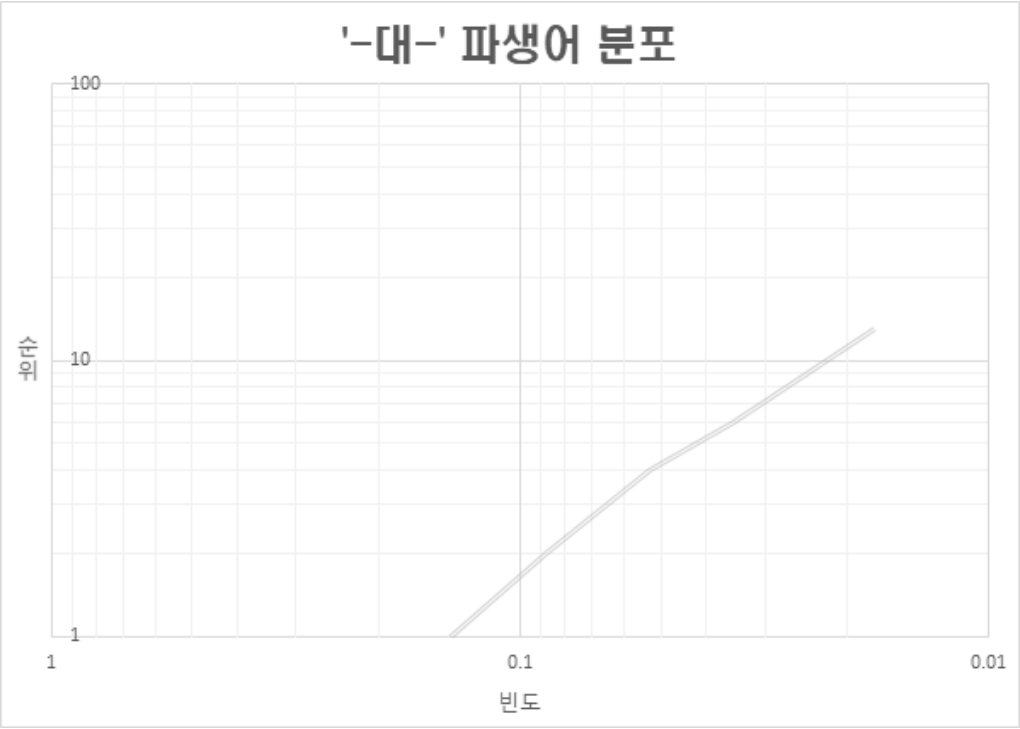
표 3.4.2. '-거리-'의 통시적 변화

이는 두 연구에서 활용한 코퍼스의 크기 차이(440만 대 300만)를 감안하더라도 급격한 변화이다. 변화의 원인을 추론하는 것은 본 연구의 주제를 벗어난다. 다만, 선행 연구를 참고하면 빈도의 급격한 감소가 접사 자체의 쇠퇴를 암시한다는 것만큼은 부정하기 힘들 듯하다. 예를 들어 Dalton-Puffer and Plag(2000:236)에서는 일반적인 접사의 분포를 따르는 접사의 경우 그 세력을 유지하도록 하는 것은 자주 사용되지는 않지만 다수인 저빈도 파생어들이라는 점을 보인 바 있다.[42] 유형 빈도가 줄어들 때 그 감소분은 대부분 저빈도어에서 발생하므로 표 3.4.2의 정보를 통해 해당 접사의 파생력이 약해지는 중이라고 추정할 수 있는 것이다. 이에 관해 4장에서는 접사의 통시적인 세력 변화와 접사 간의 경쟁이 서로에게 미치는 영향에 대해 살펴볼 것이다.

다음으로는 '-대-'의 분포를 살펴본다. '-대-'는 의미가 비슷한 '-거리-', '-이-'와 경쟁 관계에 있는 것으로 기존에 논의된 바 있다. '-이-'가 파생한 신어는 관찰되지 않았으므로 비생산적 접사를 다룬 3.3에서 살펴볼 것이다.

#### 3.2.2.2. '-대-'

42) 일반적인 접사의 분포는 소수의 고빈도 파생어와 다수의 저빈도 파생어로 구성된 분포를 의미한다.




| 파생어 | 빈도 |
|---|---|
| 우걱우걱대 | 1 |
| 틱틱대 | 1 |
| 총빈도: 57 | |

표 3.5.1. '-대-'의 신어와 빈도

'−대−'는 동작성, 상태성 어근 및 의성·의태어 어근과 결합하여 '어근이 의미하는 동작이 반복됨'의 의미를 더하는 접사이다. '−거리−'와 결합 대상은 물론이고 의미도 별다른 차이가 없다.

'−대−'의 분포에서는 저빈도와 고빈도 구간 사이의 분포 패턴의 단절을 확인할 수 없다. 따라서 저장 임계 빈도를 구할 수 없다. '−대−'가 '−거리−'와 마찬가지로 접사와 어근의 분리 현상을 보인다는 점이 선행 연구에서 지적되었다. 산출된 생산성은 다음과 같다.

P=0.03509

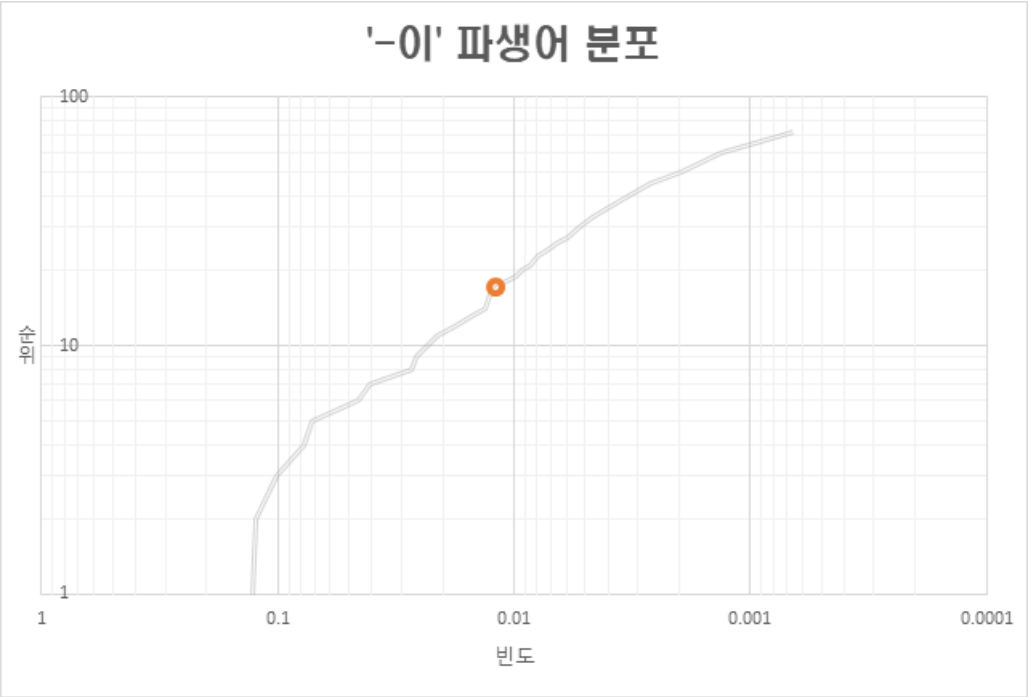
‘-대-’ 역시 ‘-거리-’와 마찬가지로 통시적으로 접사의 세력이 빠르게 약화되었다. ‘-대-’는 이광호(2009)에서 선택한 코퍼스가 수집된 2000~2005년경에 이미 세력이 상당히 위축되어 있었다는 점에서 ‘-거리-’와 다르다. ‘-거리-’와 ‘-대-’는 거의 같은 뜻으로 대부분의 어근에서 대체 관계에 있는데, 3.1.5와 3.1.6에서 살펴본 통시적 변화를 고려했을 때 이 두 접사의 파생 과정 전체가 실제 언어 사용에서 덜 사용되는 추세인 것으로 보인다. 이광호(2009)와 본 연구의 조사 결과를 비교하면 다음과 같다.

| | 이광호(2009) | 본 연구 |
|---|---|---|
| 유형 빈도 | 139 | 32 |
| 총빈도 | 451 | 57 |
| 1위 빈도 | 28 | 8 |

표 3.5.2. ‘-대-’의 통시적 변화

### 3.2.3. ‘-이’, ‘-음’, ‘-기’

#### 3.2.3.1. ‘-이’




| 파생어 | 빈도 |
|---|---|
| 떼잡이 | 1 |
| 통돌이 | 1 |
| 총빈도: 1510<br>유형 빈도: 103<br>신어:2 | |

표 3.6.1. '-이'의 신어와 빈도

'-이'는 주로 동사 및 형용사 어간, 때로는 다른 명사를 어근으로 취하여 명사를 파생하는 접미사이다. 파생 과정에서 '그러한 일을 하는 사람이나 동물, 사물' 혹은 '그러한 일에 사용되는 도구'의 의미를 더하기도 하지만 별다른 의미를 더하지 않는 경우도 많다. 예를 들어, '젖먹이'에서의 '-이'는 '그것을 하는 사람'이라는 의미를 더하지만 '턱걸이'의 '-이'는 명사 파생의 기능만을 수행한다.

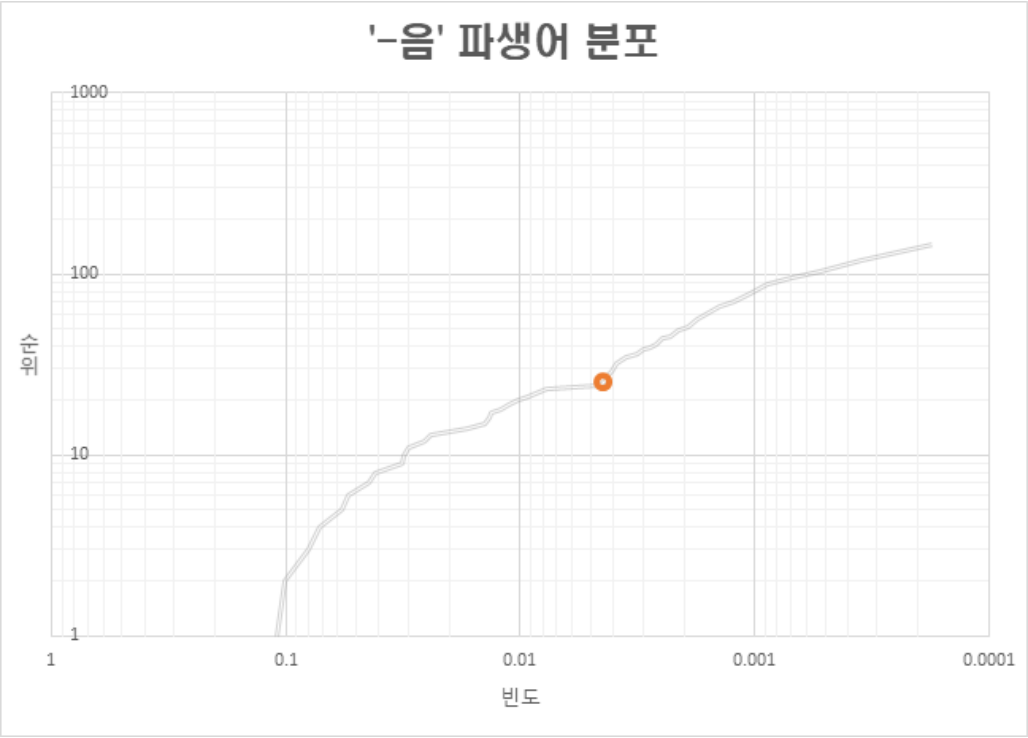
순위-빈도 그래프에 표시된 세타 파생어는 '꽂이'이며 빈도 18이다. 세타 이하인 파생어의 총빈도는 345이다. 산출된 생산성과 세타는 다음과 같다.

P=0.0058
θ=0.01158

이광호(2009)에서는 '-이'의 생산성을 0으로 계산하였지만, 이는 당시 산출 방식의 오류로 인해 과소추정된 값이다. 본 연구와 결과의 차이가 나는 것은 생산성 도출 방식이 수정되고 구어 코퍼스가 추가되었기 때문이다. 2장에서 설명하였듯 본 연구에서는 신어가 등장할 수 있는 빈도의 상한을 5까지 높이고 접사 파생어들의 분포에 따라 유동적인 세타를 도입하였다. 실제로 더 생산적으로 변화했을 수도 있지만, 완전히 비생산적인 접사가 짧은 시간 내에 생산성을 가지게 되었으리라고는 생각하기 어렵다.

동사, 형용사 어간과 결합하여 단순히 명사 파생만 하는 용도로 쓰인 '-이'는 '-음, -기'와 의미 차이는 별로 없으면서 거의 배타적으로 결합하므로 전형적인 항목 저지 관계에 있다. 이와 같은 특성은 선행 연구에서 한국어 저지 현상의 대표적인 예로 소개된 바 있다(송철의 1992: 124, 160). 그러나 신어 생산 경쟁에서는 접사의 전반적인 의미 차이가 영향을 크게 미친 것으로 보인다. 전술하였듯 '-이'의 파생어 중에는 '다듬이, 오른손잡이'처럼 어근에 별도의 의미를 더하는 단어가 '-음, -기' 파생어에 비해 많다. '-이'의 신어 역시 '사람, 도구'를 뜻하는 단어뿐이다. 다음으로는 '-음', '-기'를 차례로 다루며 각 접사 파생어의 의미적 특징을 살펴보기로 한다.

#### 3.2.3.2. '-음'




| 파생어 | 빈도 |
|---|---|
| 옆트임 | 1 |
| 총빈도: 5684<br>유형 빈도: 206 | |

표 3.7. '-음'의 신어와 빈도

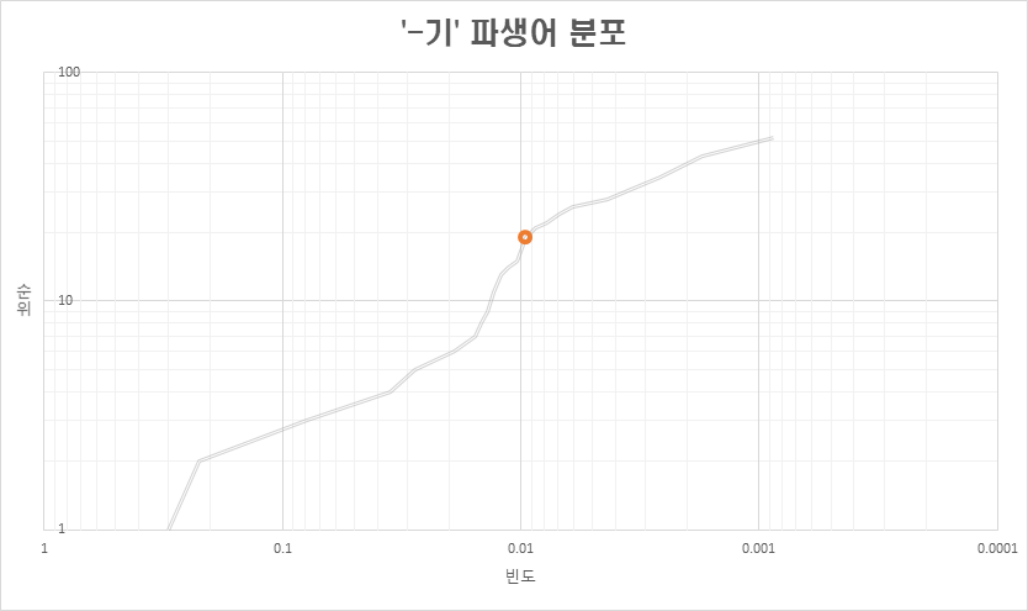
'-음'은 동사나 형용사 어간과 결합하여 명사를 파생하는 접미사이다. 순위-빈도 그래프에 표시된 세타 파생어는 '가뭄'이며 빈도 25이다. 세타 이하인 파생어의 총빈도는 999이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.001001
θ=0.0044

'-음'은 어근에 특별한 의미를 더하지 않는 경우가 많다. 그러나 '튀김, 무침'

처럼 동사 어근과 결합하면 ‘그 행위의 결과물’이라는 의미를 더하기도 한다. 또, 특별한 의미를 더하지 않는 듯한 어근에서도 단순히 명사 파생만 당하는 것이 아니라 어근의 의미에서 벗어난 파생어가 되는 경우가 매우 많다. ‘걸음, 지름, 가르침, 물음’등 많은 ‘-음’ 파생어가 여기에 해당된다. 즉, ‘걸음’은 걷는 행위 자체가 아니라 걸을 때의 동작을 의미하고, ‘지름’은 지르는 움직임이 아니라 질러 움직인 거리를 의미하며, ‘가르침, 물음’ 역시 어근 동사가 의미하는 행위가 아니라 그 동사의 내용역에 대응하는 의미가 된다. 이는 명사 파생 기능만을 가진 ‘-기’에는 없는 의미 특성으로, ‘음’ 파생어의 접사로 ‘기’를 결합해 보면 두 접사의 의미 차이가 명확하게 드러난다. ‘-음’이 파생한 신어 ‘옆트임’도 단순한 명사적 의미인 ‘트이는 과정’이 아니라 비유적 의미를 띤다.

3.2.3.3. ‘-기’



| 파생어 | 빈도 |
|---|---|
| 물빼기 | 1 |
| 대못박기 | 1 |
| 밑장빼기 | 1 |
| 총빈도: 1155<br>유형 빈도: 87<br>신어:3 | |

표 3.8. '-기'의 신어와 빈도

순위-빈도 그래프에 표시된 세타 파생어는 공동 19위인 '듣기'와 '날치기'로, 빈도 11이다. 세타 이하인 파생어의 총빈도는 189이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.01587
θ=0.00952

'-기' 역시 동사어간이나 형용사어간과 결합하여 명사를 파생한다. '-이', '-음' 파생어와 달리 코퍼스에 실제로 나타난 '-기' 파생어 대부분은 어근에 아무런 의미를 추가하지 않는 단어뿐이다.[43] 편중된 파생어 특성 때문인지 새로운 '무의미' 명사 파생은 사실상 '-기'가 전담하고 있는 듯하다. 세 건의 신어 모두 '(물을) 빼는 행위', '(대못을) 박는 행위', '(밑장을) 빼는 행위'를 의미한다.

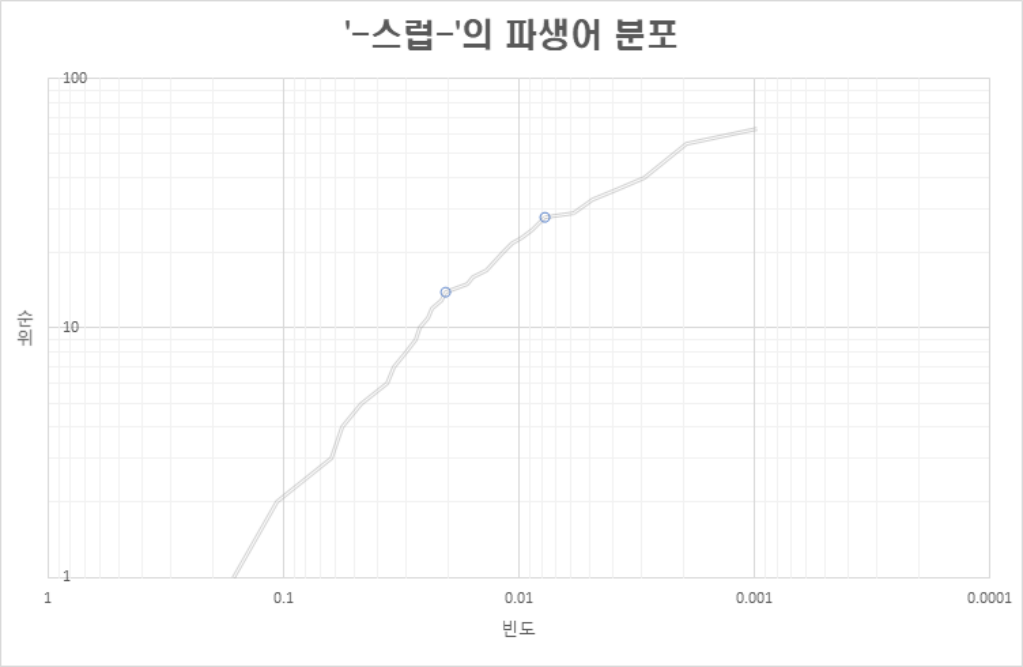
'-이, -음, -기'의 세 접사 중 '-기'의 추정 생산성이 가장 높다. 그런데 생산성 격차의 요인을 접사의 의미 차이, 즉, '-기'가 다른 두 접사와는 달리 의미를 더하지 않는다는 데에서 찾는 것은 타당하지 않다. 앞서 다루었던 '-가, -자'의 경쟁에서는 오히려 나타낼 수 있는 의미가 다양한 '-가'가 단순한 의미를 더하는 '-자'보다 훨씬 높은 추정 생산성을 보여주었기 때문이다.

'-기'의 생산성이 높은 것은 '-기'가 전담하고 있는 '의미를 더하지 않는 명사 파생'이 신어를 만들어내는 상황에서 가장 자주 필요하기 때문인 듯하다. '-이, -음, -기'가 잠재적으로 파생할 수 있는 의미영역은 서로 비슷하지만 실제로 신어를 파생할 때에는 각자의 영역을 확립하게 된 것으로 보인다.

43) 자료에 나타난 '-기' 파생어 87 항목 중 '돋보기' 하나만이 도구 명사로 파생되었다.

### 3.2.4. ‘-스럽-’, ‘-하-’, ‘-롭-’, ‘-지-’, ‘-적-’

#### 3.2.4.1. ‘-스럽-’




‘-스럽-’은 명사나 일부 어근에 붙어 형용사를 파생하는 접미사이다. ‘-스럽-’의 의미는 어근의 종류에 따라 달라지는 등의 이유로 명쾌하게 정의하기가 매우 까다롭다. 여러 논의를 종합하자면 ‘어근의 속성이 있음’ 정도로 정의할 수 있을 듯하다(김창섭 1984, 윤동원 1986, 송철의 1992). 그런데 그 의미를 크게 나누더라도 ‘-스럽-’ 파생어의 의미는 다음과 같은 두 가지로 나누어야 할 듯하다.

ㄱ. ~에 가까운 속성이 있다.
ㄴ. 그것을 대하는 주체로 하여금 ~한 상태가 되도록 하는 속성이 있다.

실제로 사용된 단어들을 살펴보면 의미의 차이를 쉽게 파악할 수 있다. 다음은 국립국어원 형태 분석 말뭉치에 등장한 '-스럽-' 파생어 전체의 목록이다. '-스럽-'의 파생어는 총 111가지가 나타났다.

| | | | | | |
|---|---|---|---|---|---|
| '대륙낭인'스럽 | 까탈스럽 | 불미스럽 | 염려스럽 | 장난스럽 | 한국스럽 |
| 가증스럽 | 낙망스럽 | 불상스럽 | 영광스럽 | 저질스럽 | 한스럽 |
| 감격스럽 | 남사스럽 | 불안스럽 | 예스럽 | 절망스럽 | 허접스럽 |
| 감동스럽 | 낭패스럽 | 비밀스럽 | 외설스럽 | 정성스럽 | 현대카드스럽 |
| 감탄스럽 | 능청스럽 | 사랑스럽 | 용맹스럽 | 조심스럽 | 혐오스럽 |
| 갑스럽 | 다행스럽 | 사치스럽 | 우려스럽 | 존경스럽 | 호들갑스럽 |
| 개탄스럽 | 당혹스럽 | 새삼스럽 | 우연스럽 | 존엄스럽 | 호사스럽 |
| 걱정스럽 | 당황스럽 | 성(聖)스럽 | 원망스럽 | 짜증스럽 | 호화스럽 |
| 게이스럽 | 만족스럽 | 소란스럽 | 위험스럽 | 챙피스럽 | 혼란스럽 |
| 경악스럽 | 맛깔스럽 | 수고스럽 | 유감스럽 | 촌스럽 | 효성스럽 |
| 고급스럽 | 멋스럽 | 수치스럽 | 유난스럽 | 충성스럽 | 후회스럽 |
| 고민스럽 | 명예스럽 | 시원시원스럽 | 유독스럽 | 치욕스럽 | |
| 고생스럽 | 밉상스럽 | 신비스럽 | 의문스럽 | 케이블스럽 | |
| 고집스럽 | 바보스럽 | 실망스럽 | 의심스럽 | 탐스럽 | |
| 고통스럽 | 변덕스럽 | 얌채스럽 | 의아스럽 | 탐욕스럽 | |
| 고풍스럽 | 보람스럽 | 양념스럽 | 익살스럽 | 통탄스럽 | |
| 곤혹스럽 | 부담스럽 | 어른스럽 | 자랑스럽 | 퇴행스럽 | |
| 공포스럽 | 분노스럽 | 억지스럽 | 자미(滋味)스럽 | 퉁명스럽 | |
| 관용스럽 | 불량스럽 | 억척스럽 | 자연스럽 | 패악스럽 | |
| 구차스럽 | 불만스럽 | 여성스럽 | 자유스럽 | 하얀스럽 | |

표 3.9.1. '-스럽-'의 파생어 목록

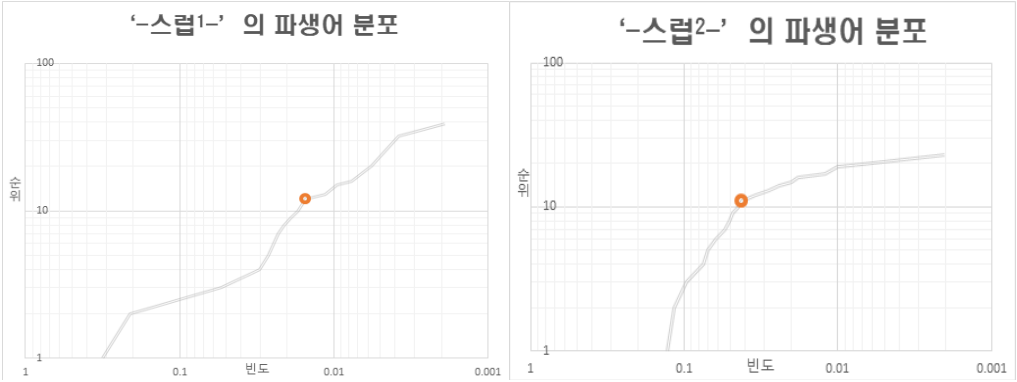
이 단어들 중 ㄱ의 의미에 가까운 것으로는 '호사스럽, 불상스럽[44], 패악스럽'등의 단어가 있고, ㄴ의 의미에 가까운 것은 '걱정스럽, 불안스럽, 의심스럽'과 같은 단어가 있다. 두 의미 모두 화자가 지각하는 어떤 대상의 속성을 나타낸다는 점에서는 동일하지만 ㄱ 부류의 어근 '호사, 불상, 패악…' 등은 순전히 그 대상의 속성이 되는 반면 ㄴ 부류의 어근 '걱정, 불안, 의심…' 등은 화자 혹은 경험주가 겪는 상태라는 점에서 중요한 차이가 있다. 단순화하자면 ㄱ부류는 '그것이 -스러운' 것이고, ㄴ부류는 '내가 (그것 때문에) -스러운' 것이다.[45] 이러한 차이 때문에 ㄴ부류의 '-스럽-' 파생어들은 그 속성과 관련

44) 신어도 단어가 사용된 문맥을 참고하여 분류하였다.
45) 물론 '-스럽-' 파생어 전체를 크게 두 부류로만 나누었으므로 이러한 구분에 잘 들어맞는 단어도 있고 그렇지 않은 단어도 있다. 가령 '수치스럽다'는 '내가 수치스럽다'로 충분하지

되는 대상을 생략하고 사용하더라도 자연스러운 경우가 많다.

이러한 의미 차이는 본고에서 택한 분포 분석 방법을 통해 확연하게 포착할 수 있다. '-스럽-' 전체의 순위-빈도 분포에는 저빈도어 부근의 일관된 패턴을 확인하기가 힘들지만 두 의미에 따라 나누어 분석할 경우 패턴의 단절점, 즉 세타가 명확하게 나타나는 것이다. 위 분포도에 작게 표시된 두 개의 미미한 단절점은 28위, 빈도 8인 '여성스럽'과 14위, 빈도 14인 '당혹스럽'이다. 이렇게 미미한 단절점은 세타로 파악하기가 부담스럽다. 그런데 빈도 분포가 이처럼 문란한 이유가 의미적으로 이질적인 두 집단이 뒤섞였기 때문일 것이라는 가설이 사실이라면 두 집단을 나누어 확인해볼 수 있다. 편의상 ㄱ의 의미를 띈 파생어를 파생하는 접사를 '-스럽[1]-', ㄴ의 의미를 띈 파생어를 파생하는 접사를 '-스럽[2]-'라고 부르기로 한다.

'-스럽[1]-', '-스럽[2]-'의 순위-빈도 분포와 파생어 목록은 다음과 같다. 신어는 목록에서 음영 처리하였다.46)




만 '의심스럽다'는 '내가 그것이 의심스럽다'와 같이 대상을 밝히지 않으면 조금 어색하거나 의미가 달라질 수 있다. 또, '자랑스럽다'와 같은 단어는 '자랑의 속성을 가지고 있음'과 같이 해석하면 ㄱ의 의미에 가깝지만 그러한 감정을 느끼는 주체에 집중하면 ㄴ으로 분류할 수도 있을 듯하다. 이것은 각 어근의 의미적 특성에 기인한 차이인 것으로 보인다. 그러나 어근이나 접사의 의미를 더 세밀하게 나누는 것은 3장의 목적에 맞지 않고, 신어나 저빈도어에서는 둘 사이의 구분이 명확한 편이므로 일단은 이 정도의 구분에서 그치기로 한다.

46) '남사스럽, 한스럽, 새삼스럽'은 분류하기가 애매하지만 ㄱ의 의미에 가까운 것으로 판단하였다. 코퍼스에 나타난 '-스럽[1]/스럽[2]-'의 신어의 빈도는 모두 1이다.

| | | | |
|---|---|---|---|
| '대륙낭인'스럽 | 불상스럽 | 자미(滋味)스럽 | 호화스럽 |
| 가증스럽 | 비밀스럽 | 자랑스럽 | 효성스럽 |
| 갑스럽 | 사랑스럽 | 자연스럽 | 총 빈 도 : 647<br>신어: 13 |
| 게이스럽 | 사치스럽 | 자유스럽 | |
| 고급스럽 | 성(聖)스럽 | 장난스럽 | |
| 고생스럽 | 소란스럽 | 저질스럽 | |
| 고집스럽 | 수고스럽 | 정성스럽 | |
| 고풍스럽 | 시원시원스럽 | 조심스럽 | |
| 관용스럽 | 신비스럽 | 존엄스럽 | |
| 구차스럽 | 얌채스럽 | 짜증스럽 | |
| 까탈스럽 | 양념스럽 | 촌스럽 | |
| 낭패스럽 | 어른스럽 | 충성스럽 | |
| 능청스럽 | 억지스럽 | 케이블스럽 | |
| 다행스럽 | 억척스럽 | 탐스럽 | |
| 맛깔스럽 | 여성스럽 | 탐욕스럽 | |
| 멋스럽 | 영광스럽 | 퇴행스럽 | |
| 명예스럽 | 예스럽 | 퉁명스럽 | |
| 밉상스럽 | 외설스럽 | 패악스럽 | |
| 바보스럽 | 용맹스럽 | 하얀스럽 | |
| 변덕스럽 | 우연스럽 | 한국스럽 | |
| 부담스럽 | 위험스럽 | 허접스럽 | |
| 불량스럽 | 유난스럽 | 현대카드스럽 | |
| 불만스럽 | 유독스럽 | 호들갑스럽 | |
| 불미스럽 | 익살스럽 | 호사스럽 | |

표 3.9.2. '-스럽[1]-'의 파생어와 신어

| | |
|---|---|
| 감격스럽 | 의심스럽 |
| 감동스럽 | 의아스럽 |
| 감탄스럽 | 절망스럽 |
| 개탄스럽 | 존경스럽 |
| 걱정스럽 | 챙피스럽 |
| 경악스럽 | 치욕스럽 |
| 고민스럽 | 통탄스럽 |
| 고통스럽 | 혐오스럽 |
| 곤혹스럽 | 혼란스럽 |
| 공포스럽 | 후회스럽 |
| 낙망스럽 | 총빈도: 372<br>신어: 2 |
| 당혹스럽 | |
| 당황스럽 | |
| 만족스럽 | |
| 보람스럽 | |
| 분노스럽 | |
| 불안스럽 | |
| 수치스럽 | |
| 실망스럽 | |
| 염려스럽 | |
| 우려스럽 | |
| 원망스럽 | |
| 유감스럽 | |
| 의문스럽 | |

표 3.9.3. '-스럽[2]-의 파생어와 신어

파생어를 의미에 따라 나누어 분석하니 빈도 분포 패턴의 단절을 확연하게 알아볼 수 있다. '−스럽[1]−'의 순위−빈도 그래프에 표시된 세타 파생어는 '여성스럽'이며, '−스럽[2]−'의 순위−빈도 그래프에 표시된 세타 파생어는 '당혹스럽'이다. '여성스럽'과 '당혹스럽'은 '−스럽−' 전체의 순위−빈도 그래프에서 약한 단절점으로 나타났던 바로 그 파생어들이다. 각자의 분포에서는 단절이 명확하지만 한 데 뒤섞으니 마치 파장이 서로를 간섭하듯 분포 패턴이 흐려져 버렸던 것이다.

이상의 결과를 통해 다음과 같은 사실을 알 수 있다. 첫째, 한 접사의 의미가 둘 이상이면 세타도 의미당 하나씩 여러 개 나타날 수 있다. 둘째, 각 의미의 파생어 집합은 각자의 개별적인 빈도 분포를 갖는, 전체 파생어 집합의 독립적인 부분 집합으로 해석할 수 있다.

다음으로는 '−스럽−'의 산출된 생산성, 저장 임계 빈도를 제시하겠다. '−스럽[1]−', '−스럽[2]−'의 의미 차이가 아주 크지는 않지만 일관되게 나타나며 분포

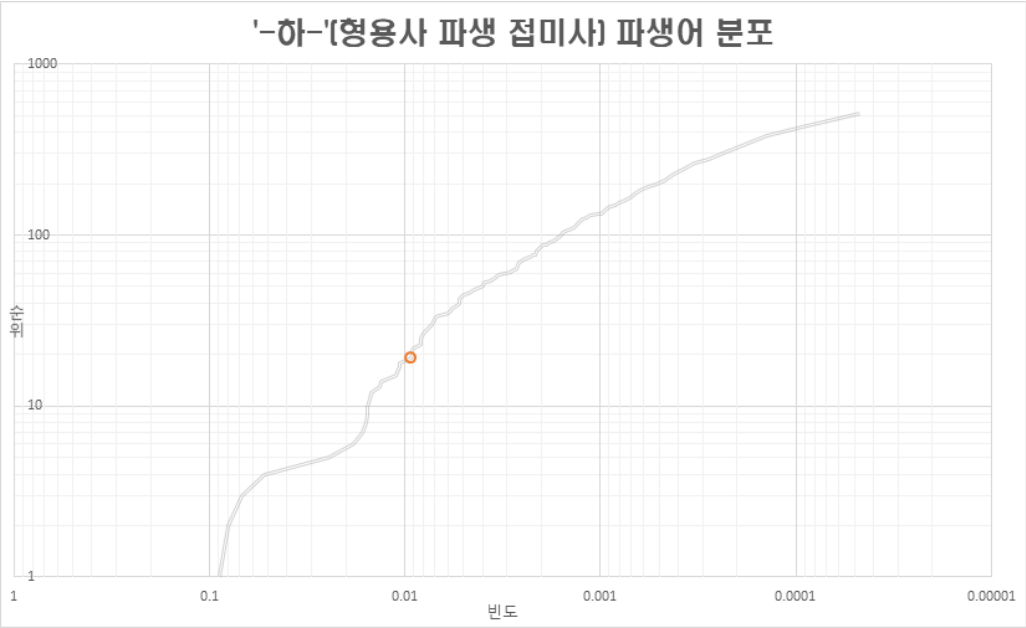
상 확연한 차이를 보여주므로 여기서는 둘을 분리하여 계산한 값도 함께 제시한다. 다만, 다른 접사들과 일관되게 처리하기 위해 이후로는 '-스럽-' 전체의 값을 사용하기로 한다.[47] '-스럽-' 전체의 생산성과 세타를 P·θ, '-스럽[1]-'과 '-스럽[2]-'의 생산성과 세타를 각각 $P_1 \cdot \theta_1$, $P_2 \cdot \theta_2$라 하자. $P_1 \cdot \theta_1$을 계산하는 기준이 된 세타 파생어는 '여성스럽'이고 빈도 8, 세타 이하인 파생어의 총빈도는 173이다. $P_2 \cdot \theta_2$를 계산하는 기준이 된 세타 파생어는 '당혹스럽'이고 빈도 21, 세타 이하인 파생어의 총빈도는 351이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

$P=0.01461$, $\theta=0.0079$
$P_1=0.104$, $\theta_1=0.01524$
$P_2=0.01626$, $\theta_2=0.04251$

한편, '-스럽-'은 어근에 따라 '-답-'[48], '-하-', '-롭-', '-지-', '-적(的)[49]-' 등과 유사한 의미를 가지며, 따라서 많은 어근과의 결합에서 경쟁 관계에 있다. '-스럽-'이 포함된 저지쌍 내의 저지현상을 다룰 때에는 앞서 다루었던 의미·분포 특성을 고려해야 할 것이다. 저지는 동의어 간에 발생하기 때문이다. '-스럽-'의 저지에 대해서는 4장에서 다룰 것이다. 다음으로는 '-스럽-'과 경쟁하는 접사를 차례로 살펴본다.

#### 3.2.4.2. '-하-(형용사 파생 접미사)'

47) '-스럽[1]-'과 '-스럽[2]-'의 의미 차이는 4장에서 접사 간 경쟁과 저지현상의 분석에 활용할 것이다.
48) 이 때의 '-답-'은 두 가지의 서로 다른 '-답-' 중 '~의 속성을 풍부히 지님'의 뜻을 가진 것이다. 3.1.6 참고.
49) '-적(的)-'은 '-스럽-, -롭-, -하-, -지-' 등과 직접 대치되지는 못하고 서술격 조사 '-이-'와 결합한 뒤에야 경쟁 관계에 놓이게 된다. 그러나 본 연구에서는 이와 같은 차이는 크지 않은 것으로 보고 접사 간 경쟁이나 생산성, 저지를 논의할 때 함께 다루었다.




| 파생어 | 빈도 | 파생어 | 빈도 | 파생어 | 빈도 |
|---|---|---|---|---|---|
| 감성감성하 | 1 | 소녀소녀하 | 1 | 프랙티컬하 | 1 |
| 그루브하 | 1 | 스테레오티피컬하 | 1 | 플랜티테이하 | 1 |
| 꺼묵꺼묵하 | 1 | 슴덩슴덩하 | 1 | 하드하 | 1 |
| 꼬롬꼬롬하 | 1 | 시스테믹하 | 1 | 후덜덜하 | 1 |
| 넙작하 | 1 | 의리의리하 | 1 | 진덤진덤하 | 1 |
| 네셔널하 | 1 | 익스트림하 | 1 | 샤이하 | 2 |
| 다기능하 | 1 | 잘박잘박하 | 1 | 심볼릭하 | 2 |
| 따들따들하 | 1 | 쥬시하 | 1 | 크리미하 | 2 |
| 뜨끈뜨끈하 | 1 | 진덤하 | 1 | 마이너하 | 3 |
| 러블리하 | 1 | 컨서버티브하 | 1 | 총빈도: 21234<br>신어:54 | |
| 로얄하 | 1 | 코어하 | 1 | | |
| 리액티브하 | 1 | 크리스피하 | 1 | | |
| 몽글몽글몽글하 | 1 | 크리티컬하 | 1 | | |
| 민질민질하 | 1 | 탱글하 | 1 | | |
| 반담하 | 1 | 통글통글하 | 1 | | |
| 빠듯빠듯하 | 1 | 퍼들퍼들하 | 1 | | |
| 빠짝하 | 1 | 페미닌하 | 1 | | |
| 뿔뚝뿔뚝하 | 1 | 포슬포슬하 | 1 | | |
| 살큰살큰하 | 1 | 푸시시하 | 1 | | |
| 설컹설컹하 | 1 | 프라그마틱하 | 1 | | |

표 3.10. '-하-'의 신어와 빈도

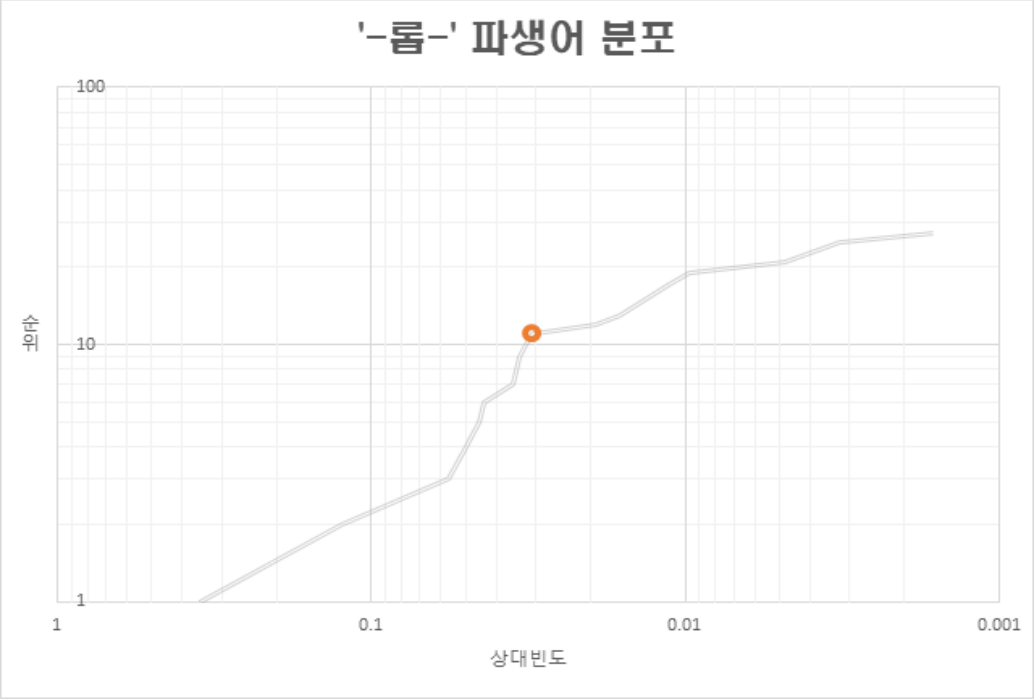
형용사를 파생하는 '-하-'는 의성·의태어 부사나 상태성 명사 등 다양한 대

상과 결합하며 별다른 의미를 더하지 않는 접미사이다. 외국어의 형용사는 '-하-' 파생을 거친 뒤에야 한국어에서 형용사로 사용할 수 있다. 이러한 파생법은 신어를 생산하는 데 매우 활발하게 적용되고 있다.

순위-빈도 그래프에 표시된 세타 파생어는 '불안하'이고 빈도 198이다. 세타 이하인 파생어의 총빈도는 10729이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.00503
θ=0.00932

3.2.4.3. '-롭-'




| 파생어 | 빈도 |
|---|---|

| | |
|---|---|
| 간교롭 | 1 |
| 총빈도: 620 | |

표 3.11.1. '-롭-'의 신어와 빈도

'-롭-'은 모음으로 끝나는 명사나 명사성 어근과 결합하여 '~의 속성이 있음'의 뜻을 더하고 형용사를 파생하는 접미사이다. 순위-빈도 그래프에 표시된 세타 파생어는 '지혜롭'이고 빈도 19이다. 세타 이하인 파생어의 총빈도는 119이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.0084
θ=0.03065

'-롭-'은 전통적으로 생산성이 낮은 접사로 분류되어 왔고(송철의 1992), 계량적 연구에서도 생산성이 가장 낮은 축에 속하는 것으로 조사된 바 있다(차준경 1995, 이광호 2009). 게다가 유형 빈도 35, 총빈도 620으로 세력도 상당히 미약한 편이다. 고빈도 구간은 다음과 같이 '-스럽-, -적(的)' 등의 경쟁 접사를 완전히 저지하거나 양적으로 압도하는 파생어들로만 이루어져 있다.[50]

| 파생어 | 빈도 |
|---|---|
| 이롭 | 20 |
| 조화롭 | 21 |
| 여유롭 | 22 |
| 해롭 | 22 |
| 정의롭 | 27 |
| 풍요롭 | 28 |
| 순조롭 | 31 |
| 평화롭 | 35 |
| 흥미롭 | 77 |
| 자유롭 | 218 |

표 3.11.2. 빈도가 저장 임계점을 넘은 '-롭-' 파생어

이렇게 거의 배타적으로 결합하는 소수 어근의 파생어들이 파생어 집합 전

50) 접사들 간의 구체적인 경쟁 양상은 4장에서 다룬다.

체를 과점하는 것은 신어 생산 경쟁에서 밀리고 있는 접사들의 전형적인 분포 형태이다. 다시 말해 '-롭-'은 고빈도로 결합하는 일부의 어근과만 한 덩어리로 나타나는 경향이 강하며, 접사의 기능은 잃어가고 있는 것으로 보인다.

3.2.4.4. '-지-'

| 파생어 | 빈도 |
|---|---|
| 고급지 | 1 |
| 총빈도: 207 | |

표 3.12.1. '-지-'의 신어와 빈도

'-지-'는 주로 명사를 어근으로 하여 '그런 속성이 있음'의 뜻을 더하고 형용사를 파생하는 접미사이다. 사용되는 빈도가 높지 않은데도 신어가 포착되어 생산성은 비교적 높게 산출되었다. 그러나 포착된 신어 '고급지'는 최근 들어 유행어처럼 사용된 단어인데다 유형 빈도가 20 미만으로 매우 낮아 값이

왜곡되었을 것이므로 실제 생산성이 이처럼 높다고 보기는 어렵다. 순위-빈도 그래프에 표시한 세타 파생어는 '건방지'이고 빈도 7이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

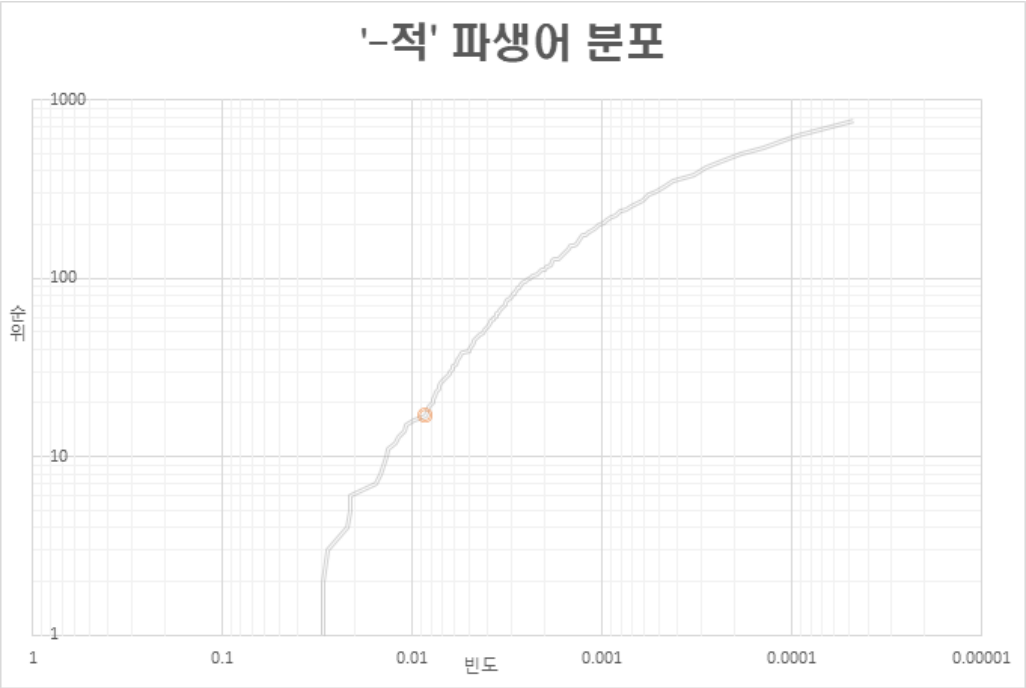
P=0.02083
θ=0.03382

'-지-' 역시 '-롭-'과 마찬가지로 극소수의 고빈도 파생어가 파생어 집합의 대부분을 차지한다. 최고순위 파생어인 '멋지'는 전체 파생어의 60퍼센트에 달하는 극단적으로 높은 비율을 차지하고 있다. 이러한 분포를 통해 '-지-' 역시 '-롭-'과 마찬가지로 접사 간 경쟁에서 밀려나고 있음을 알 수 있다. 다음 표에서 고빈도 구간의 분포를 확인할 수 있다.

| 파생어 | 빈도 |
|---|---|
| 차지 | 8 |
| 야무지 | 9 |
| 기름지 | 10 |
| 값지 | 17 |
| 멋지 | 123 |

표 3.12.2. 저장 임계점을 넘은 '-지-' 파생어와 그 빈도

### 3.2.4.5. '-적(的)'




| 파생어 | 빈도 | 파생어 | 빈도 | 파생어 | 빈도 | 파생어 | 빈도 |
|---|---|---|---|---|---|---|---|
| 개간적 | 1 | 불량적 | 1 | 영합주의적 | 1 | 트로트적 | 1 |
| 비공정적 | 1 | 브랜드적 | 1 | 예전적 | 1 | 패러독스적 | 1 |
| 구단적 | 1 | 비방적 | 1 | 오기적 | 1 | 포르노적 | 1 |
| 국외부적 | 1 | 비상시적 | 1 | 오리엔탈리즘적 | 1 | 포퓰리스트적 | 1 |
| 권리적 | 1 | 소급사관적 | 1 | 오타쿠적 | 1 | 학사적 | 1 |
| 극예술적 | 1 | 사기극적 | 1 | 원칙법적 | 1 | 한식적 | 1 |
| 남미적 | 1 | 사범적 | 1 | 유언비어적 | 1 | 혼(魂)적 | 1 |
| 뉴스적 | 1 | 사이코패스적 | 1 | 일정적 | 1 | 뉴라이트적 | 1 |
| 단테적 | 1 | 생계적 | 1 | 작정적 | 1 | 화해주의적 | 1 |
| 대합적 | 1 | 생협적 | 1 | 장송적 | 1 | 행정만능주의적 | 1 |
| 독식적 | 1 | 서시적 | 1 | 적대국적 | 1 | 주위적 | 1 |
| 뒷감당적 | 1 | 서예사적 | 1 | 전문직적 | 1 | 공영적 | 2 |
| 록(rock)적 | 1 | 소명적 | 1 | 평가절하적 | 1 | 성희롱적 | 2 |
| 막사적 | 1 | 수법적 | 1 | 정반합적 | 1 | 열외적 | 2 |
| 메타텍스트적 | 1 | 심증적 | 1 | 제로썸적 | 1 | 외교관적 | 2 |
| 풍기문란적 | 1 | 아일랜드적 | 1 | 진영적 | 1 | 헤라클레이토스적 | 2 |
| 방목적 | 1 | 야기적 | 1 | 차베스적 | 1 | 총빈도: 21174 | |
| 법가적 | 1 | 야상적 | 1 | 총력적 | 1 | 신어:88 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 법고창신적 | 1 | 어드미니스트레이션적 | 1 | 컬쳐적 | 1 | |
| 법사회학적 | 1 | 엔엘적 | 1 | 콘텐츠적 | 1 | |
| 변형주의적 | 1 | 엔엘피적 | 1 | 클리셰적 | 1 | |
| 보람적 | 1 | 엔지니어링적 | 1 | 통섭적 | 1 | |

표 3.13. '-적(的)'의 신어와 빈도

'－적(的)'은 주로 명사에 붙어 '그런 성격을 띠는'의 의미를 더하고 명사를 파생하는 한자어 접미사이다. '－적' 파생어는 다음과 같은 이유에서 명사로 분류해야 한다(김창섭 1996:186).

ㄱ. 어미나 조사 없이 단독형으로 쓰일 수 있다.
ㄴ. 계사 '－이－'를 취할 수 있다.
ㄷ. 부정문에서 'X的(－이－) 아니다'의 형식을 취한다.
ㄹ. 구격 조사 '－으로'를 취할 수 있다.

김창섭(1996)에서는 이러한 특성들이 오히려 '－적'의 형용사적 기능을 드러내는 것으로 해석할 수 있다고 하였는데, 이는 타당한 듯하다. 우선, '－적' 파생어는 일반적인 명사에 비해 취할 수 있는 조사에 제약이 많다. '적' 파생어는 격조사 '이', '의', '과', '을', '에' 등과의 결합이 불가능하다(노명희 2005:189). 다음으로, 실제 쓰임에 있어서도 '－적' 파생어는 '－이－'와 결합한 'X적이－' 형태로 형용사로 쓰이는 경우가 많다.[51]

이처럼 '－적'은 의미와 기능 면에서 형용사와 비슷한 점이 많아 다른 형용사 파생 접사와 자주 비교된다. 김규철(1981), 김창섭(1984), 김창섭(1996) 등에서 '－적'이 '－스럽－, －답－, －되/롭－, －하－' 등과 배타적으로 분포한다고 지적한 바 있으며, 이광호(2009)에서는 '－적', '－스럽－', '－롭－'의 저지현상을 다루기도 했다. 전술하였듯 본 논의에서는 '－적－, －스럽－, －답－, －롭－, －하－' 사이의 경쟁에 대해 4장에서 논의할 것이다.

다음은 '－적'의 생산성과 저장 임계 빈도를 살펴본다. 순위－빈도 그래프에 표시된 세타 파생어는 '근본적'이고 빈도 179이다. 세타 이하인 파생어의 총빈

51) 코퍼스에서 '-적' 파생어는 총 21174회 출현하였는데, 이 중 'X적이' 형태로 나타난 것은 7640회이다.

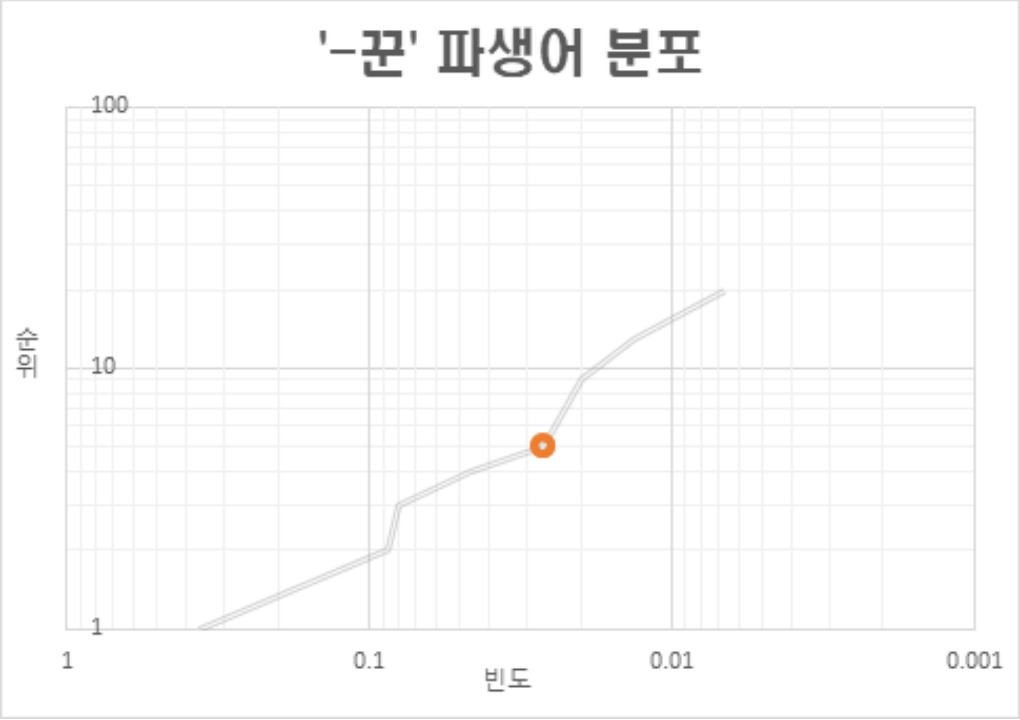
도는 15266이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.00576
θ=0.00845

매우 생산적이라는 기존의 직관적인 평가와는 달리 그다지 생산적인 편이 아니다. 세타와 비교한 상대적 생산성도 매우 낮다.

### 3.2.5. ‘-꾼’, ‘-쟁이’

#### 3.2.5.1. ‘-꾼’



| 파생어 | 빈도 |
|---|---|
| 방구꾼 | 1 |
| 로비꾼 | 1 |
| 달리기꾼 | 1 |
| 유혹꾼 | 1 |
| 명품꾼 | 1 |
| 총빈도: 150<br>신어:5 | |

표 3.14. '-꾼'의 신어와 빈도

'-꾼'은 명사에 붙어 '그것을 (버릇처럼·직업적으로) 하는 사람', '그것이나 그것의 특성을 많이 가진 사람'의 의미를 더하고 명사를 파생하는 접미사이다. 빈도가 낮지만 신어 형성 능력이 매우 뛰어나다. 산출된 생산성과 저장 임계 빈도를 먼저 제시하면 다음과 같다. 세타 파생어는 공동 5위로 '파수꾼, 시위꾼, 심부름꾼, 밀수꾼'이며 빈도 4이다. 세타 이하인 파생어의 총빈도는 63이다.

P=0.07937
θ=0.02667

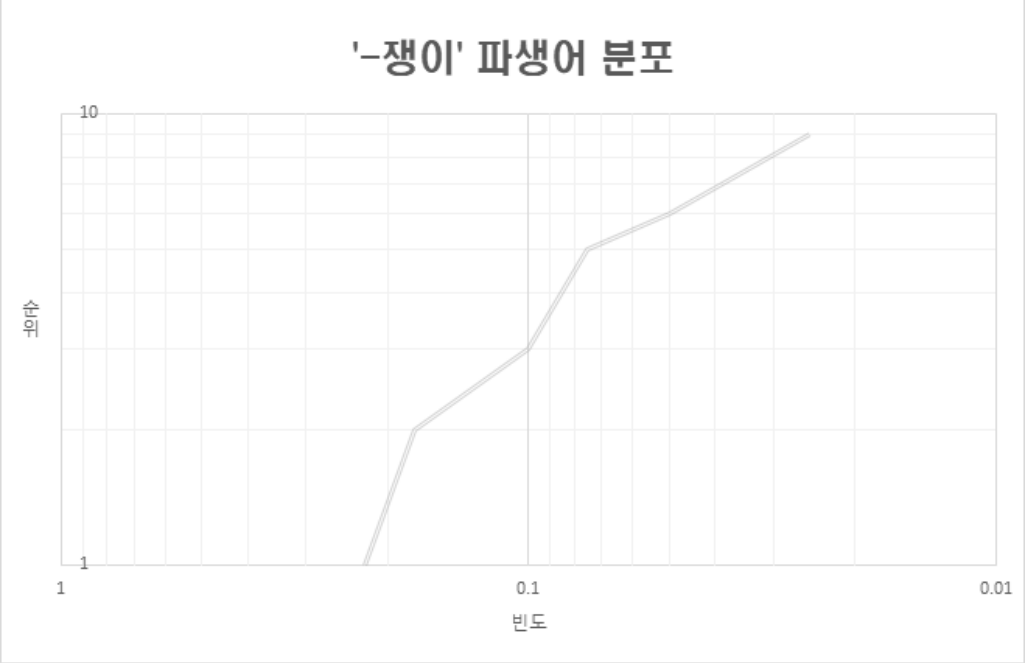
0.08에 달하는 P는 본 연구에서 조사한 접사 중에서도 높은 편이고, 이광호(2009) 당시의 수치에 비해서도 매우 높아진 것이다.[52] '-장이, 쟁이-'와 같이 의미가 비슷한 다른 접미사들이 통시적으로 빠르게 쇠퇴하였음을 고려하면 '-꾼'의 이와 같은 세력 강화는 '~하는 사람'의 의미영역을 점차 흡수해 온 결과가 아닐까 한다. 즉, 원래는 '-꾼'에게도 경쟁 접사가 있었지만 그들의 세력이 점차 약화되면서 이 의미를 독점적으로 표현하는 접사가 된 것이다. 공시적으로 생산적인 경쟁 접사가 없는 '-꾼'을 본 연구의 조사 대상에 포함한 이유는 이처럼 공시적 분포가 통시적인 경쟁 과정을 함축하고 있기 때문이다.

한편, '-꾼'은 의미적으로 앞서 다루었던 '-가(家), -자(者)'와 거의 같지만 경쟁 관계에 있다고 볼 수 없다. 결합 가능한 어근의 범위에서 차이가 나기 때문이다. '-꾼'은 '-가, -자'와 달리 고유어 및 외래어 어근과 결합할 수 있

52) 이광호(2009)에서 계산한 P는 0.01278이다. 지금의 P는 이광호(2009)의 방식에 따라 계산하더라도 0.05263에 달하므로 4배 이상 증가하였다.

다. 이런 차이 때문인지 '-가, -자'는 한자어 어근의 파생만을 담당하고 '-꾼'은 고유어, 외래어 및 '-가, -자'가 결합하지 않는 한자어 어근만을 파생하는 것으로 나타난다.[53] 따라서 또, '-꾼'과 '-가, -자'는 엄밀히는 의미에서도 차이가 있다. '-꾼'은 부정적 의미가 있는 어근과 결합하거나 그러한 의미가 없는 어근에 부정적인 의미를 더할 수 있지만 '-가, -자'는 그러한 의미가 없다.

3.2.5.2. '-쟁이'




| 파생어 | 빈도 |
|---|---|
| 사치쟁이 | 1 |

53) 간혹 '약탈꾼-약탈자'와 같이 결합 가능한 어근이 겹치는 경우도 있지만 파생어의 의미가 다르다.

| 배구쟁이 | 1 |
|---|---|
| 총빈도: 40<br>유형 빈도: 15<br>신어:2<br>P=0.05 | |

표 3.15,1. '-쟁이'의 신어와 빈도, 생산성

'－쟁이'는 '－꾼'과 마찬가지로 일부 명사 뒤에 붙어 '어근과 관련된 속성이나 기술을 지닌 사람'의 의미와 때에 따라 낮잡아 보는 부정적 의미를 더하고 명사를 파생하는 접미사이다. 여전히 신어 생산 능력은 어느 정도 유지하고 있다. 그러나 분포 패턴이 뚜렷하지 않고 유형 빈도가 너무 낮아서 세타를 계산하기 어렵다. 표 3.15.1의 P는 파생어 전체에 대한 신어의 비율로 계산한 것이다.

'－쟁이'는 기존에도 세력이 약한 편이었으나 시간이 흐름에 따라 표 3.15.2, 3.15.3과 같이 꾸준히 쇠퇴하고 있다. 통시적인 변화와 공시적 상태를 모두 고려하면 '－쟁이'는 사실상 사멸 단계에 접어든 것으로 보인다. '－쟁이'는 원래 자신과 결합했거나 충분히 새로 결합할 수 있는 '방구, 명품, 로비'와 같은 어근의 파생을 전부 '－꾼'에게 양보하고 있다. 기존의 고빈도 파생어 역시 배타적 저지력을 유지하고는 있으나 코퍼스의 규모나 구성의 차이를 감안하더라도 다음과 같이 전체적으로 훨씬 덜 쓰이는 추세이다.[54]

| | 이광호(2009) | 본 연구 |
|---|---|---|
| 유형 빈도 | 41 | 16 |
| 총빈도 | 201 | 53 |
| 1위 빈도 | 21 | 13 |

표 3.15.2. '-쟁이'의 통시적 변화

| 파생어 | 이광호(09) | 본 연구 |
|---|---|---|
| 영감쟁이 | 5 | 0 |
| 오입쟁이 | 5 | 0 |
| 글쟁이 | 5 | 2 |
| 고집쟁이 | 6 | 0 |
| 중매쟁이 | 7 | 0 |
| 수다쟁이 | 8 | 0 |
| 흉내쟁이 | 9 | 0 |
| 양복쟁이 | 11 | 0 |
| 멋쟁이 | 12 | 2 |
| 거짓말쟁이 | 13 | 13 |

54) 이광호(2009)에서 사용한 코퍼스는 문어 440만 어절, 본 연구에서 사용한 코퍼스는 문어와 구어 혼합 300만 어절(문어 200만 어절, 구어 100만 어절) 규모이다.

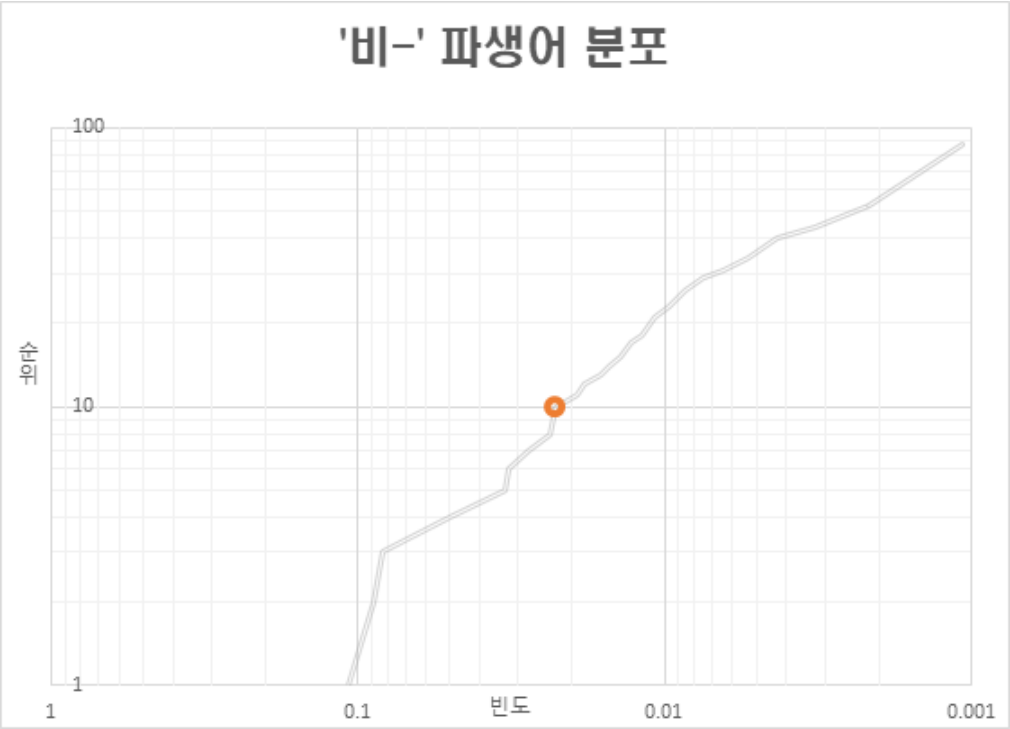
| | | |
|---|---|---|
| 빚쟁이 | 15 | 2 |
| 겁쟁이 | 15 | 6 |
| 점쟁이 | 21 | 4 |
| 월급쟁이 | 21 | 12 |

표 3.15.3. '-쟁이' 고빈도 파생어의 통시적 변화

'-쟁이'는 세력이 빠르게 쇠퇴하면서 생산성에서의 우위도 '-꾼'에게 넘겨준 것으로 보인다. 차준경(1995), 이광호(2009)의 조사에서는 '-쟁이, -장이'의 생산성이 압도적으로 높았다. 과거의 연구에 따르면 세 접사의 생산성은 '-쟁이(0.03)' > '-장이(0.021)' > '-꾼(0.0128)'(이광호 2009)의 순서였지만 본 연구의 조사 결과 '-꾼(0.079)' > '-쟁이(0.05)' > '-장이(0)'로 바뀌었다. 특히 '-장이'는 약 15년 전에 구축된 코퍼스에서는 신어를 생산하고 49회나 출현하였던 것에 비해 본 연구의 조사 대상 코퍼스에서는 단 1회만 나타나면서 급격한 쇠퇴 단계에 접어든 것으로 보인다.[55]

### 3.2.6. '비(非)-'

55) 본 연구의 조사 결과에서 '-꾼'과 '-쟁이'의 생산성은 크게 차이 나지는 않으므로 '-꾼'과 '-쟁이'의 순위가 완전히 역전되었다고 단언하기는 어려울 것이다. 선행 연구와 본 연구는 생산성 산출 방식뿐 아니라 코퍼스의 구성에서도 큰 차이가 있기 때문이다. 다만 두 접사가 변화하는 방향만큼은 비교적 극적으로 엇갈렸다고 봐도 무리가 없을 듯하다.




| 파생어 | 빈도 | 파생어 | 빈도 |
|---|---|---|---|
| 비개착 | 1 | 비철강 | 1 |
| 비경연 | 1 | 비판사 | 1 |
| 비고용직 | 1 | 비합성어 | 1 |
| 비공영 | 1 | 비허가 | 1 |
| 비교육인 | 1 | 비헌법 | 1 |
| 비동구권 | 1 | 비호지킨 | 1 |
| 비박지원 | 1 | 비문재인 | 1 |
| 비범죄 | 1 | 비무슬림 | 1 |
| 비사진적 | 1 | 비식별 | 1 |
| 비수정주의 | 1 | 비거래 | 1 |
| 비실기 | 1 | 비입원 | 1 |
| 비아프리카 | 1 | 비교육인 | 1 |
| 비안철수 | 1 | 비검찰 | 2 |
| 비외교관 | 1 | 비고시 | 2 |
| 비원자력 | 1 | 비둔산 | 2 |
| 비유전자 | 1 | 비전기식 | 2 |
| 비임상 | 1 | 비박근혜 | 2 |
| 비자민당 | 1 | 비보장 | 2 |
| 비재무 | 1 | 비공학 | 3 |

| | | |
|---|---|---|
| 비정책 | 1 | 총빈도: 933<br>신어:47 |

표 3.16. '비-'의 신어와 빈도

'비(非)-'는 명사와 결합하여 대체로 '~이 아님'의 뜻을 더하는 한자어 접두사이다. '무(無)-, 반(反)-, 미(未)-, 몰(沒)-' 등의 부정 접두 한자어와는 의미가 다르다. '불(不)-'과는 의미나 결합 대상의 차이가 거의 없는 듯하며 그렇게 알려져 있다. 그러나 파생 전후의 의미를 면밀히 비교해 보면 두 접사 사이의 체계적인 의미 차이를 발견할 수 있다. '비-'는 대부분의 파생어에서 '~이 아님'이라는 단순 부정의 의미만을 추가한다. 반면, '불-'은 파생하면 정도 반의어(gradable antonym)가 되는 어근과 결합하는 경우가 많다. 예컨대 '불이익'과 같은 파생어는 '이익이 아닌 것'이 아니라 '손해'를 의미한다.

그런데 어근에 따라 단순 부정이 곧 상보 반의어(complementary antonym)를 파생하는 경우에는 두 접사의 파생어 사이에 의미 차이가 없다. 가령, '완전'과 같은 단어는 의미적으로 완전하거나 완전하지 않은 두 가지 경우만 존재하기 때문에 단순 부정만으로도 상보 반의어를 만들게 된다. 이러한 어근에서는 두 접사의 의미적 차이가 중화된다고 할 수 있다(Plag 2003:82). 접사의 기본적인 의미가 다르더라도 의미적 차이가 중화된 경우에는 경쟁이 일어날 수 있다. 이에 관한 내용은 4장에서 다룬다.

순위-빈도 그래프에 표시된 세타 파생어는 '비과세'이고 빈도 21이다. 세타 이하인 파생어의 총빈도는 474이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.09916
θ=0.02251

### 3.2.7. 경쟁하지 않는 접사

3.2.7에서는 공시적으로 경쟁하는 접사는 없지만 분석할 만한 가치가 있다고 생각되는 몇몇 접사들의 정보를 제시한다. 분석에 포함한 것은 생산성이 아주

높을 것으로 예상되거나 3.1의 접사들과 경쟁하는 것으로 볼만한 접사들, 대표적인 한자어 접미사들이다. 의미가 유사하여 경쟁하는 접사보다는 독립적인 의미영역을 가진 접사들의 수가 훨씬 많으므로 경쟁하지 않는 접사를 분석 대상에 포함해야만 한국어 접사의 전체적인 분포 특성을 파악할 수 있다. 경쟁하지 않는 접사들의 지표는 유형 저지의 양상을 파악할 때 저지 접사와 피저지 접사를 구분하는 시금석과 같은 역할을 할 수 있으므로 본 연구에서는 특별히 중요하다.

3.2.7.1. '-답-'

| 파생어 | 빈도 | 파생어 | 빈도 |
|---|---|---|---|
| SK답 | 1 | 수도승답 | 1 |
| 총장답 | 1 | 상남자답 | 1 |
| 중독자답 | 1 | 진희님답 | 1 |
| 산업답 | 1 | 수상자답 | 1 |

| | | | |
|---|---|---|---|
| 리더답 | 1 | 김지영답 | 1 |
| 의원님답 | 1 | 막걸리답 | 1 |
| 트럼프답 | 1 | 종교단체답 | 1 |
| 사립답 | 1 | 전북답 | 1 |
| 외고답 | 1 | 기업인답 | 1 |
| 지하철답 | 1 | 현악기답 | 1 |
| 콘서트답 | 1 | 건반악기답 | 1 |
| 회담답 | 1 | 성악가답 | 1 |
| 국민차답 | 1 | 방문답 | 1 |
| 폴크스바겐답 | 1 | 이력답 | 1 |
| 희극답 | 1 | 나미답 | 1 |
| 승부답 | 1 | 매출답 | 1 |
| 지상파답 | 1 | 공단답 | 1 |
| 도우미답 | 1 | 격식답 | 1 |
| 관광특구답 | 1 | 황제답 | 1 |
| 화제작답 | 1 | 총빈도: 128 | |
| | | 신어: 38 | |

표 3.17. '-답-'의 신어와 빈도

'−답−'은 한국어의 대표적인 형용사 파생 접미사이다. 그런데 '~의 속성이 풍부함'의 의미를 갖는 '−답[1]−'과 '~의 자격을 갖추고 있음'을 의미하는 '−답[2]−'은 혼동될 수 있으나 형태만 같고 전혀 다른 접사이다. '−답[1]−'은 기원적으로 '−롭−', '−되−'와 함께 동일한 형태소의 이형태였다가 통시적 변화의 결과 분화된 접사로, 현대 한국어에서는 세력이 급속히 약해져 완전히 사멸하는 단계에 있다.[56] 이런 접사는 빈도 분석이 불가능하다. 그러므로 본 논의에서는 두 접사 중 '−답[2]−'만을 다루고 '−답−'으로 칭하기로 한다.

'−답−'은 이광호(2009)의 조사에서 매우 높은 생산성을 보였는데, 이러한 생산성을 아직까지 유지하고 있는 것으로 보인다. 본 연구의 조사 대상 접사 중 생산성이 가장 높았다. 순위−빈도 그래프에 표시된 세타 파생어는 '나라답'으로, 빈도 4이다. 세타 이하인 파생어의 총빈도는 98이다. 산출된 생산성

56) '-답[1]-'은 이광호(2009)의 조사 결과에서 이미 다음과 같이 극히 제한된 파생어 계열을 보여준다.

| 파생어 | 시답(實-) | 참답 | 꽃답 | 정답 | 아름답 |
|---|---|---|---|---|---|
| 빈도 | 51 | 14 | 14 | 29 | 1474 |

본 연구의 조사 결과 빈도가 전체적으로 더욱 감소한 것을 관찰할 수 있었다.

| 파생어 | 참답 | 정답 | 꽃답 | 아름답 |
|---|---|---|---|---|
| 빈도 | 2 | 3 | 4 | 270 |

'-답[1]-'은 오래 전부터 이미 접사로서의 단어 파생 능력을 완전히 잃었고, 그 파생어들은 전부 화석화되었으며 어휘부 내의 저장 강도가 계속해서 약해지고 있는 것으로 보인다.

과 저장 임계 빈도는 다음과 같다.

P=0.38776
θ=0.03125

‘-답-’이 높은 생산성을 보이는 이유로 결합 가능한 어근의 범위가 넓다는 점을 생각해볼 수 있다. ‘-답-’은 자립성이 있는 거의 모든 명사와 결합할 수 있어 결합 가능한 어근의 범위가 넓은 편이기 때문이다.

그러나 ‘-답-’의 생산성이 압도적으로 높은 데 비해 결합은 다른 접사들보다 월등히 자유롭다고 보기 어렵다. 실제 사용에서는 어근과의 결합에 제약이 오히려 큰 편이다. 어근이 ‘행복, 죽음, 운동’ 등의 추상명사일 때보다 ‘군인, 선생, 교수’와 같은 구체명사일 때 결합이 자연스럽기 때문이다(송철의 1992:215). 표 3.17의 신어 목록에서도 이를 확인할 수 있다. 대부분의 신어는 구체명사 혹은 고유명사를 어근으로 취하고 있으며 이는 고빈도어에서도 마찬가지이다.

‘-답-’의 높은 생산성의 원인은 의미를 일부 이상 공유하는 생산적인 경쟁 접사가 없다는 데 있는 듯하다. ‘-답$^{2}$-’의 의미와 같거나 유사한 접사는 현대국어에서 쓰이고 있지 않다.[57] 어떤 의미를 표현할 방법이 다양하다면 신어가 출력되는 경로도 분산되므로 생산성이 낮아진다. 이것은 그 의미가 현실에서 아무리 자주 필요하더라도 마찬가지다. 3.1.1에서 3.1.5에 걸쳐 다룬 접사들의 의미는 ‘명사 파생’, ‘그러한 성질이 있음’과 같이 언어 생활에서 상당히 흔하게 사용되는데도[58] 생산성은 대체로 낮은 편이다.

#### 3.2.7.2. ‘-끼리’

57) ‘-스럽-’이나 ‘-롭-’과 자주 비교되는 ‘-답-’은 ‘-답$^{1}$-’이다.
58) 같은 의미끼리 묶으면 코퍼스에서 전부 10000회 이상 출현하였는데, 코퍼스의 전체 규모가 300만 어절임을 감안하면 아주 높은 빈도이다.

| 빈도 | 파생어 | 빈도 | 파생어 |
|---|---|---|---|
| 감독끼리 | 1 | 정파끼리 | 1 |
| 고기끼리 | 1 | 종중원끼리 | 1 |
| 교섭단체끼리 | 1 | 주전끼리 | 1 |
| 그룹끼리 | 1 | 중상위권끼리 | 1 |
| 노숙인끼리 | 1 | 중앙위원끼리 | 1 |
| 본교끼리 | 1 | 중진끼리 | 1 |
| 수감자끼리 | 1 | 최고경영진끼리 | 1 |
| 영부인끼리 | 1 | 프로끼리 | 1 |
| 우리식끼리 | 1 | 화가끼리 | 1 |
| 원내대표끼리 | 1 | 너네끼리 | 3 |
| 부대표끼리 | 1 | 총빈도: 114<br>신어: 24 | |
| 이종사촌끼리 | 1 | | |
| 이통사끼리 | 1 | | |
| 전끼리 | 1 | | |

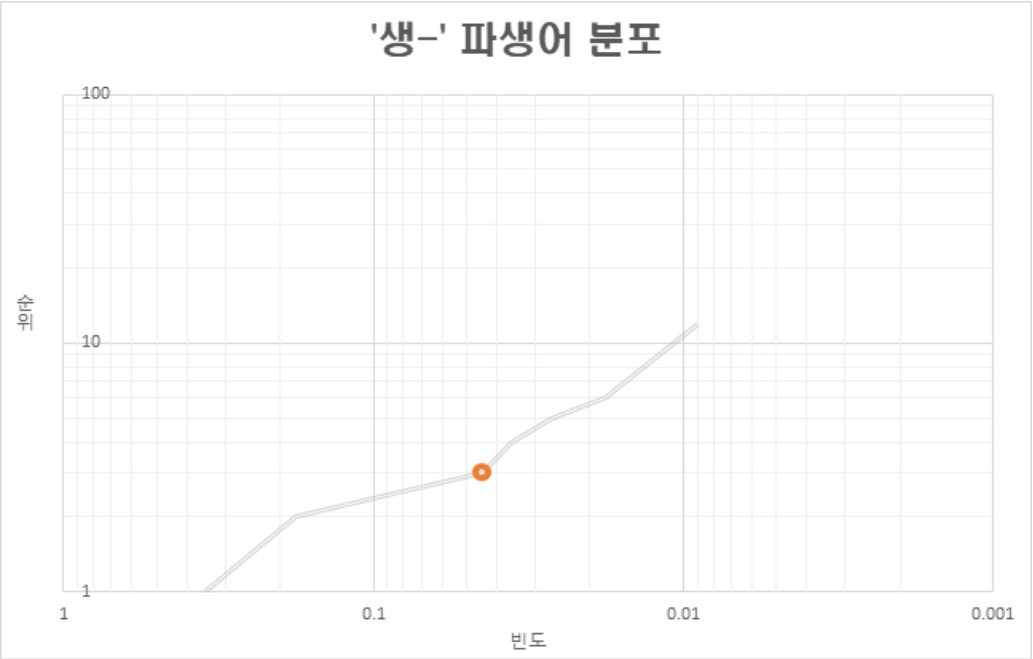
표 3.18. '-끼리'의 신어와 빈도

'-끼리'는 복수성을 가진 대다수 명사나 명사구에 결합하여 '그 부류만이 서로 함께'의 의미를 더하고 부사를 파생하는 접미사이다. 생산성이 높을 것으로 예상되는 접사를 비교 집단에 포함시키기 위해 조사하였다.[59] 파생 순위-빈

도 그래프에 표시된 세타 파생어는 '가족끼리'이며 빈도 18이다. 세타 이하인 파생어의 총빈도는 74이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.32432
θ=0.1579

3.2.7.3. '생(生)-'




| 파생어 | 빈도 |
|---|---|
| 생밀가루 | 1 |
| 생목소리 | 1 |
| 생꽁치 | 1 |
| 생수제비 | 1 |

59) '-끼리'는 NP에도 결합하기 때문에 접사로 보지 않을 수도 있다. 본 연구에서는 구와 결합한 경우는 제외하고 나머지 '-끼리' 결합형은 파생어로 간주하였다. 단, '원내 대표'와 같이 사전에 단독으로 등재된 단위는 한 단어로 처리하였다.

| | |
|---|---|
| 생청국장 | 1 |
| 생바이러스 | 1 |
| 생보리 | 1 |
| 생딜(dill) | 1 |
| 총빈도: 112<br>신어:8 | |

표 3.19. ‘생-’의 신어와 빈도

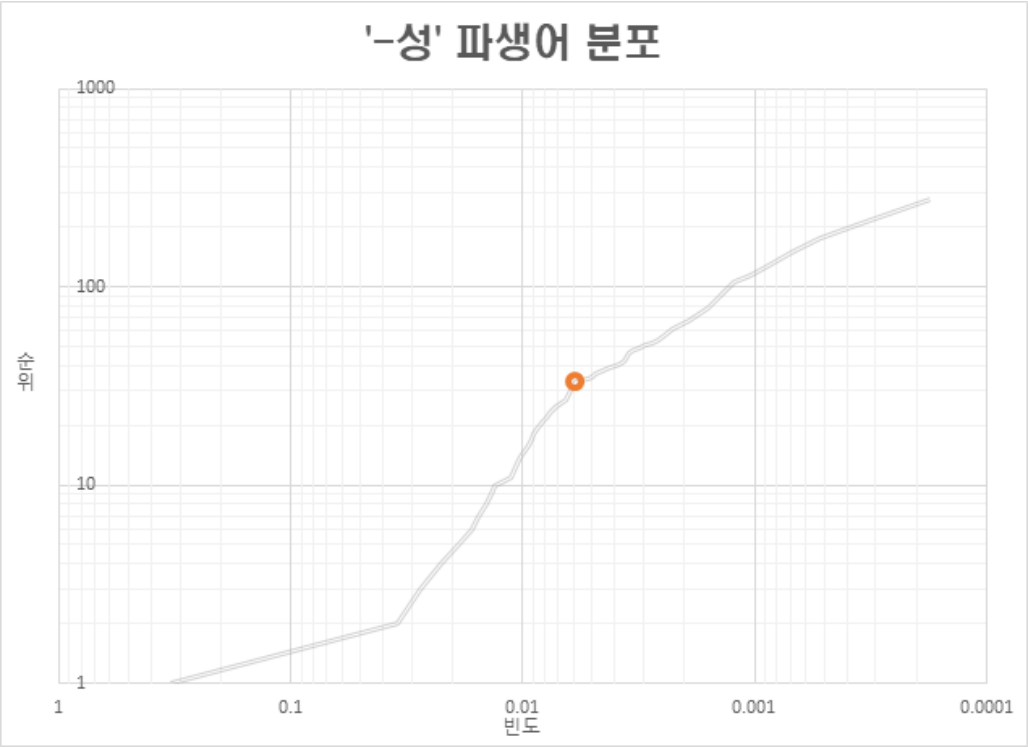
‘생(生)-’은 명사에 결합하여 ‘원래 상태 그대로의’, ‘순전한’, ‘억지스러운’, ‘실제로 낳은’ 등의 다양한 의미를 더하는 한자어 접두사이다. 한자어라는 인식이 거의 사라져 가는 접사로, ‘生’에서 분화된 의미가 많이 추가되었으며 결합 대상도 다양해졌다. ‘생-’은 한자어뿐 아니라 고유어나 외래어·외국어와도 활발히 결합하며 신어를 생산한다.

순위-빈도 그래프에 표시된 세타 파생어는 ‘생지옥’이고 빈도 5이다. 세타 이하인 파생어의 총빈도는 53이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다

P=0.15094

θ=0.04464

### 3.2.7.4. ‘-성(性)’




| 파생어 | 빈도 | 파생어 | 빈도 |
|---|---|---|---|
| 견제성 | 1 | 증진성 | 1 |
| 논평성 | 1 | 지불성 | 1 |
| 방도성 | 1 | 진성성 | 1 |
| 보은성 | 1 | 참고인성 | 1 |
| 빈도성 | 1 | 철면피성 | 1 |
| 쇄신성 | 1 | 최소성 | 1 |
| 스트레치성 | 1 | 커터성 | 1 |
| 시대착오성 | 1 | 토픽성 | 1 |
| 액션성 | 1 | 통용성 | 1 |
| 양자성 | 1 | 핑계성 | 1 |
| 예금성 | 1 | 호칭성 | 1 |
| 오타쿠성 | 1 | 박스성 | 1 |
| 옹호성 | 1 | 아프타성 | 1 |
| 유럽성 | 1 | 괴물성 | 1 |
| 유류성 | 1 | 저해성 | 1 |
| 이글성 | 1 | 코드성 | 1 |
| 임기성 | 1 | 노동자성 | 2 |
| 자기애성 | 1 | 총 빈도: 5729<br>신어:38 | |

| | | |
|---|---|---|
| 예우성 | 1 | |
| 중성성 | 1 | |

표 3.20. '-성'의 신어와 빈도

'-성(性)'은 명사나 명사성 어근과 결합하여 '그러한 성질'이라는 뜻을 더하는 한자어 접미사이다. 의미의 폭이 상당히 넓고, 고유어를 통한다면 통사적 구성으로 길게 표현해야 할 의미를 한 단어로 간단히 표현할 수 있기 때문에 현대 한국어에서 널리 사용된다. '-적(的)'이나 '-성(性)' 등의 파생어가 유독 높은 빈도로 사용되는 것 역시 이러한 표현의 편리성 때문인 듯하다.

'-성' 파생어에는 수식적 용법과 명사적 용법이 있으며, 파생어들을 두 가지 용법이 모두 가능한 부류와 명사적 용법으로만 쓰이는 부류로 나눌 수 있다는 사실이 밝혀진 바 있다(노명희1998, 2005). 가령 다음과 같은 '-성' 파생어는 명사적 용법으로만 사용된다는 것이다.

● 가능성, 강인성, 건전성, 견고성, 결백성, 명료성, 모호성

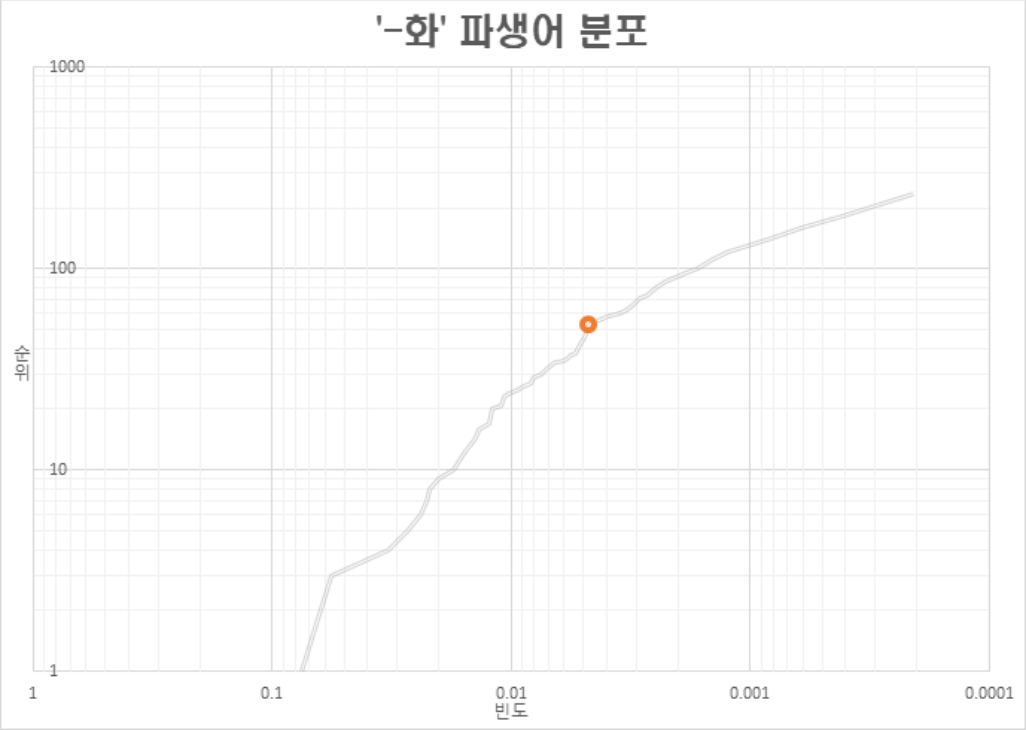
이광호(2009)에서는 이러한 기능적 차이가 의미상의 차이를 수반하며 따라서 빈도 분포나 생산성과 같은 수치상의 특성에도 체계적으로 영향을 준다고 주장하였다.

그러나 노명희(1998, 2005)에서 '-성' 파생어를 구분한 것은 전형적인 용법에 대한 서술이라고 보아야 한다. 적절한 상황이나 맥락만 주어진다면 위와 같은 파생어들도 수식적으로 사용할 수 있다. 단지 그러한 상황이 실현될 확률이 극히 희박할 뿐이다. 전형적인 용법에 어긋나더라도 코퍼스에서 관찰되었다면 그대로 언어의 모습으로 인정하는 것이 계량적 연구의 본질이다. 순위-빈도 분포에서도 한국어 화자들이 두 가지 용법에 속하는 파생어들을 다르게 취급한다는 증거는 찾아볼 수 없다. 따라서 이와 같은 현상은 본 연구에서 고려할 대상이 아니다.

순위-빈도 그래프에 표시된 세타 파생어는 '신뢰성'이고 빈도 34이다. 세타 이하인 파생어의 총빈도는 1796이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.02116
θ=0.00593

3.2.7.5. ‘-화(化)’




| 파생어 | 빈도 | 파생어 | 빈도 |
|---|---|---|---|
| 검사화 | 1 | 자급화 | 1 |
| 경영화 | 1 | 장수명화 | 1 |
| 경찰화 | 1 | 제복화 | 1 |
| 공급지화 | 1 | 대강화 | 1 |
| 그린화 | 1 | 사행화 | 1 |
| 급행화 | 1 | 스마트화 | 1 |
| 기원화 | 1 | 과도화 | 1 |
| 다제화 | 1 | 빅데이터화 | 1 |
| 다지화 | 1 | 모범생화 | 1 |

| | | | |
|---|---|---|---|
| 덤핑화 | 1 | 소프트웨어화 | 1 |
| 동결화 | 1 | 제삼자화 | 1 |
| 바이오가스화 | 1 | 포퓰리즘화 | 1 |
| 백화점화 | 1 | 검찰화 | 2 |
| 산소화 | 1 | 다종화 | 2 |
| 선거화 | 1 | 총빈도: 4840<br>신어:36 | |
| 솔루션화 | 1 | | |
| 수림화 | 1 | | |
| 수익화 | 1 | | |
| 위원화 | 1 | | |
| 일반고화 | 1 | | |

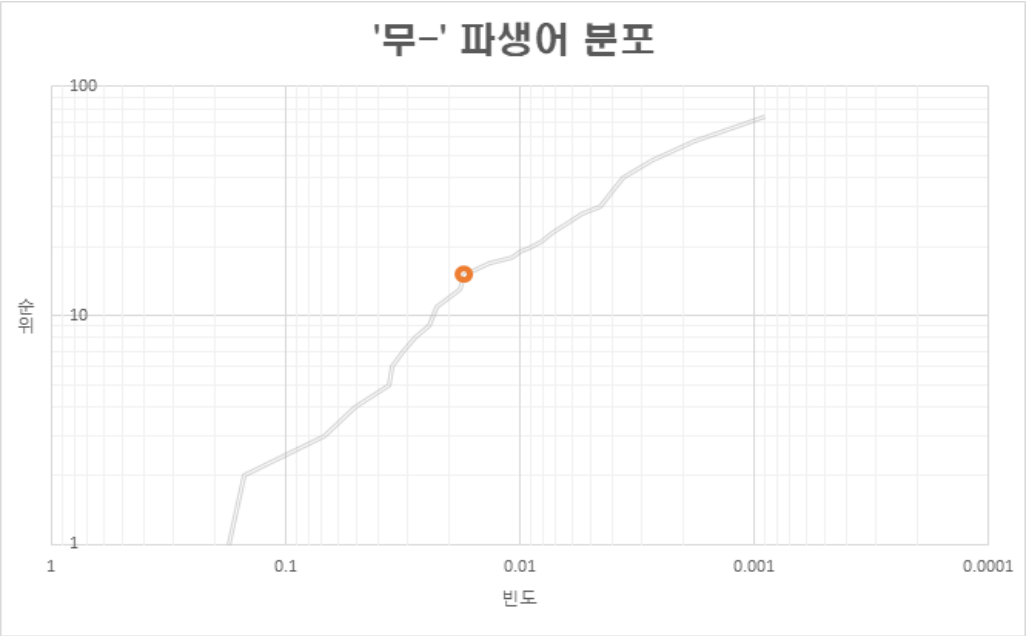
표 3.21. '-화'의 신어와 빈도

'－화(化)'는 명사나 명사성 어근과 결합하여 '그렇게 만들거나 됨'의 의미를 가진 서술성 명사를 파생하는 한자어 접미사이다. 원래 '－하/되－' 와의 결합이 불가능한 명사를 서술어로 사용할 수 있게 해주므로 한국어에서 현대 한국어에서 활발하게 사용된다.

순위－빈도 그래프에 표시된 세타 파생어는 '장기화'이고 52회 출현하였다. 세타 이하인 파생어의 총빈도는 1341이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.02685
θ=0.01074

### 3.2.7.6. '무(無)－'




| 파생어 | 빈도 |
|---|---|
| 무간 | 1 |
| 무볼넷 | 1 |
| 무정량 | 1 |
| 무제초제 | 1 |
| 무피안타 | 1 |
| 무항생제 | 1 |
| 무호르몬 | 1 |
| 무화학 | 1 |
| 무실기 | 1 |
| 무회전 | 1 |
| 무관용 | 2 |
| 무흉터 | 2 |
| 총빈도: 1107<br>신어:14 | |

표 3.22. '무-'의 신어와 빈도

'무(無)-'는 명사와 결합하여 '그것이 없음'의 의미를 더하는 한자어 접두사이다. '무-' 파생어는 어근의 성격에 따라 그 기능적 범주를 다르게 바꿀 수 있다는 특성이 있다(노명희 1998, 2005). 기능적 범주가 달라진다는 것은 어근의 '-하-' 결합 가능 여부가 '무-' 파생 이후에 뒤바뀐다는 것이다. 따라서

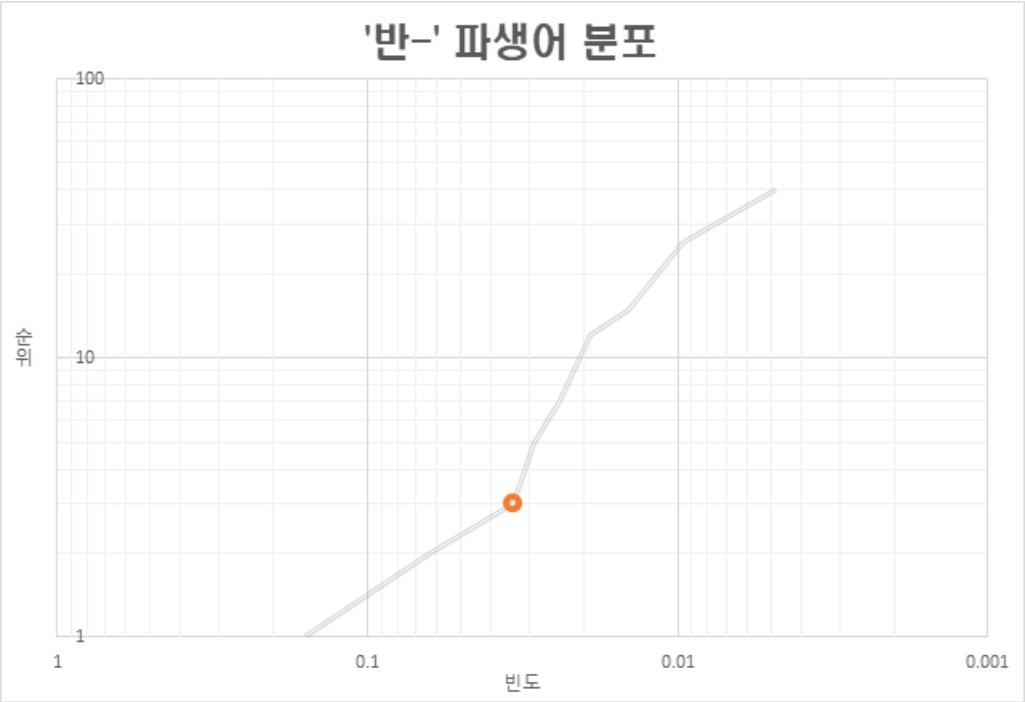
‘무-’ 파생어는 어근과 파생어의 ‘-하-’ 결합 가능 여부(+/-)에 따라 총 네 가지 부류로 구분할 수 있게 된다. 이광호(2009)에서는 이러한 기능적 차이에 따라 파생어가 양적으로 다르게 행동한다고 주장하였다.

그러나 어떤 파생어가 그것을 다시 파생할 수 있는 접사의 종류에 따라 다른 분포를 보인다는 주장은 근거가 부족한 것으로 보인다. 앞서 살펴보았듯 파생어는 의미 차이에 따라 다르게 행동한다는 정도만이 계량적으로 입증 가능하다. 전술한 ‘무-’의 기능적 차이는 큰 의미 차이를 유발하지는 않는 것으로 보인다. 순위-빈도 그래프에서도 빈도 분포상의 교란은 찾아볼 수 없다. 따라서 본 논의에서는 ‘무-’의 기능적 차이를 고려하지 않고 분포를 분석하였다.

순위-빈도 그래프에 표시된 세타 파생어는 ‘무의미’이고 빈도 19이다. 세타 이하인 파생어의 총빈도는 325이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.04308
θ=0.01716

3.2.7.7. ‘반(反)-’




| 파생어 | 빈도 |
|---|---|
| 반촘스키 | 1 |
| 반기능 | 1 |
| 반한류 | 1 |
| 반혁신 | 1 |
| 반엠비 | 1 |
| 반카스트 | 1 |
| 반무르시 | 1 |
| 반히틀러 | 1 |
| 반문재인 | 1 |
| 반토건 | 1 |
| 반인도주의 | 1 |
| 반횡령 | 1 |
| 반낭비 | 1 |
| 반복지 | 1 |
| 반트럼프 | 3 |
| 총빈도: 208<br>신어:17 | |

표 3.23. ‘반-’의 신어와 빈도

‘반(反)-’은 명사와 결합하여 ‘그것에 반대되는/반대하는’의 뜻을 더하는 한

자어 접두사이다. 전통적으로 부정 접두사로 분류되지는 않지만 어근의 뜻과 대립하는 의미를 파생한다. 이러한 의미적 유사성 때문에 의미적으로 유(有)와 무(無)로 나뉘는 어근 중 일부에서는 '비(非)-' 파생어와 경쟁하기도 한다.

순위-빈도 그래프에 표시된 세타 파생어는 '반체제'이고 빈도 7이다. 세타 이하인 파생어의 총빈도는 162이다. 산출된 생산성과 저장 임계 빈도는 다음과 같다.

P=0.10494
θ=0.03365

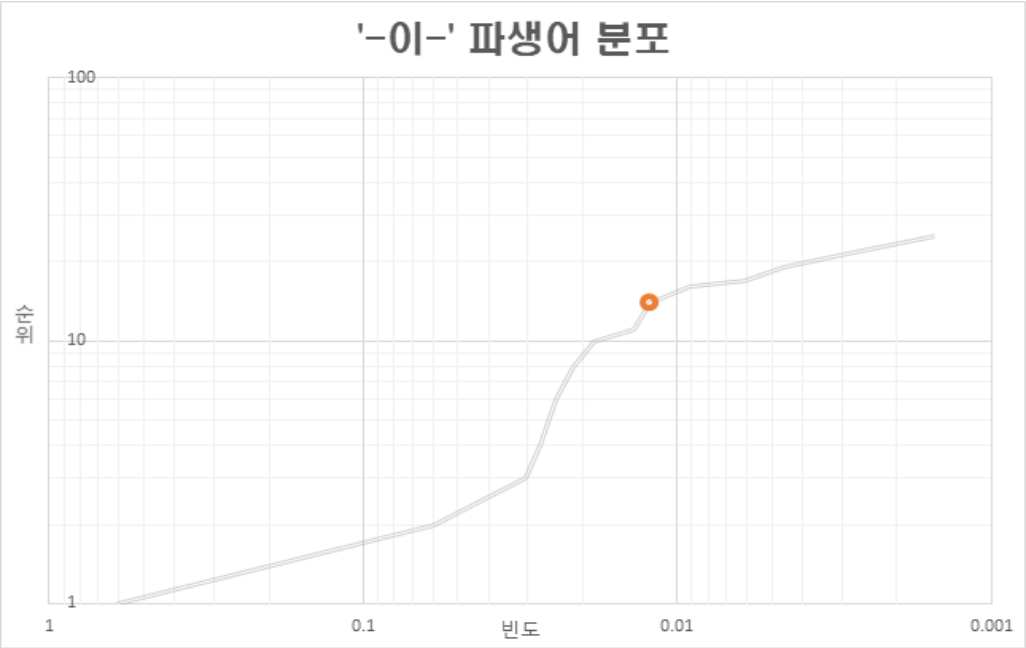
## 3.3. 비생산적 접사와 신생 접사

3.2에서는 비생산적 접사와 신생 접사의 파생어와 빈도를 정리할 것이다. 처리 결과 생산성이 0으로 나온 접사들을 비생산적 접사로 분류하였다. 비생산적 접사에는 빈도 구조상으로는 생산적 접사와 크게 차이 나지 않는 것들도 있고, 총 빈도나 유형 빈도가 너무 낮아서 분석이 불가능한 '사멸 접사'도 있다.[60][61] 신생 접사는 접사처럼 사용된 지 오래되지 않은 접사들로, 파생어들이 전부 신조어이다. 빈도 구조나 파생어의 규모 면에서는 사멸 접사와 구분하기 힘들다. 신생 접사는 사전에 등재되지 않으므로 코퍼스에 나타난 접사 중 일부를 정리하였다. 사멸 접사와 신생 접사들은 순위-빈도 분석 없이 파생어와 파생어별 빈도의 목록만 간단히 제시하겠다.

### 3.3.1. '-이(동사 파생 접미사)-'

60) P=0으로 나타난 접사가 실제로 생산성이 전혀 없다고 섣불리 결론지을 수는 없다. 전술하였지만 코퍼스를 통해 산출한 지표에는 오차가 있을 수밖에 없기 때문이다. 단 한 건의 신어 출현에 따라 생산적/비생산적 접사가 나뉠 수 있는 생산성이 낮은 접사에서는 특히 주의해야 한다. 따라서 3.2에서 다루는 접사 중 '-이-', '불(不)-' 등은 '생산성이 매우 낮은 접사'로 보아야 실제 언어 현실에 맞을 것이다.

61) 유형 빈도에 대한 설명은 각주 12 참고.




| |
|---|
| 총빈도: 655 |
| 유형 빈도: 40 |
| 저장 임계점 순위: 14 |
| 저장 임계점 빈도: 8 |

표 3.24. '-이-' 파생어의 빈도 정보

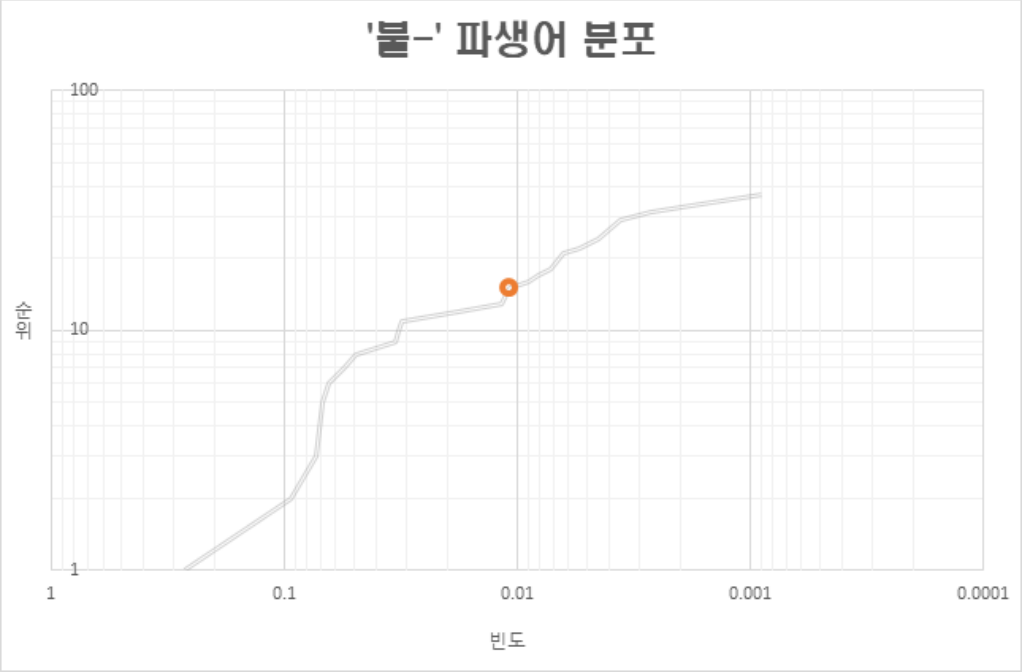
'-이-'는 의성·의태어 어근과 결합하여 '그런 동작이 반복됨'의 의미를 더하고 동사를 파생하는 접미사이다. 파생하는 거의 모든 어근에서 '-거리-' 혹은 '-대-'와 경쟁한다. 즉, 거의 모든 어근에서 '-거리-, -대-'와 의미 차이 없이 바꿔 쓸 수 있다.

코퍼스에서는 신어가 하나도 나타나지 않았으나 '-이-'가 실제로 완전히 생산성을 잃었다고 보기는 힘들다. 고빈도 구간과 구별되는 저빈도 구간을 명확히 확인할 수 있기 때문이다. 코퍼스의 크기가 더 컸다면 신어를 포착할 수 있었을 듯하다. 다만 경쟁 접사의 파생을 완전히 저지할 수 있는 '움직-(빈도

401)', '망설-(빈도 39)' 등의 극소수 어근에 기형적으로 집중된 분포를 고려하면 실제 생산성 역시 아주 낮을 것이다. 이는 배타적으로 결합하는 소수 어근의 파생어들이 파생어 집합 전체를 독점하는 빈도 분포를 통해서도 추론할 수 있다.

'-이-'가 '-거리-, -대-'와 의미가 거의 같은데도 생산성을 잃고 위축되는 이유는 결합 제약에 있는 듯하다.

### 3.3.2. '불(不)-'




| |
|---|
| 총빈도: 1113 |
| 유형 빈도: 47 |
| 저장 임계점 순위: 15 |
| 저장 임계점 빈도: 12 |

표 3.25.1. '불-' 파생어의 빈도 정보

‘불(不)-’은 명사와 결합하여 ‘~가 아님’, ‘~하지 아니함’, ‘~에 어긋남’ 등의 의미를 더하는 한자어 접두사이다. 많은 어근에서 ‘비(非)-’와 경쟁할 수 있지만 단순 부정의 의미만을 더하는 ‘비-’에는 없는 의미 때문에 완전히 동의적이라고 보기는 힘들다.[62] 즉, 두 접사는 전면적인 경쟁 관계에 있지 않으며, 동의 관계가 성립하는 일부 파생어에서만 경쟁한다. ‘불-’이 경쟁에서 밀리고 있는 다른 접사들보다 비교적 건재한 것은 이처럼 우세한 경쟁 접사와의 경쟁이 제한되었기 때문인 듯하다. 그럼에도 불구하고 ‘불-’은 표 3.24.2와 같이 빠르게 쇠퇴하고 있다.

| | 이광호(2009) | 본 연구 |
|---|---|---|
| 유형 빈도 | 2479 | 1113 |
| 총빈도 | 88 | 46 |
| 1위 빈도 | 611 | 296 |

표 3.25.2. ‘불-’의 통시적 변화

### 3.3.3. 기타 접사들의 파생어와 빈도

| | 접사 | 유형빈도 | 총빈도 | (파생어, 빈도) |
|---|---|---|---|---|
| 비생산적·사멸 접사 | -장이 | 1 | 1 | (거짓말장이, 1) |
| | -수(手) | 12 | 5651 | (교수, 2874), (선수, 2305), (가수, 273), (고수, 109), (무용수, 37), (백수, 12), (저격수, 11), (조수, 8), (조타수, 8), (소방수, 7), (적수, 6), (기수(騎手), 1) |
| | 맨- | 12 | 45 | (맨발, 10), (맨손, 7), (맨땅, 7), (맨얼굴, 4), (맨살, 4), (맨몸, 4), (맨입, 3), (맨바닥, 2), (맨눈, 1), (맨밥, 1), (맨정신, 1), (맨다리, 1) |
| | 몰(沒)- | 3 | 5 | (몰염치, 1), (몰지각, 2), (몰상식, 2) |
| | 뒤- | 8 | 143 | (뒤흔들, 33), (뒤덮, 28), (뒤섞, 25), (뒤바뀌, 23), (뒤엉키, 15), (뒤틀, 15), (뒤얽, 3), (뒤젓, 1) |
| | 짓- | 7 | 39 | (짓누르, 15), (짓밟, 14), (짓이기, 5), (짓뭉개, 2), (짓무르, 2), (짓찧, 1) |
| | 치- | 5 | 153 | (치솟, 97), (치닫, 45), (치밀, 9), (치달리, 1), (치받, 1) |

62) ‘불(不)-’과 ‘비(非)-’의 의미 차이에 대해서는 3.1.6에서 다루었다.

| | | | | |
|---|---|---|---|---|
| | 휘- | 8 | 162 | (휘두르, 87), (휘말, 44), (휘젓, 20), (휘감, 4), (휘날리, 3), (휘몰아치, 2), (휘둥그렇, 1), (휘몰, 1) |
| | 드- | 3 | 13 | (드넓, 5), (드높, 5), (드세, 3) |
| | 들- | 1 | 1 | (들쑤시, 5) |
| | 풋- | 3 | 66 | (풋고추, 42), (풋마늘, 23), (풋냉이, 1) |
| | 처- | 4 | 6 | (처박히, 2), (처박, 2), (처넣, 1), (처던지, 1) |
| | 왕- | 6 | 8 | (왕방울, 1), (왕벚꽃, 1), (왕모기, 1), (왕초보, 1), (왕언니, 3), (왕뾰루지, 1) |
| 신생 접사 | 갓(god)- | 2 | 3 | (갓티스트, 2), (갓형석, 1) |
| | 꿀- | 4 | 8 | (꿀맛, 5), (꿀벅지, 1), (꿀잠, 1), (꿀팁, 1) |
| | -짱 | 2 | 7 | (얼짱, 5), (몸짱, 2) |

표 3.26. 비생산적 접사와 신생 접사의 파생어와 빈도. 밑줄 친 파생어는 신어이다.

표 3.26에는 저빈도 비생산적 접사와 사멸 접사, 신생 접사 일부의 정보를 정리하였다. 저빈도 비생산적 접사는 파생어에서 신어가 전혀 관찰되지 않은 접사 중 빈도가 매우 낮은 접사이고, 사멸 접사는 신어가 나타났지만 사실상 거의 쓰이지 않는 것으로 보이는 초저빈도 접사를 의미한다. 코퍼스의 수집 방식에 따라 접사당 총 1회 정도의 신어는 포함이 될 수도 있고 그렇지 않을 수도 있으므로 저빈도의 비생산적 접사와 사멸 접사 사이에는 큰 차이가 없다. 신생 접사는 한국어에서 새로이 접사 기능을 수행하게 된 접사이다. 신생 접사에 의한 파생어는 대부분 유행어의 성격이 짙은데다 비속어이기 때문에 접사로 인정하기 어려운 면이 있다. 그러나 접사가 생겨나고 사라지는 모습을 모두 포착해보고자 분석 대상에 포함하였다.

많은 고유어 접두사가 비생산적이거나 사멸 단계에 있다는 점이 눈에 띈다. '뒤−, 짓−, 휘−, 처−, 드−, 들−'과 같이 유사한 의미를 공유하는 접사들이 한꺼번에 위축되기도 하고, '처−'와 같이 비속어스러운 단어를 파생하는 접사가 세력을 잃기도 한다. 그런데도 신조어 접두사는 생겨나고 있으므로 고유어 접두 파생 자체가 위축되는 것으로는 볼 수 없을 듯하다. 그러나 전반적으로 고유어 접두사에 의한 파생의 세력이 전반적으로 매우 약하다는 점은 부정하기 힘들 듯하다. 공시적으로 한국어의 생산적 접두 파생은 대부분 한자어 접두사에 의해 이루어지고 있는 것이다.

# 4. 한국어의 저지현상

## 4.1. 도입

저지가 유무적 현상이 아니라는 것은 주지의 사실이다. 심지어 같은 뜻을 가진 파생어가 있더라도 다른 접사가 같은 어근과 결합하여 새로 파생어를 만들기도 한다. 그러나 언어 사용자는 보통 같은 뜻을 표현하기 위해 굳이 여러 형태를 사용하지 않으며, 이 때 동의적 형태 중 사용할 단어를 정하는 것은 경험에 의존하는 경향이 있다. 따라서 코퍼스를 통해 대규모로 수집한 파생어의 빈도 자료에서는 의미가 같거나 유사한 대부분의 파생어 간에 절대적인 저지가 일어나거나 한 접사의 파생형이 압도적으로 많이 나타난다. 이렇게 한 접사가 다른 접사들을 압도하는 힘의 근원이 무엇인지에 대해 오랫동안 많은 논의가 있었으나 대체로 저지는 파생어의 빈도에 의해 일어난다는 견해가 지배적이었다. 어떤 파생어의 빈도가 높을수록 어휘부에 저장된 강도 역시 높아질 것이므로 어떤 의미가 어휘부에서 구체적인 형태로 출력될 때 경쟁 접사를 앞서기 쉬워진다는 것이다. 3장에서는 이런 이론의 타당성을 검증하기 위해 각 접사의 의미를 명확히 하고 파생어의 저장 지점인 저장 임계 빈도를 접사마다 구체적인 수치로 도출했다. 4장에서는 3장의 결과를 활용하여 저지에 대한 기존의 논의가 얼마나 타당한지를 살펴보고, 그렇지 않은 경우엔 어떻게 설명할 수 있을지에 대해 살펴볼 것이다.

## 4.2. 유형 저지

선행 연구에서 정립한 유형 저지(type blocking)의 정의에 따르면 유형 저지는 한 접사가 다른 접사의 출현을 완전히 막는 현상이다. 이러한 저지는 전혀 관찰할 수 없다. 접사의 의미가 비슷하더라도 일부 파생어에서는 의미가 달라지기 마련이고, 실제로는 쌍형어도 많이 사용하기 때문이다.

4.2에서 다룰 유형 저지는 이러한 기존의 개념과 상당히 다르다. 의미가 비슷하여 적어도 일부 파생어에서 의미의 차이가 중화될 수 있는 접사들 사이에는 경쟁이 성립하는데, 이 때 한 접사가 다른 경쟁 접사들의 생산성을 억제하게 된다. 유형 저지의 피저지 접사들은 잠재 생산성보다 낮은 생산성을 띠게 되며, 저지당하지 않는 접사들과는 다른 분포를 보이게 된다. 본 연구에서는 이러한 경쟁 접사 사이의 신어 생산 방해 현상을 유형 저지로 정의한다.

유형 저지에 대해 본격적으로 논하기 전에 4.2.1에서 우선 유형 저지를 포착할 방법을 설명하고 그 방법의 타당성을 검증할 것이다. 4.2.2에서는 이렇게 포착된 유형 저지의 모습을 통해 그 본질을 살필 것이다.

### 4.2.1. 저장 임계 빈도와 생산성의 상관관계 검증

| 접사 | θ | P | θ 순위 | P 순위 |
|---|---|---|---|---|
| -가(家) | 0.005549 | 0.018519 | 17 | 12 |
| -자(者) | 0.010913 | 0.006536 | 12 | 15 |
| -이(명사파생) | 0.011921 | 0.005797 | 11 | 16 |
| -음 | 0.004398 | 0.001 | 19 | 19 |
| -기 | 0.009524 | 0.015873 | 13 | 13 |
| -스럽- | 0.020448 | 0.042735 | 9 | 8 |
| -하(형용사)- | 0.009325 | 0.005033 | 14 | 18 |
| -롭- | 0.030645 | 0.008403 | 6 | 14 |
| -지- | 0.033816 | 0.020833 | 3 | 11 |
| -적(的) | 0.008454 | 0.005764 | 15 | 17 |
| -꾼 | 0.026667 | 0.079365 | 7 | 6 |
| 비(非)- | 0.02246 | 0.098947 | 8 | 5 |
| -답- | 0.03125 | 0.37255 | 5 | 1 |
| -끼리 | 0.157895 | 0.324324 | 1 | 2 |
| 생(生)- | 0.044643 | 0.150943 | 2 | 3 |
| -성(性) | 0.005237 | 0.022701 | 16 | 10 |
| -화(化) | 0.004733 | 0.021158 | 18 | 9 |
| 무(無)- | 0.017164 | 0.043077 | 10 | 7 |
| 반(反)- | 0.033654 | 0.104938 | 4 | 4 |

표 4.1. 생산적 접사의 θ-P 분포표

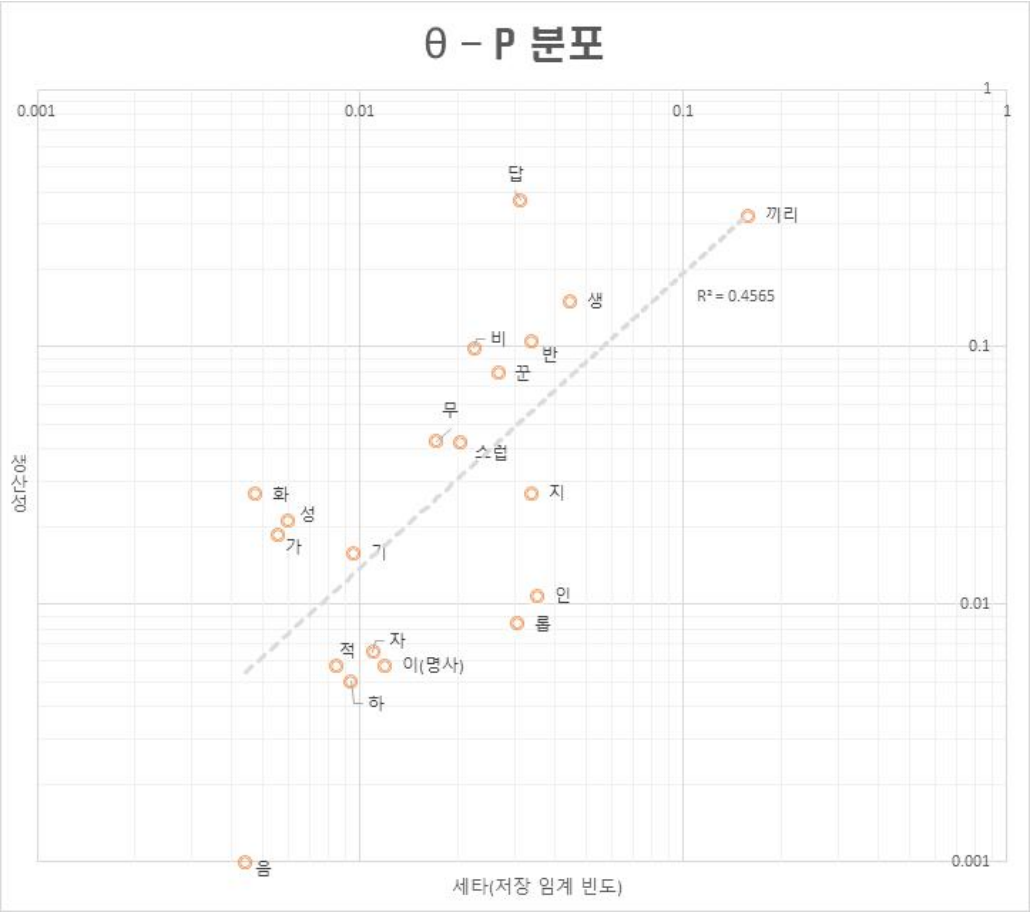
표 4.1은 생산적 접사의 생산성 추정치(P)와 저장 임계 빈도 추정치(θ), 두

변수[63]의 접사별 순위 정보를 정리한 것이다. 두 변수의 순위인 'θ 순위, P순위'는 두 값을 모두 구할 수 있었던 접사 집합 내에서의 해당 접사의 θ, P 크기 순위를 의미한다.

신어를 관찰하지 못해 생산성이 0으로 산출되었거나(P=0) θ를 구할 수 없었던 접사는 상관관계 분석에서 제외하였다. P=0인 접사를 제외하지 않으면 하나의 값에 너무 많은 θ가 대응하게 되기 때문이다. 이런 값들은 이상치가 된다. 이상치를 분석에 포함시키면 자료가 왜곡되어 올바른 분석을 할 수가 없다. P가 0인 접사를 제외하는 또 다른 이유는 전술하였듯 P가 실제 생산성의 추정치에 지나지 않기 때문이다. 극도로 생산성이 낮은 접사가 파생한 신어도 P=0으로 나타날 수 있으므로 P값만을 보고 실제 생산성이 전혀 없다고 단정할 수 없다. 특히 본 연구에서 채택한 것과 같은 소규모 코퍼스에서는 생산성이 매우 낮은 경우 신어를 관찰하기 어렵다.

표 4.1에서 θ와 P에 어느 정도의 상관관계가 있음을 짐작할 수 있다. 두 변수의 변화 추이를 확인하기 위해서는 분포를 시각적으로 나타내면 편리하다. 그래프 3.1에는 2차원 분포도로 표 4.1의 정보를 표시하였다. 세로축에는 추정치 산출 방식이 여러 선행 연구에 의해 확립된 P를 표시하였고, θ는 본 연구에서 산출 방식을 새롭게 고안한 지표이므로 독립 변수로 보아 가로축에 두었다. 평면을 가로지르는 점선은 거듭제곱법에 의한 추세선이다. 추세선은 P와 θ가 이동하는 전반적인 경로를 나타낸 선이다.

63) P와 θ는 각 접사마다 특별히 갖는 값이므로 각각 하나의 변수로 취급할 수 있다.




그래프 4.1. 세타-생산성 분포

두 변수의 시각적 분포로부터 θ와 P 사이에 실제로 어느 정도 비례 관계가 있음을 조금 더 확신할 수 있다. 두 값의 편차가 큰 접사도 종종 있지만, 대체로 한 값이 크면 다른 값도 커지는 경향은 확실한 듯하다. 표시된 추세선도 확실히 우상향하는 모습을 보여준다.

θ와 P가 비례하는 것은 언어학적으로는 당연하지만 수학적으로는 그렇지 않다. 언어학적 측면에서 보면, 어떤 접사의 생산성이 높을수록 접사와 어기를 분리해서 인식하는 경향이 강해지므로 파생어 전체가 어휘부에 저장되기도 힘들어질 것이다. 그렇다면 두 변수는 양의 상관관계에 있게 된다. 반면, 본 연

구의 공식에 따르면[64] θ와 P는 반비례 관계에 있으므로 수학적으로 이 관계는 예측에 어긋나는 결과이다.

따라서 표 4.1과 그래프 3.1에 정리된 통계 자료에 나타난 상관관계가 통계적 우연에 따른 것은 아닌지 검증하는 절차가 필요하다. 통계학적 검증 결과 관계가 유의(significant)하지 않다면, 저장 임계 빈도와 생산성에 관한 실증적·이론적 선행 연구가 모두 틀렸거나 본 연구에서 채택한 저장 임계 빈도 및 생산성 산출 방법에 심각한 결함이 있는 것이다.

반대로 생산성과 저장 임계 빈도의 상관관계가 유의한 것으로 밝혀지면 이를 분석에 활용할 수 있다. 두 변수의 상관관계를 활용하면 저장 임계 빈도를 구체적인 비교 기준으로 삼아서 생산성의 높고 낮음을 입체적이고 정확하게 평가할 수 있다. 비교 기준이 없으면 단순히 어떤 접사의 생산성을 다른 접사의 전반적인 생산성과 비교하는 수준에 그칠 수밖에 없다. 가령, 기존의 분석 방식으로는 '-답-'과 '-끼리' 모두 다른 접사에 비해 월등히 높은 생산성을 나타내므로 비정상적인 부류로 처리해야 한다는 결론에 만족해야 한다. 그러나 θ-P 분포를 활용하면 '-끼리'의 생산성은 세타에 알맞은 수준이고 '-답-'의 생산성만이 전반적인 추세보다 상당히 높은 편임을 알 수 있다. 상대적 생산성의 크기에 따라 접사들을 분류함으로써 신어 생산에서의 유형 저지 양상도 바로 파악할 수 있다. 통계 결과의 언어학적 의미를 논하고 유형 저지의 양상을 분석하기 전에 검증을 먼저 진행하기로 한다.

검증을 위해 수립한 귀무가설과 대립가설은 다음과 같다. 유의수준은 비자연과학 분야 연구의 관례에 따라 0.05로 설정한다.

$H_0$ : θ와 P 사이에는 아무런 상관관계가 없다.

$H_A$ : θ와 P 사이에는 양의 상관관계가 있다.[65]

---

64) 본 연구의 P 산출 공식은 다음과 같다.

$$P = \frac{N(\text{신어의 빈도의 합})}{F_\theta(\text{빈도가 } \theta \text{ 이하인 파생어들의 빈도의 합})}$$

공식의 도출 과정은 2.3에서 설명하였다.

65) 따라서 단측검정을 실시해야 한다.

검정통계량을 계산하기 위한 상관계수로는 켄달의 등위상관계수(Kendall's tau, τ)을 선택하였다. 상관계수의 추정치는 다음과 같이 계산되었다(소수점 아래 일곱번째 자리에서 반올림하였다).[66] 수치가 1에 가까울수록 양의 상관관계가 강력하다는 의미이다.

$\tau_h$=0.442105

적률상관계수 r은 가장 보편적으로 쓰이는 상관계수이지만 변수가 변하는 정도를 직접 비교하기 때문에 θ와 P를 분석하는 데 적합하지 않다. τ는 한 변수의 순위가 높거나 낮을 때 다른 변수의 순위도 비슷한 위치에 있는지를 가늠하기 위해 사용하는 등위상관계수이다. P와 θ는 절대적인 크기보다는 서로 간의 상대적인 크기가 더 유의미한 지표이므로 r보다는 τ가 더 적합하다. 또, τ는 순위의 변화 방향까지를 엄격하게 따지기 때문에 다른 등위상관계수에 비해서도 가설검증의 타당성이 더 높다. 따라서 본 연구에서는 $\tau_h$를 검정통계량으로 선택하고 가설검정을 실시하였다.

귀무가설 하의 표준편차, 검정통계량을 표준편차를 통해 표준화한 Z-통계량, Z-통계량에 해당하는 P-값은 다음과 같이 산출되었다.

표준편차 = 0.167153
Z-통계량 = 3.043745
P-값 = 0.001168

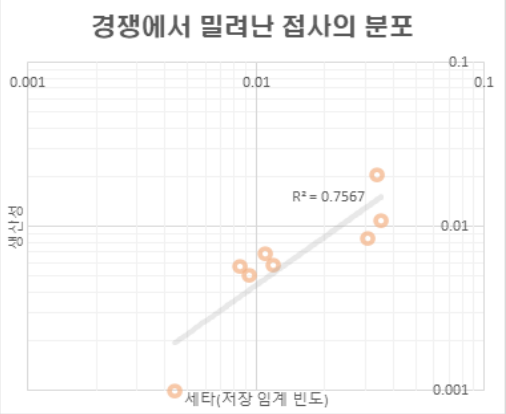
P값이 유의수준 0.05보다 작으므로 귀무가설을 기각하고 대립가설을 채택할 수 있다. 즉, 유의수준 0.05에서 θ와 P 사이에 양의 상관관계가 있다는 주장은 통계적으로 타당하다. 주목할 점은 P-값이 극도로 작다는 점이다. 이는 유의수준 0.05가 아니라 0.005와 같은 매우 엄격한 기준에서도 대립가설을 채

66) 아래 계산된 값들은 모집단인 '실제 발화된 말 전체의 집합'이 아니라 모집단의 표본을 통해 얻은 수치이므로 상관계수의 추정치이다. 실제 값이 아닌 추정치에는 기호 위에 '^'을 붙이거나 다른 아래 첨자를 붙여서 표시하는 것이 관례이다. 여기서는 아래 첨자 h를 사용하였다.

택할 수 있는 크기이다. 즉, 통계학적으로 θ와 P는 거의 확실하게 양의 상관관계에 있다.

### 4.2.2. 접사 간 저지 양상

저장 임계 빈도와 생산성 사이의 관계가 검증되었으므로 두 변수의 값이나 분포가 가지는 의미를 해석해볼 수 있다. θ-P 분포를 보면, 경쟁에서 밀려난 접사들은 다른 접사에 비해 확연히 낮은 생산성을 보인다. 회색으로 표시된 추세선을 경계로 하여 그 위쪽에 분포하는 접사들은 경쟁하지 않거나 신어 생성 경쟁에서 우위를 점하고 있는 접사들이다.[67] 추세선 아래에는 경쟁에서 밀려난 접사들만이 자리 잡고 있다. 두 부류의 접사를 따로 떼어 보면 두 부류의 접사들이 다음과 같이 거의 직선상에 분포하므로 상대적 생산성(=생산성/세타)이 같은 부류 내에서 일관되게 나타남을 알 수 있다.




그래프 4.2. 저지당하지 않은 접사의 분포　　그래프 4.3. 피저지 접사의 분포

추세선에 표시된 $R^2$ 값은 앞서 설명한 적률상관계수 r의 설명력을 나타내는 수치로, 종속 변수(생산성)의 변화를 독립 변수(저장 임계 빈도)가 얼마나 잘

67) 이런 접사들은 대체로 저지현상에서도 다른 접사의 침범을 막거나 다른 접사의 영역에 쉽게 침투하는 것으로 보이지만 양상이 일관적이지는 않다. 이에 관한 내용은 4장에서 다룬다.

설명하는지를 나타낸다. 접사 전체(0.4616)에서보다 설명력이 월등히 높아져 각 집단이 기존에 비해서나 절대적으로나 상당히 높은 군집성을 띤다는 것을 알 수 있다. τ 역시 각각 0.7, 0.79로 상관성이 훨씬 높다. 이를 통해 전체 접사 집합에서 변수의 설명력이 낮았던 것은 접사들이 이질적인 두 집단으로 이루어져 있었기 때문이며, 그 두 집단은 경쟁에서 밀려난 접사와 나머지 접사들이라는 것을 알 수 있다. 낮은 생산성을 보이는 집단이 경쟁에서 밀려난 접사라는 것은, 경쟁 관계에 있지 않은 '-답-', '-끼리'와 같은 접사들은 하나도 피억제 접사 접사와 같은 분포를 보이지 않았다는 점에서 추론할 수 있다. 경쟁하지 않는 접사들이 경쟁하는 두 집단 간의 대조에서 통제집단의 역할을 한 셈이다.

이처럼 '-지-, -롭-, -하-, -적(的), -이, -음, -자(者)'의 신어 생성은 생산적인 경쟁 접사 '-스럽-, -기, -가(家)'에 의해 체계적으로 억제된다.

경쟁 접사들 사이에 의미가 균일한 정도는 제각각이지만 의미 차이가 있더라도 어근에 따라 그 차이가 중화될 수 있음을 앞서 서술하였다. 이 정도의 의미적 관련성만으로도 언어 사용자에게는 접사 간의 의미적 연관성이 각인되고 접사의 신어 파생 가능성에도 영향을 줄 수 있는 듯하다. 즉, 일단 '-스럽-, -하-, -롭-, -지-, -적'이 모두 비슷한 의미를 띠는 접사로 인식하게 되면, 새로운 단어를 만들 때에는 기존에 신어 생산에 상대적으로 훨씬 자주 사용했던 '-스럽-'을 선택할 확률이 높아지는 것이다.[68]

이처럼 경쟁 관계에 있는 접사가 경쟁 상대의 신어 파생을 억제하는 현상을 일종의 저지로 볼 수 있다. 같은 의미영역 내의 다른 접사가 파생하는 것을 방해한다는 점과 저지당하는 쪽의 출현이 완전히 불가능하지는 않다는 점에서 저지의 정의에 들어맞기 때문이다. 그런데 저지하는 접사가 어떤 신어를 파생했을 때 꼭 그 항목에 대해서 파생 방해가 일어나는 것은 아니므로 기존에 유일하게 실존하는 현상으로 인정받아온 항목 저지(token blocking)에는 해당한다고 볼 수 없다. 접사가 다른 접사의 파생을 전반적으로 억누른다는 점에서는 오히려 기존 논의에서 존재 자체가 부정되었던 유형 저지(pattern blocking)에 가까운 듯하다.[69]

---

68) 이것은 얼핏 순환논증처럼 보일 수 있으나, 저지현상은 별다른 이유 없이 '원래 이 형태를 사용하기 때문에' 사용하는 현상이다.

그런데 앞서 제시한 통계적 분석 결과를 유형 저지로 분석하기 위해서는 경쟁 접사로 설정한 접사들이 실제로 경쟁한다는 것을 밝힐 수 있어야 한다. 접사 간의 의미가 다르다면 경쟁이 성립할 수 없고, 따라서 저지로도 볼 수 없다.

본 연구에서 양의 생산성을 가지면서 경쟁 관계에 있다고 본 '-자(者), -가(家), -인(人)', '-이, -음, -기', '-스럽-, -하-, -롭-, -지-, -적(的)', '-꾼, -쟁이, -장이', '비(非)-, 불(不)-' 등[70]의 묶음은 그 구성원의 의미영역이 완전히 일치하지는 않지만 일부의 의미에서 동시에 파생이 가능하다. 직관적으로 의미가 다르게 느껴지는 이들 접사를 경쟁 및 저지 관계에 있다고 본 것은 바로 이런 이유에서이다. 모든 접사가 파생 가능한 의미에 대해, 단지 언중의 언어 사용의 관성 때문에 일부 접사의 신어 파생이 제한되었다면 이는 저지 현상으로 볼 수 있다. 1장에서 소개한 저지의 동의성 조건과 생산성 조건을 만족하기 때문이다. 다음에서는 각 경쟁 집단이 각자의 의미영역에서 구체적으로 어떻게 경쟁하여 앞선 통계 결과가 나올 수 있었는지를 차례로 살펴보기로 한다.

1) '-자(者), -가(家), -인(人)'

각 접사의 파생어는 크게 고빈도어, 저빈도어, 신어로 나눌 수 있다. 고빈도어에서는 그 접사의 전형적인 용법을 확인할 수 있고, 저빈도어에서는 신어는 아니지만 신어처럼 새로 조합해서 쓰는 말과 기존에 사용되던 것이지만 모종의 이유로 퇴화되어 낮은 빈도로 쓰이는 단어들을 확인할 수 있으며, 신어에서는 신어 파생에 있어 그 접사가 어떻게 사용되고 있는지를 확인할 수 있다. 신어 파생에서의 경쟁과 저지를 제대로 파악하려면 이들 모두를 살펴야 한다. 특히 저빈도어는 수가 부족하기 쉬운 신어를 대신해 저빈도 구간의 분포를 보여줄 수 있으므로 꼭 분석 대상에 포함해야 한다. 그런데 유형 빈도가 적게는 100에서 많게는 1000을 넘는 접사의 파생어 모두를 살펴보는 것은 비효율적

69) 유형 저지와 항목 저지의 정의에 대해서는 1.3에서 정리하였다.

70) '-거리-, -대-, -이-'는 독특한 분포 형태 때문에 저장 임계 빈도를 구할 수 없고, 의미 특성으므로 다른 경쟁 묶음들과 함께 논의하기에 어색한 면이 있다. 4.2에서는 이들의 의미 특성을 간단히만 언급할 것이다.

이므로 일부를 추려내어 비교하는 편이 바람직하다. 따라서 고빈도어는 세타 이상의 파생어, 저빈도어는 빈도가 1이고[71] 신어가 아닌 단어 중 무작위로 뽑은 20개만을 살펴보기로 한다.

접사의 용법을 제시하려면 그 의미를 먼저 구분해야 한다. 노명희(2005)에서는 '-자(者), -가(家)'의 의미가 비슷하지만 다른 점도 있다고 하였다. 예를 들어, '-가'의 경우 다음의 목록 (1)에서는 [+전문성, +직업성]이므로 '-자'와 유사하지만 (2)는 [-직업성]이어서 구별된다는 것이다.

(1) 극작가, 도예가, 만화가, 무용가, 미술가, 사업가, 예술가, 웅변가, 음악가, 점성가, 점술가, 정치가, 조각가, 체육가, 감정가, 작곡가, 작명가, 탐정가, 평론가
(2) 사교가, 전략가, 만담가, 이론가, 전문가, 전술가, 정략가, 책략가, 외교가

이러한 구분은 대체로 타당한 것으로 보이지만 단어 개별적으로는 이론의 여지가 있다. 가령, 표준국어대사전에 따른 '웅변가'의 정의는 '웅변을 잘하는 사람'이므로 (1)보다는 (2)에 포함되어야 한다고 주장할 수도 있다. '외교가' 역시 '외교관'과 비슷하게 [+직업성]에 가까운 것으로 보이기도 한다.

그런데 이처럼 접사나 개별 파생어의 미묘한 의미 차이를 샅샅이 밝히는 것은 본 논의에서 불필요하다. 의미가 심각하게 다르지 않다면 경쟁하는 접사들이 의미영역을 분담하는 모습을 관찰할 수 있는 것으로 충분하다. 실제로 표 4.2와 표 4.3을 보면 (1)과 (2)의 의미를 구분하지 않고 [+직업성]이라는 공통점만을 확인한 채 파생어를 분류하더라도 의미에 따른 접사의 파생 비율 차이를 쉽게 알아볼 수 있다(의미 ②).

의미 ①은 '~하는 사람', '~한 사람' 정도의 의미를 갖는 파생어를 묶은 것으로, 역시 이질적인 집합으로 보아 일일이 따져볼 수도 있지만 어떤 행위를 단순히 수행하였다는 공통점에 집중하여 하나의 의미로 처리하였다. 한편, 목록 (2)의 파생어들 중 도저히 [+직업성]이나 [+반복성]의 의미를 찾을 수 없는 것들은 오히려 분리해 내어서 '~을 (많이) 가진 사람', '~의 특성을 가진 사

71) 조건을 만족하는 항목이 부족할 경우 빈도2, 빈도3...의 순으로 추가하였다.

람'과 묶었다.[72] 두 접사뿐 아니라 '-인' 사이의 분업을 잘 보여줄 수 있기 때문이다. 의미를 필요에 따라 나누거나 합치지 않고 더 세밀한 구분을 추구하더라도 앞선 계량적 조사 결과의 해석은 크게 달라지지 않는다. 이에 대해서는 자료를 살펴본 뒤에 설명하겠다.

우선 세 접사가 어근에 공통적으로 더할 수 있는 의미를 정리하면 크게 ① '~하는 사람, ~한 사람', ② '~이 직업이거나 직업처럼 익숙하게 하는 사람', ③ '그러한 (특성을 가진) 사람' 정도로 나눌 수 있다.[73] '-인'은 ④ '그러한 집단에 속한 사람'이라는 특수한 의미도 파생할 수 있는 것으로 보인다. 의미별로 각 접사가 파생한 단어의 목록은 다음과 같다.

| '-가' | 고빈도어 | 저빈도어 | 신어 |
|---|---|---|---|
| 의미 ① | 활동가 | 감상가 | - |
| 의미 ② | 안무가, 비평가, 조각가, 만화가, 사진가, 정치가, 분석가, 운동가, 투자가, 연구가, 사업가, 기업가, 작곡가, 예술가, 평론가, 건축가, 소설가, 화가, 작가 | 경영가, 기고가, 등반가, 선동가, 연주가, 문학가, 여행가, 수장가, 웅변가 | 교습가, 반역가, 인권가, 탐서가 |
| 의미 ③ | 애호가, 전문가 | 재력가, 이상가, 야심가, 책략가, 지략가, 달변가, 애주가, 재산가, 명망가, 호사가 | 야망가 |
| 의미 ④ | - | | |

표 4.2. '가'의 의미별 파생어 목록

72) '-가, -자' 파생어들은 [+직업성]에서 기준을 조금 완화하여 [+대가성]과 같은 자질을 제하고 남은 [+반복성]만을 통하면 의미 칸에 분류되지 못한 잉여 항목이 없어지므로 이를 활용하였다.

73) 본 논의에서는 경쟁 접사 사이의 의미 차이만을 따지면 충분하기 때문에 의미를 최대한 간추린 것이다.

| ‘-자’ | 고빈도어 | 저빈도어 | 신어 |
|---|---|---|---|
| 의미 ① | 탈북자, 가입자, 투자자, 지도자, 소비자, 근로자 | 방청자, 내담자, 적대자, 운반자, 복직자, 보고자, 조언자, 상근자, 교신자, 작사자, 초대자 | 개통자, 단식자, 모의자, 상환자, 시구자, 약정자, 운송자, 운행자, 시상자, 종결자, 포획자, 주력자, 재범자, 채용자 |
| 의미 ② | 사업자, 노동자 | - | - |
| 의미 ③ | 당사자, 대상자, 당국자, 유권자, 피해자, 후보자, 관계자 | 계급자, 이중인격자, 미만자, 유발자, 상속자, 치료자, 판단자, 지위자, 예비자 | 위험자, 결함자, 내사자, 면책자, 전형자, 제한자, 산화자, 협론자, 감면자, 면세자, 신규자, 영주자, 추정자, 파병자, 휴가자 |
| 의미 ④ | - | | |

표 4.3. ‘자’의 의미별 파생어 목록

| ‘-인’ | 고빈도어 | 저빈도어 | 신어 |
|---|---|---|---|
| 의미 ① | 대변인 | 기소인, 탈북인, 출가인, 대행인, 진정인, 공증인, 발행인 | 해국인, 감례인, 중계인, 계량인, 측량인, 하역인 |
| 의미 ② | 시인, 정치인, 연예인 | 소작인, 예능인 | 공무인 |
| 의미 ③ | 장애인, 당선인, 직장인, 민간인, 일반인 | 동시대인, 르네상스인, 본토인, 그리스도교인 | 미납인, 오지인, 도외인 |
| 의미 ④ | - | 무슬림인, 바둑인, 이화인, 탁구인, 희극인, 배구인, 무용인 | 한수원인, 뮤지컬인, 성균인, 합창인, 검도인, 수협인, CJ인, KAIST인, 한센인 |

표 4.4. ‘인’의 의미별 파생어 목록

세 접사의 파생어 목록을 통해 각 접사의 의미영역이 겹치지만 전형적인 의미, 즉 고빈도어에서 주로 사용되는 의미가 다르고 신어 파생에서 주로 담당하는 의미는 그것과는 또 다르다는 것을 알 수 있다. ‘-가’는 의미 2가 전형적이고 신어 파생도 주로 그 의미영역 내에서 이루어진다. ‘-자’는 나머지 의미 1, 3의 파생을 담당하고 있다. 따라서 ‘-인’은 전형적 의미가 1~3에 해당하는데도 신어 파생에서는 특수 의미인 의미 4를 주로 담당하고 있다는 것을 알 수 있다.[74)]

전형적인 의미가 다르다는 것은 접사에 그러한 의미가 전혀 없다는 것이 아니라 언중들의 언어 사용 습관에 따라 주로 어떤 접사를 어떠한 의미로 사용하는 경향이 강하다는 뜻이다. 의미론적 제약, 음운론적 제약과 같은 언어 내적인 제약이 없으며 의미의 중화까지 가능한데도 '단지 그렇게 사용해왔기 때문에 그러한' 것이므로 이 역시 약한 성격의 저지로 볼 수는 있을 듯하다.

이처럼 접사의 전형적인 용법이 상당히 다르고, 그러한 차이가 저빈도어와 신어에서 그대로 유지되는 것으로 보이므로 신어 파생에서도 저지의 성격이 약하다고 볼 수 있을 것이다. 그러나 본 논의에서는 두 가지 측면에서 이러한 역할 분담을 유형 저지의 원인으로 본다. 첫째, 두 접사 '-가, -자'가 의미 ①~③을 모두 가지는데도 다른 접사의 존재로 인해 특정 의미를 자유롭게 파생할 수 없게 된 것이므로 그 자체로 고전적인 의미에서의 저지에 해당된다. 이는 '-인'의 파생어 목록을 추가해 보면 더 잘 드러난다. 이 접사는 의미 ①~③을 모두 파생할 수 있고 전형적인 용법도 이 세 가지 의미 안에 머무르지만, 저빈도어에서 신어 파생으로 갈수록 자신만이 추가할 수 있는 의미 ④에 더욱 집중하고 있다. 이는 어떤 단어를 만들어 쓰는 상황이 오면 접사의 선택을 관성에 따랐기 때문인 듯하다. 둘째, 위와 같은 분석에 따르면 분업에 따라 전체적인 세력이나 빈도 분포에 비해 상대적으로 신어를 많이 파생하는 의미를 담당한 접사는 생산적인 쪽에, 그렇지 않은 의미를 담당한 접사는 비생산적인 쪽에 머무르게 된 셈이다. 경쟁에 따라 신어 파생에서의 분업이 발생하고, 분업으로 인해 생산성에 차이가 나게 되었으므로 본 논의에서 상대적 생산성을 통해 정의한 유형 저지의 측면에서도 저지로 볼 수 있다. 다음으로 살펴볼 나머지 경쟁 집단에서도 비슷한 형태의 간섭이 나타난다.

한편, 이러한 해석에 기반한다면 접사의 의미를 세밀한 구분이 본 논의에서 큰 가치가 없는 이유도 자연스레 이해할 수 있다. 세밀하게 구분할수록 접사 간에 겹치는 의미는 적어지겠지만 그 중 일부만 중복되더라도 경쟁에 따른 분업에 의해 생산성 격차가 발생하였다는 논리는 성립할 수 있기 때문이다.

---

74) '-인'의 신어 중 '감례인, 계량인, 측량인, 하역인, 중계인' 등은 모두 법조문을 인용한 두 문장에서만 나타난 것이므로 '-인'의 신어 파생의 전형적인 사례라고 볼 수 없다. 그러나 일관된 기준에 따라 자료를 처리해야 하므로 신어로 분류하였다.

2) '-이, -음, -기'

세 접사 중 둘 이상이 파생 가능한 의미는 크게 ① '그러한 행위', ② '①에 해당하지 않는 단순 명사 파생', ③ '척도 명사 파생'('-이', '-기'), ④ '그러한 행위의 결과물'이 있다. '-이'는 ⑤ '그것을 (직업적으로 또는 직업처럼 잘) 하는 사람' 또는 '그렇게 사용되는 도구' 등의 특수한 의미를 더할 수 있다. '-기'는 특별한 용법이 없다.

①과 ②의 차이는 의미적으로 아주 미미하지만 차례로 '-기'와 '-음'이 담당하고 있으므로 그들의 역할 분담을 확실히 나타내기 위해 분리하였다. 일반적인 논의에서라면 ⑤의 '사람'과 '도구'는 분리해야 바람직하겠으나 둘 다 '-이'가 전담하는 특수 의미로 분류할 수 있으므로 굳이 분리하지 않았다.

| '-이' | 고빈도어 | 저빈도어 | 신어 |
|---|---|---|---|
| 의미 ① | 몰이, 맞이, 나들이, 살이, 물갈이, 되풀이, 벌이, 놀이, 풀이 | 멱살잡이, 떨이, 얽이, 드잡이, 마수걸이, 바릇잡이 | - |
| 의미 ② | - | - | - |
| 의미 ③ | 넓이, 길이, 높이, 깊이 | 음높이 | - |
| 의미 ④ | 구이 | - | |
| 의미 ⑤ | 먹이, 걸이 | 턱받이, 재떨이, 다듬이, 반닫이, 살붙이, 왼발잡이, 구두닦이, 넝마주이, 총알받이 | 떼잡이, 통돌이 |

표 4.5. '이'의 의미별 파생어 목록

| ‘-음’ | 고빈도어 | 저빈도어 | 신어 |
|---|---|---|---|
| 의미 ① | 물음, 탈바꿈, 걸음, 믿음, 만남, 맞춤, 싸움, 잠, 죽음 ,흐름, 움직임, 모임, 꿈, 느낌 | 높임, 앞가림, 뒷받침, 뒷막음, 구김, 단잠 | 태움 |
| 의미 ② | - | 노여움, 간지러움, 설레임 | - |
| 의미 ③ | - | | |
| 의미 ④ | 구이, 그림 | 튀김, 무침, 볶음, 찜 | - |
| 의미 ⑤ | - | | |

표 4.6. ‘음’의 의미별 파생어 목록

| ‘-기’ | 고빈도어 | 저빈도어 | 신어 |
|---|---|---|---|
| 의미 ① | 듣기, 날치기, 뛰기, 차기, 버티기, 부풀리기, 가로채기, 물타기, 걷기, 뒤집기, 힘겨루기, 끝내기, 몰아주기, 쓰기, 달리기, 보기 | 굽기, 곱하기, 돌려짓기, 상모돌리기, 뒤치기, 메치기, 곁누르기, 잡기, 딱지치기, 술래잡기, 구슬치기, 도움닫기, 돌림감기, 당기기, 보물찾기, 즐겨찾기 | 물빼기, 대못박기, 밑장빼기 |
| 의미 ② | - | | |
| 의미 ③ | 굵기, 크기, 세기 | 빠르기, 기울기, 거칠기 | - |
| 의미 ④ | - | | |
| 의미 ⑤ | - | | |

표 4.7. ‘기’의 의미별 파생어 목록

세 접사 모두 전형적인 용법에서 ‘그러한 행위’의 의미를 파생한다. 즉 언어 사용자가 인식하는 이 명사 파생 접미사들의 기본 의미는 의미 ①에 가깝다고 할 수 있다. 하지만 그러한 파생 의미는 저빈도어와 신어에서 완전히 달라지게 된다. 신어 파생에 있어 ‘-이’는 자신의 특수 의미인 의미 ⑤를, ‘-기’는 경쟁 계열 전체의 기본 의미인 ①을 담당하게 되었다. ‘-기’가 다른 접사들의 생산성을 압도한 이유가 여기에 있다고 볼 수 있다. 특히 ‘-음’은 넓은 의미에서는 ①에 해당하는 의미를 파생할 수 있으나 저빈도어나 신어에서까지도 단순한 ‘그러한 행위’의 의미가 아니라 비유적이거나 특수한 의미를 더하는 역

할에 국한되었다. 이러한 의미의 특수화는 어떤 접사에서든 고빈도 파생어에서는 흔히 나타나지만 저빈도어나 신어에서도 유지되는 경우는 드물다. '-음'이 유독 낮은 생산성을 나타냈던 이유는 여기에 있는 듯하다.

주어진 코퍼스에서 신어로 나타나지 않은 의미들은 새로운 단어를 만들 수요가 적은 것들이다. 여기에는 흔히 쓰이긴 하지만 새로운 어휘의 편입이 매우 드문 경우도 있고 의미의 수요 자체가 적은 경우도 있다. 의미 ③의 척도명사는 전자에 해당하고 의미 ②는 후자에 해당한다.

3) '-스럽-, -하-, -롭-, -지-, -적(的)'

이들 접사는 모두 ① '그러한 성질이 있음(단순 형용사 파생)'의 의미 기능을 가진다. 일부 파생어에서는 ② '어근이 가질 것으로 예상되는 성질이 있음'이라는 의미를 더하기도 한다. 이 의미는 '-답-'의 그것과 비슷하지만, '-답-'은 실제로 그 어근에 해당되는 대상이 그러한 성질도 가지고 있음을 의미하는 반면 '-스럽-' 계열은 어근과 대상이 반드시 같은 집합 내에 있을 것을 전제하지는 않는다는 차이가 있다(송철의 1992). 또, '-스럽-'은 3장에서 서술하였듯 ③ '그것을 대하는 주체로 하여금 ~한 상태가 되도록 하는 속성이 있음'이라는 특수한 의미를 더할 수 있다. 이 의미는 다른 맥락에서는 의미 ①과 구분할 필요가 거의 없으나 '-하-'의 대조에서 그 차이가 명확히 드러나므로 여기에서는 구분하기로 한다.

한편, '-롭-, -지-'는 이미 사실상 거의 사용되지 않아 사멸 단계에 접어들었고 '-하-, -적'은 공시적으로 대단히 높은 결합력을 가지고 있으므로 둘을 분리하여 다룰 필요가 있다. 주어진 코퍼스에서 '-롭-', '-지-'의 유형 빈도는 각각 36, 17로, 나타난 파생어 대부분이 저빈도와 고빈도를 막론하고 해당 접사의 전형적인 용법에 편입되어 있기 때문이다.[75] '-스럽-'은 가장 다양한 의미를 파생하고 제약도 적어서 경쟁 계열 내의 다른 모든 접사와 신어 파생에서 경쟁하는 모습을 보인다.

---

75) 그럼에도 불구하고 일부 항목에서는 불완전한 저지가 일어나기도 한다. 이는 4.3에서 다루었다.

| '-스럽-' | 고빈도어 | 저빈도어 | 신어 |
|---|---|---|---|
| 의미 ① | 고급스럽, 부담스럽, 자연스럽, 자랑스럽, 조심스럽 | 호사스럽, 관용스럽, 구차스럽, 낭패스럽, 명예스럽, 밉상스럽, 바보스럽, 불량스럽, 불상스럽, 소란스럽, 외설스럽, 우연스럽 | 게이스럽, 시원시원스럽, 얌채스럽, 양념스럽, 유독스럽, 퇴행스럽, 갑스럽 |
| 의미 ② | - | 어른스럽 | 대륙낭인스럽, 케이블스럽, 현대카드스럽, 자미스럽, 하얀스럽, 한국스럽 |
| 의미 ③ | 당혹스럽, 당황스럽, 우려스럽, 걱정스럽, 의심스럽, 만족스럽, 곤혹스럽, 고통스럽, 혼란스럽 | 절망스럽, 감동스럽, 감탄스럽, 경악스럽, 보람스럽, 불안스럽, 염려스럽 | 낙망스럽, 분노스럽 |

표 4.8. '스럽'의 의미별 파생어 목록

| '-적' | 고빈도어 | 저빈도어 | 신어 |
|---|---|---|---|
| 의미 ① | 근본적, 부정적, 실질적, 긍정적, 상대적, 지속적, 경제적, 법적, 기본적, 개인적, 본격적, 세계적, 대표적, 적극적, 구체적, 사회적, 정치적 | 가격적, 가변적, 결합적, 경고적, 관료주의적, 교파적, 관조적, 다자적, 단절적, 다문화적, 대폭적, 독설적, 동물적, 마케팅적, 모험적, 반자동적, 실존적, 양면적 | 개간적, 비공정적, 구단적, 국외부적, 권리적, 남미적, 뉴스적, 메타텍스트적, 성희롱적, 혼적, 주위적, 공영적, 정반합적, 작정적, 원칙법적(등 70항목, 74회 출현) |
| 의미 ② | - | 강원도적, 만화적 | 아일랜드적, 록(rock)적, 헤라클레이토스적, 외교관적, 뉴라이트적, 엔엘피적, 한식적, 트로트적, 차베스적, 사이코패스적, 서예사적, 남미적, 단테적 |

표 4.9. '적'의 의미별 파생어 목록

| '-하-' | 고빈도어 | 저빈도어 | 신어 |
|---|---|---|---|
| 의미 ① | 불안하, 치열하, 불편하, 건강하, 유리하, 행복하, 특별하, 정확하, 유일하, 이상하, 강력하, 유명하, 안녕하, 불과하, 부족하, 다양하, 가능하, 중요하, 필요하 | 무가치하, 강령하, 꿀꿀하, 드라이하, 랜덤하, 부스스하, 순결하, 숭악하, 스무드하, 스페셜하, 실하, 악독하, 우뚝하, 에스닉하, 우량하, 포멀하, | 감성감성하, 그루브하, 꺼묵꺼묵하, 꼬롬꼬롬하, 네셔널하, 다기능하, 따들따들하, 러블리하, 로얄하, 탱글하(등 49항목, 54회 출현) |
| 의미 ② | - | | |

표 4.10. '하'의 의미별 파생어 목록

'-적'과 '-하-'는 같은 어기에 대해 대체 관계에 있지 않으므로 서로 직접 대립한다기보다는 각자가 '-스럽-'과 일대일 경쟁을 하는 것으로 보아야 한다. 큰 틀에서 보면 '-스럽-'은 신어 파생에서 '-적'에 비해 의미 ②에 특화되어 있다. '-스럽-'의 전형적인 의미는 ③에 가깝고 저빈도 구간에서도 이 의미로 많이 사용되지만 신어에서는 이 의미가 상대적으로 적은 비율을 차지한다. '-적'은 '-하-'와 함께 의미 ①의 신어 파생을 전담하고 있다. 그런데 '-스럽-'과 달리 고유어 어근을 거의 취하지 못하는 약한 제약을 가진 듯하다. '보람적'과 같은 고유어 어근 파생어가 존재하긴 하지만 이는 직관적으로나 통계적으로나 비전형적인 용법으로 판단되며, 대다수의 파생어가 한자어 혹은 외래어만을 어근으로 삼는다. '-하-'는 '-스럽-'과 의미적 측면에서 구별된다. '-하-' 파생어는 어근에 관계 없이 대상의 상태를 나타내는 의미로만 나타나지만 '-스럽-' 파생어는 의미 ③으로도 사용되는 것이다. '-스럽-, -하-, -적'은 이처럼 접사 사이의 의미적 분업과 제약이 골고루 영향을 미쳐 생산성에 차이를 보인 것으로 해석할 수 있다.

'-지-'와 '-롭-'은 공히 의미 ①만을 파생할 수 있다. 그런데 이들은 이미 파생어 전체에 대한 수요가 줄어들었기 때문에 '-스럽-' 혹은 '-하-', '-적'과 온전히 경쟁한다고 보기는 어렵다.[76] 그러나 저지의 정의에 따르면 여전히 생산성을 유지하고 있고 넓은 의미의 동의성 조건도 만족하므로 유형 저지 묶음에 포함하였다.

76) 항목 저지에서도 다른 접사 파생어들이 이들의 파생 영역을 심각하게 침범하는 모습을 확인할 수 있다.

4) 나머지 경쟁 접사들

나머지 접사는 경쟁하는 계열 내의 하나 이상의 접사가 생산성이 없는 것으로 나타난 것들이다. 이들은 생산성 조건을 만족하지 않는 것처럼 보이지만 코퍼스를 통해 측정된 생산성이 실제 생산성을 의미하지는 않으므로 이들의 생산성 격차도 저지의 결과로 보고자 한다. 경쟁 접사의 확연한 쇠퇴와 함께 그들의 신어 파생에서의 의미영역을 어느 한 접사가 흡수하고 있다고 해석할 수 있다.

'-꾼, -쟁이, -장이'는 '그것을 (버릇처럼·직업적으로) 하는 사람', '그것이나 그것의 특성을 많이 가진 사람'이라는 의미영역을 나누어 가진다. 이 중 두 의미를 골고루 더할 수 있는 '-꾼'이 경쟁에서 우위를 차지하였음을 앞서 확인한 바 있다.

'비(非)-, 불(不)-'에서는 '비-'가 경쟁에서 우위를 점하고 있다. 전술하였듯 '비-'는 대부분의 어근에 단순한 부정의 의미만을 더하는 반면 '불-'은 파생의 결과 정도 반의어가 되는 경우가 상대적으로 훨씬 많은데, 이 의미 차이가 신어 파생 능력의 격차를 야기한 것으로 보인다.

'-거리-, -대-, -이-'는 특수한 경우이다. 이들은 의미가 거의 같으며 대부분의 어근에서 교체가 가능하다. '-거리-, -대-'에서는 '-이-'와 달리 접사와 어근이 분리되는 성질이 강한데, 이것이 이들의 분포 형태를 가른 결정적인 특징인 듯하다.[77]

## 4.3. 항목 저지

저지에 관한 선행 연구에서는 모두 항목 저지(token blocking)를 다루었다. 따라서 항목 저지는 그 특성이 알려진 바가 많다. 가장 널리 알려진 항목 저지의 특징은 저지력이 일반적으로 빈도에 비례하는 저장 강도에 의해 결정된다는 점이다.[78] 의미가 아주 비슷하고 결합이 가능하더라도 접사의 배타적 결

77) '-거리-'와 '-대-'에서는 분포의 뚜렷한 단절점(세타)을 찾을 수 없었다.
78) 저지력을 빈도의 함수로 이해할 수 있다는 논의는 1.2에서 간단히 다루었다.

합은 기본적으로 지켜지지만, 어근을 공유하는 경우에도 빈도가 높은 파생어일수록 저지가 압도적으로 이루어지는 경우가 대부분이다. 특히 본 연구의 조사 결과 저장 임계 빈도를 넘은 파생어는 거의 절대적인 저지력을 보여주었다.

다른 파생어의 출현을 막는 것뿐 아니라 저지[79]를 거스르는 데에도 빈도가 큰 영향을 미치는 것으로 보인다. 접사의 빈도가 매우 높은 경우에는 항목에 따라 기본적인 저지의 방향을 무시하고 고빈도로 나타나기도 한다. 접사 자체는 빈도가 높지 않더라도 개별 항목의 빈도가 높다면 저지를 무시할 수 있다. 빈도에 의한 저지 방해에 대해서는 논의된 바가 없는데 이런 현상 역시 빈도에 좌우되는 항목 저지의 일반적인 경향에서 벗어나는 것은 아니다.

항목 저지는 경쟁하는 접사끼리 짝을 지어서 비교할 것이다. 항목 저지에서 직접 저지에 참여하는 것은 물론 각 항목들이지만, 이는 접사들이 여러 항목에 걸쳐 파생하고자 경쟁하는 것으로도 해석할 수 있기 때문이다(Gardani, F. et al. 2019). 각 접사는 마치 고유의 파생 영역을 가진 것처럼 행동한다. 대부분 항목에서 경쟁 접사의 파생을 막고 일부 항목에서만 파생을 허용하는 것이다. 이런 관점에서 볼 때 상대 접사의 파생을 허용한 항목은 접사 간 경쟁의 최전선에 있다고 비유할 수 있다.

다음으로, 항목 저지를 다룰 때에는 두 접사의 결합형이 모두 출현하는 항목을 위주로 비교할 것이다. 두 접사의 결합형이 중복되지 않는 항목은 단순히 경쟁 항목이 먼저 존재했기 때문에 출현하지 못했거나 아예 경쟁이 성립하지 않는 경우이므로 설명의 여지가 별로 없다. 게다가 빈도에 의한 저지와 단순한 존재에 의한 저지를 구분할 기준을 정하기도 매우 어렵다.[80]

4.3.1에서는 일반적인 유형의 항목 저지를 살펴본다. 일반적인 저지란 빈도

---

79) 기본적인 저지의 방향은 4.2에서 살펴본 유형 저지의 방향으로 쉽게 파악할 수 있다. 신어 생산 경쟁에서 승리한 접사는 경쟁 접사 집단 중 의미가 제일 무표적이고 결합에 제약도 적은 쪽에 속하기 때문에 항목 저지에서도 저지어를 파생하는 접사가 된다.

80) 피저지어의 빈도가 0이고 저지어의 빈도가 n이라 하자. 가령 n이 100정도로 크다면 둘 사이의 저지를 빈도에 의한 것으로 설명하기 쉽고 n=1에서는 존재에 의한 저지로 설명하기 편하겠지만, 두 저지가 만나는 빈도가 어디일지는 알 수 없다. 사실 1회 출현하면서도 완전한 저지가 가능한 파생어는 현재의 빈도가 낮을 뿐 예전에는 높은 빈도로 사용되어 어휘부에 확고하게 저장된 항목일 가능성이 높다. 따라서 본 연구에서는 두 현상을 본질적으로 같은 것으로 보고 4.3.1.1에서 함께 다룬다.

와 그에 따른 어휘부 저장 강도에 의해 설명 가능한 저지를 의미한다. 4.3.2에서는 한국어 항목 저지의 예외인 '거리-대' 및 '스럽-적-하'의 경쟁을 다룰 것이다.

### 4.3.1. 일반적인 항목 저지

일반적인 항목 저지는 빈도에 의한 저지를 의미한다. 빈도에 의한 저지란 빈도가 높은 저지 항목의 어휘부에서의 저장 강도가 강해서 저지력이 확보되는 저지를 의미한다. 코퍼스에 나타난 빈도가 높지 않더라도 완전한 저지를 보여주는 항목은 기존의 높은 빈도에 의해 이미 어휘부에 깊이 저장되었을 가능성이 높다. 이런 관점에서는 항목 저지의 대부분이 빈도에 의한 저지에 해당하게 된다. 본 연구에서 빈도에 의한 저지로 분류한 관계는 '-스럽-'-'-하-', '-스럽-'-'-적(的)', '-거리-'-'-이-', '-대-'-'-이-'를 제외한 모든 경쟁 접사 사이의 저지이다. 4.3.1.1에서는 이 중 다룰 만한 특징이 있는 일부 저지쌍들을 대상으로 분석을 진행하고자 한다.

'-가'-'-자'의 저지는 대부분 피저지어의 출현을 완전히 막는 형태로 나타난다. 두 접사의 배타적 출현은 부분적으로는 접사의 의미 차이 때문이다.[81] '-가'와 '-자'는 파생했을 때 '~하는 사람' 또는 '~의 특성이 있는 사람'과 같이 '사람'에 최소한의 의미만을 더하는 어근과는 공통적으로 결합할 수 있다. 하지만 '-가'는 좀 더 구체적인 의미를 추가할 수 있다. 예를 들어, '여행가', '자문가'와 같은 파생어의 '-가'에는 '~를 직업으로 하는 사람' 혹은 '~를 직업처럼 자주·반복적으로 하는 사람'이라는 의미가 있다. 이러한 파생어의 어근은 '-자'와 결합하지 않거나, 결합하더라도 의미가 달라진다.

그런데 '-가'와 '-자'가 모두 결합할 수 있을 법한 어근에서도 두 접사는 거의 배타적으로 결합한다. 앞서 제시한 표 4.2에서 이를 확인할 수 있다. 이 목록은 '-가'의 파생어 중 의미상 직업으로 삼거나 꾸준히 하는 것이 불가능한 어근의 결합 양상이다. 이 어근들은 이론적으로는 '-자'와 결합하는 데 아무

81) 전술하였듯 경쟁하는 접사 사이에 의미 차이가 있다는 것은 접사들의 의미영역이 완전히 일치하지 않는다는 뜻이지 아예 전반적으로 다르다는 뜻이 아니다. 거의 의미의 차이를 느낄 수 없는 항목이 자주 나타나기 때문에 두 접사를 경쟁한다

런 문제가 없지만 ‘-가’와 결합하여 거의 모든 항목에서 경쟁 파생어를 완전히 저지하였다.

한편, ‘-가’, ‘-자’의 파생 영역이 이처럼 확고한데도 ‘-자’는 ‘전문’과 같은 전형적인 ‘-가’ 파생어 어근과 결합하는 데 성공했다는 점이 눈에 띈다. 유형 저지에서 열세에 있던 접사가 항목 저지에서는 오히려 더 강한 세력을 갖는 것처럼 보이는 이런 현상은 다른 접사 간 경쟁에서도 어렵지 않게 찾아볼 수 있다.

| | -가 | -자 |
|---|---|---|
| 이상(理想) | 1 | 0 |
| 야심 | 1 | 0 |
| 야망 | 1 | 0 |
| 책략 | 1 | 0 |
| 지략 | 1 | 0 |
| 반역 | 1 | 0 |
| 달변 | 1 | 0 |
| 애주 | 1 | 0 |
| 재산 | 1 | 0 |
| 명망 | 2 | 0 |
| 호사 | 2 | 0 |
| 애연 | 2 | 0 |
| 독설 | 2 | 0 |
| 독지 | 3 | 0 |
| 애호 | 20 | 0 |
| 전문 | 1152 | 1 |

표 4.11. ‘-가’-‘-자’의 경쟁

| | -스럽- | -롭- |
|---|---|---|
| 명예 | 1 | 15 |
| 호화 | 1 | 3 |
| 자유 | 2 | 218 |
| 신비 | 3 | 9 |
| | -스럽- | -지- |
| 고급 | 28 | 1 |

표 4.12. ‘-스럽-’-‘-롭-’, ‘-스럽-’-‘-지-’의 경쟁

표 4.12의 ‘-스럽-’-‘-지-’의 저지 양상에서도 유형 저지에서 ‘-스럽-’에 비해 심각하게 위축되었던 ‘-지-’가 오히려 ‘-스럽-’의 파생 영역을 위협하는 듯한 모습을 확인할 수 있다. 전형적인 ‘-스럽-’ 파생어인 ‘고급스럽’을 ‘-지-’가 파생한 것이다. 그러나 ‘고급지’는 순간적인 필요에 의해 생성된 단어가 아니라 최근 들어 크게 유행한 단어로, 기존 코퍼스에 나타나지는 않았지만 짧은 기간 동안 특별히 빠른 속도로 어휘부에 저장되었다고 보아야 할 것이다. 즉, 신조어이지만 신어는 아닌 것이다. 따라서 ‘-지-’가 ‘-스럽-’의 파생 영역을 실질적으로 침범하고 있다고 보기는 어렵다.

‘-스럽-’-‘-롭-’ 경쟁에서는 전형적인 ‘-롭-’ 파생 어근들에 ‘-스럽-’이

결합하고 있다. 이 중 '신비'는 특이한 경우인데, 같은 의미로 사용된 네 개의 접사 파생형이 동시에 관찰되었기 때문이다. '−롭−'은 전반적으로 저지력을 잃어가고 있는 것으로 보인다.

| | -스럽- | -롭- | -적(的) | -하- |
|---|---|---|---|---|
| 신비 | 3 | 9 | 1 | 4 |

표 4.13. '신비' 파생 경쟁

한편, 전술하였듯 '−스럽−'은 의미에 따라 '−스럽$^1$−'과 '−스럽$^2$−'로 나눌 수 있는데, 각각이 다른 접사와 경쟁한다. '−스럽$^1$−'은 '~에 가까운 성질이 있다', '−스럽$^2$−'은 '그것을 대하는 주체로 하여금 ~한 상태가 되도록 하는 성질이 있다' 정도의 의미이다. 표 4.5, 4.6에는 '−스럽$^1$−'과 '−스럽$^2$−'의 파생어 목록을 정리하였다.[82] 파생어 목록을 통해 '−스럽$^1$−'은 '−롭−', '−지−', '−하−', '−적'과 경쟁하고 '−스럽$^2$−'는 '−적'과 경쟁한다는 것을 알 수 있다. '−스럽−'과 '−하−, −적' 사이의 저지는 '−스럽−'의 의미에 따라 나누어 분석할 것이다. 이들의 저지 양상은 다른 접사들과 조금 다르게 나타나므로 4.3.1.2에서 따로 다루기로 한다.

| | | | |
|---|---|---|---|
| 호사스럽 | 유독스럽 | 자유스럽 | 조심스럽 |
| '대륙낭인'스럽 | 자미(滋味)스럽 | 통명스럽 | 자랑스럽 |
| 케이블스럽 | 존엄스럽 | 고집스럽 | |
| 현대카드스럽 | 충성스럽 | 까탈스럽 | |
| 게이스럽 | 퇴행스럽 | 능청스럽 | |
| 관용스럽 | 패악스럽 | 비밀스럽 | |
| 구차스럽 | 하얀스럽 | 신비스럽 | |
| 낭패스럽 | 한국스럽 | 억지스럽 | |
| 명예스럽 | 허접스럽 | 억척스럽 | |
| 밉상스럽 | 호들갑스럽 | 유난스럽 | |
| 바보스럽 | 호화스럽 | 장난스럽 | |
| 불량스럽 | 효성스럽 | 짜증스럽 | |
| 불상스럽 | 갑스럽 | 탐욕스럽 | |
| 소란스럽 | 저질스럽 | 고생스럽 | |
| 시원시원스럽 | 가증스럽 | 고풍스럽 | |

| | |
|---|---|
| 절망스럽 | 당혹스럽 |
| 감동스럽 | 당황스럽 |
| 감탄스럽 | 우려스럽 |
| 경악스럽 | 걱정스럽 |
| 낙망스럽 | 의심스럽 |
| 보람스럽 | 만족스럽 |
| 분노스럽 | 곤혹스럽 |
| 불안스럽 | 고통스럽 |
| 염려스럽 | 혼란스럽 |
| 의문스럽 | |
| 의아스럽 | |
| 챙피스럽 | |
| 치욕스럽 | |
| 통탄스럽 | |
| 원망스럽 | |

82) '-스럽-'의 하위 의미에 따른 접사의 빈도 분포적 특성에 대해서는 3.1에서 분석하였다. 이밖에 4장에서 다루는 접사의 분포나 빈도에 관한 모든 내용은 3장에서 이미 다룬 것이다.

| | | |
|---|---|---|
| 얌채스럽 | 맛깔스럽 | 수고스럽 |
| 양념스럽 | 사치스럽 | 탐스럽 |
| 외설스럽 | 어른스럽 | 성(聖)스럽 |
| 우연스럽 | 예스럽 | 변덕스럽 |
| 위험스럽 | 용맹스럽 | 익살스럽 |
| 여성스럽 | 영광스럽 | 부담스럽 |
| 멋스럽 | 촌스럽 | 자연스럽 |
| 불만스럽 | 불미스럽 | 남사스럽 |
| 사랑스럽 | 다행스럽 | 한스럽 |
| 정성스럽 | 고급스럽 | 새삼스럽 |

표 4.14. '-스럽[1]-'의 파생어

| |
|---|
| 혐오스럽 |
| 후회스럽 |
| 개탄스럽 |
| 공포스럽 |
| 존경스럽 |
| 고민스럽 |
| 감격스럽 |
| 수치스럽 |
| 실망스럽 |
| 유감스럽 |

표 4.15. '-스럽[2]-'의 파생어

'-꾼'-'-쟁이'-'-장이'는 완전히 배타적으로 출현한다. 이들 사이의 신어 파생 경쟁은 이미 '-꾼'이 압도하고 있으며 이는 앞서 매우 위축된 '-쟁이, -장이'의 세력을 통해 확인한 바 있다. 이런 경향은 항목 저지에서도 확인할 수 있다. 표 4.7은 코퍼스에 나타난 '-쟁이' 파생어의 과거와 현재 빈도이다. 이를 통해 마치 유형 저지인 듯해 보였던 '-쟁이'-'-꾼'의 관계가 사실은 항목 저지에 해당한다는 것을 알 수 있다. 파생어의 대부분은 과거에 훨씬 높은 빈도로 사용되다가 단어 자체가 쓰임이 줄어든 것이므로 파생어 각각의 어휘부 저장 강도가 높았을 것이기 때문이다. 파생어의 저장 강도와 저지력에 의한 저지는 항목 저지에 해당하는 현상이다. 완전 배타적으로 출현하는 대부분의 다른 저지 관계들도 이런 이유에서 항목 저지로 분류하게 된다.

| 파생어 | 이광호(09) | 본 연구 |
|---|---|---|
| 욕쟁이 | 4 | 1 |
| 뻥쟁이 | 0 | 1 |
| 약쟁이 | 0 | 1 |
| 참견쟁이 | 0 | 1 |
| 사치쟁이 | 0 | 1 |
| 배구쟁이 | 0 | 1 |
| 멋쟁이 | 12 | 2 |
| 글쟁이 | 5 | 2 |
| 빚쟁이 | 15 | 2 |
| 겁쟁이 | 15 | 3 |
| 점쟁이 | 21 | 4 |
| 욕심쟁이 | 4 | 4 |
| 거짓말 | 13 | 7 |
| 월급쟁이 | 21 | 9 |

표 4.16. '-쟁이' 파생어의 통시적 빈도 변화

‘-이’-‘-음’-‘-기’ 는 거의 배타적으로 출현한다. 코퍼스에 나타난 세 접사의 파생어 가운데 쌍형어로 나타난 것은 ‘놀이(빈도 117)’-‘놀음(빈도 2)’뿐이다. ‘놀음’은 보통 ‘놀이’와 다른 의미로 쓰이지만 여기서는 ‘야바위 놀음’, ‘혐한 놀음’과 같이 쓰여 ‘놀이’와 거의 동일한 의미를 보였다. 그런데 ‘-이, -음, -기’는 이처럼 서로 철저하게 저지하면서도 의미가 중화되는 어근은 많다는 특징이 있다. 세 접사는 신어 파생에서뿐 아니라 전반적인 파생어 사용에서 의미영역이 이미 거의 분화된 것으로 보인다. 즉, 이들은 이론적으로는 거의 같은 의미의 파생어를 만들 수 있지만, 실제 사용에서는 서로 다른 의미를 파생하는 어근만을 분리하여 담당한다. 파생을 막는 요인이 없는데도 경쟁하는 파생어의 존재만으로 자유롭게 파생하지 못하였으므로 전형적인 저지로 볼 수 있다.

| | 비 | 불 |
|---|---|---|
| 공정 | 1 | 81 |
| 활성 | 1 | 1 |

표 4.17. ‘비-’-‘불-’의 경쟁

‘비-’와 ‘불-’은 기본적인 의미가 다르지만 많은 어근에서 의미 차이가 중화될 수 있는데도 거의 배타적으로 출현한다. ‘불-’은 파생하면 정도 반의어(gradable antonym)가 되는 어근과 결합하는 경우도 많지만, 단순 부정이 곧 상보 반의어(complementary antonym)을 뜻하는 경우에는 ‘비-’와의 의미 차이가 중화된다. ‘비-’는 대부분의 경우 단순 부정의 의미만을 추가하기 때문이다. 가령, 자료에서 ‘불X’ 형태로만 나타난 ‘간섭, 확정, 기소, 출마, 구속, 투명, 가능’ 등은 ‘비-’로 파생하더라도 아무런 의미 차이가 없지만 ‘비X’ 형태를 완전히 저지하였다. 이를 통해 두 접사의 배타적인 분포가 ‘그렇게 사용하기 때문에 사용하는’, 저장에 의한 항목 저지임을 짐작할 수 있다.

### 4.3.2. 예외적인 항목 저지

#### 4.3.2.1. ‘-거리’, ‘-대-’, ‘-이-’

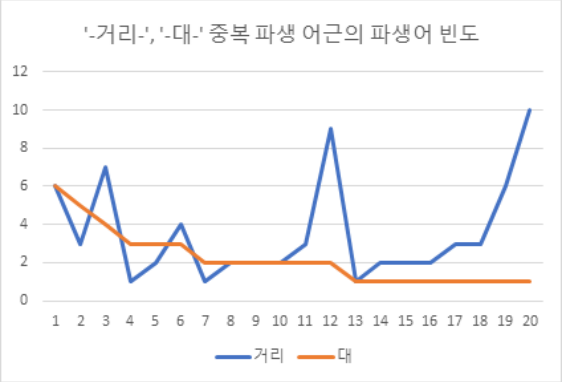
‘-거리-’와 ‘-대-’는 파생어의 의미가 거의 동일하지만 사실상 서로에 대해 아무런 저지를 하지 않는다. 표 4.18은 자료에 나타난 33가지 ‘-대-’ 파생어 전체의 어근 목록이다. 모든 어근이 ‘-스럽-’과 결합 가능하며 ‘-스럽-’으로 파생하더라도 의미 차이가 느껴지지 않는다.

| | | | |
|---|---|---|---|
| 북적북적 | 깔깔 | 넘실 | 키득 |
| 쩝쩝 | 주춤 | 빈정 | 휘청 |
| 우걱우걱 | 살랑 | 낄낄 | 들락 |
| 징징 | 삐걱 | 웅성 | 칭얼 |
| 으르렁 | 굽신 | 틱틱 | 끼룩 |
| 흥청 | 킁킁 | 바둥 | 투덜 |
| 비아냥 | 허우적 | 꿈틀 | 미적 |
| 허둥 | 수군 | 북적 | |

표 4.18. ‘-대-’ 파생어의 어근 목록

이 어근들 중 20가지는 실제로 자료에서 ‘-거리-’ 파생어로도 나타난다. 두 접사 간의 배타적 결합이 거의 무너진 가운데 ‘-대-’가 ‘-거리-’에 비해 상대적으로 저지를 당하는 편이라는 사실을 표 4.19와 그래프 4.4를 통해 알 수 있다.

| | 거리 | 대 |
|---|---|---|
| 꿈틀 | 6 | 6 |
| 수군 | 3 | 5 |
| 허우적 | 7 | 4 |
| 흥청 | 1 | 3 |
| 허둥 | 2 | 3 |
| 비아냥 | 4 | 3 |
| 바둥 | 1 | 2 |
| 투덜 | 2 | 2 |
| 북적 | 2 | 2 |
| 킁킁 | 2 | 2 |
| 깔깔 | 3 | 2 |
| 삐걱 | 9 | 2 |
| 넘실 | 1 | 1 |
| 낄낄 | 2 | 1 |
| 빈정 | 2 | 1 |




그래프 4.4. ‘-거리-’-‘-대-’의 경쟁

| | | |
|---|---|---|
| 웅성 | 2 | 1 |
| 미적 | 3 | 1 |
| 틱틱 | 3 | 1 |
| 들락 | 6 | 1 |
| 휘청 | 10 | 1 |

표 4.19. '-거리-'-'-대-'의 경쟁

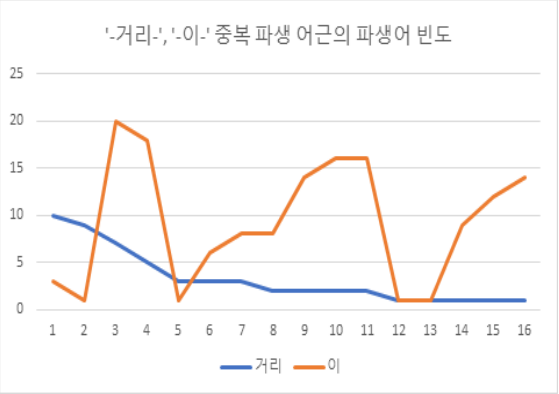
흥미로운 것은 역시 비슷한 의미를 파생하는 '-이-'와의 경쟁에서는 '-이-'가 두 접사의 파생어를 전반적으로 압도하는 모습을 보인다는 점이다. 표 4.20, 표 4.21, 그래프 4.5에서 이것을 확인할 수 있다.

| | 거리 | 이 |
|---|---|---|
| 휘청 | 10 | 3 |
| 삐걱 | 9 | 1 |
| 반짝 | 7 | 20 |
| 다독 | 5 | 18 |
| 절뚝 | 3 | 1 |
| 깜빡 | 3 | 6 |
| 출렁 | 3 | 8 |
| 서성 | 2 | 8 |
| 울먹 | 2 | 14 |
| 들썩 | 2 | 16 |
| 북적 | 2 | 16 |
| 간질 | 1 | 1 |
| 홀짝 | 1 | 1 |
| 들먹 | 1 | 9 |
| 끄덕 | 1 | 12 |
| 술렁 | 1 | 14 |

표 4.20. '-거리-'-'-이-'의 경쟁

| | 대 | 이 |
|---|---|---|
| 북적 | 2 | 16 |
| 휘청 | 1 | 3 |
| 삐걱 | 2 | 1 |
| 웅성 | 1 | 1 |

표 4.21. '-대-'-'-이-'의 경쟁



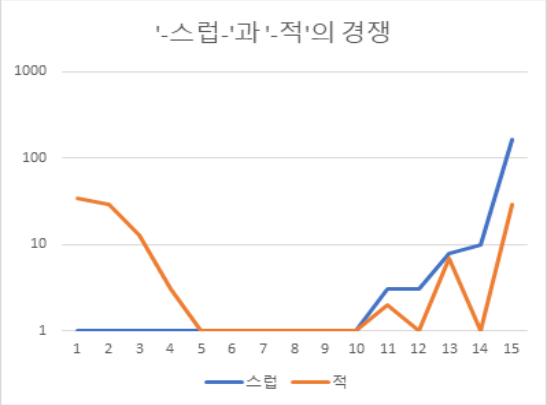

'-이-'가 다른 두 접사를 항목 저지에서 압도하는 것은 특이하지 않다. '-이-'는 생산성이 거의 없어서 P=0으로 조사되었는데, 이처럼 생산성이 낮은 접사의 파생어가 더 생산적인 접사의 파생어를 잘 저지하는 것은 일반적인 항목 저지의 경향과 일치하는 것이다. 그러나 같은 의미로 파생될 수 있으며 파생

가능한 어근 집합이 거의 동일한데도 경쟁하는 파생어가 아닌 접사에 따라 저지의 양상이 달라졌다는 점은 매우 독특하다. ‘-거리-’, ‘-대-’, ‘-이-’ 파생어 사이의 저지는 개별 항목이 아닌 접사의 특성이 저지의 양상을 좌우하므로 항목 저지의 예외가 된다.

#### 4.3.2.2. ‘-스럽-’, ‘-적(的)’, ‘-하-’

| | 스럽 | 적 |
|---|---|---|
| 한국 | 1 | 34 |
| 감동 | 1 | 29 |
| 퇴행 | 1 | 13 |
| 절망 | 1 | 3 |
| 보람 | 1 | 1 |
| 불량 | 1 | 1 |
| 외설 | 1 | 1 |
| 우연 | 1 | 1 |
| 치욕 | 1 | 1 |
| 저질 | 1 | 1 |
| 혐오 | 3 | 2 |
| 신비 | 3 | 1 |
| 여성 | 8 | 7 |
| 감격 | 10 | 1 |
| 자연 | 167 | 29 |

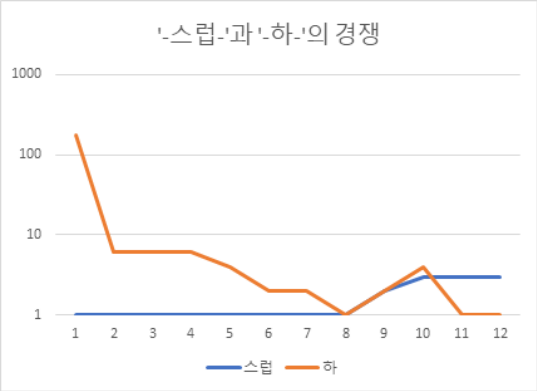
표 4.22. ‘-스럽-’-‘-적’의 경쟁




그래프 4.5. ‘-스럽-’-‘-적’의 경쟁

| | 스럽 | 하 |
|---|---|---|
| 위험 | 1 | 177 |
| 불량 | 1 | 6 |
| 시원시원 | 1 | 6 |
| 우연 | 1 | 6 |
| 구차 | 1 | 4 |
| 소란 | 1 | 2 |
| 존엄 | 1 | 2 |
| 허접 | 1 | 1 |
| 용맹 | 2 | 2 |
| 신비 | 3 | 4 |
| 퉁명 | 3 | 1 |
| 유난 | 3 | 1 |

표 4.23. ‘-스럽-’-‘-하-’의 경쟁




그래프 4.6. ‘-스럽-’-‘-하-’의 경쟁

표 4.22, 4.23, 그래프 4.5, 4.6에서 ‘-스럽-’과 ‘-하-, -적’의 파생어 간

저지가 접사에 따라 비대칭적으로 일어난다는 것을 알 수 있다. 경쟁하는 항목에서 '-하-'와 '-적'의 고빈도 파생어는 '-스럽-' 파생어를 완벽하게 저지하는 반면 '-스럽-'의 파생어는 거의 저지력이 없는 것처럼 보인다. 이들 사이의 저지도 항목별로 독립적인 저지가 이루어지지 않고 접사에 따라 저지의 양상이 확연히 달라지기 때문에 일반적인 항목 저지로 보기 어렵다. 이 접사들 사이에 비대칭적인 저지가 일어나는 원인으로 '-하-'와 '-적'이 아주 활발하게 사용 된다는 점, 즉 빈도가 매우 높다는 점을 생각해볼 수 있으나 이것 역시 접사에 의한 설명이므로 항목 저지에 대한 설명에 적합하지 않다. 따라서 '-스럽-, -적, -하-' 파생어 사이의 항목 저지 역시 예외적인 것으로 볼 수 있다.

# 5. 결론

본 연구에서는 한국어에서 나타나는 저지의 모습과 그러한 모습을 띠게 된 원리를 밝히는 데 집중하였다. 저지의 본질을 밝히기 위해 우선 저지와 밀접하게 관련된 수치인 저장 임계 빈도와 생산성을 접사별로 구했다. 한국어 접사의 고유한 저장 임계 빈도를 찾는 작업은 본 연구에서 처음으로 수행한 것인데, 이는 선행 연구를 참고하여 새로 고안한 계량언어학적 빈도 분석 방법에 힘입은 것이다. 이 빈도 분석 방법을 통해 선행 연구의 한계를 극복하고 언어학적으로 합당한 생산성을 계산할 수 있었다. 다음으로는 이렇게 구한 생산성과 저장 임계 빈도의 상대적 분포를 분석하여 접사들 간에 일종의 신어 생산 경쟁과 저지가 존재한다는 것을 밝힐 수 있었다. 이러한 저지현상은 이전에 알려지지 않았던 형태이지만, 유형 저지라는 기존의 개념에 잘 들어맞을 뿐 아니라 이 개념은 용어로 존재만 해왔고 실체는 없는 상태였으므로 그대로 사용하기로 한 것이다.

저지현상은 공시적으로 경쟁하는 여러 접사가 서로 관계를 맺는 현상이므로 고차원적인 분석 방법이 필요했다. 이처럼 거시적인 언어 구조는 파생어를 빈도에 따라 단순히 나열하는 1차원적인 해석으로는 인지할 수조차 없는 경우가 많다. 따라서 본 연구에서는 언어 자료가 갖는 통계적 특성을 적극적으로 활용하여 나름의 분석 도구를 만들어낸 것이다.

본 연구는 기존에 잘 알려진 언어학적 현상을 새로운 방식으로 분석하는 과정이었으므로 주된 아쉬움도 방법론에 남는다. 가장 크게 아쉬운 점은 동원한 말뭉치의 크기가 너무 작고 내부 텍스트의 다양성도 부족하다는 것이다. 계량적 방법론의 힘은 동원할 수 있는 자료의 양과 질에 좌우되기 때문이다. 말뭉치 구축이 진행됨과 함께 같은 주제를 두고 여러 번 연구하더라도 의미 있는 결과물을 얻을 수 있을 것이다.

그러나 본 연구의 일부를 참고하여 후속 연구를 진행한다면 가장 흥미로운 주제는 단어의 저장·분석 지점에 관한 것이 아닐까 한다. 파생어만이 아니라 용언의 활용형이나 복합어의 형성 등 국어 연구의 많은 주제가 이와 관련된

다. 인지실험이나 뇌과학적 접근을 통한 연역적인 연구가 가능하다면 계량적으로 바라본 본 연구와 서로의 약점을 보완하는 상승효과를 얻을 수 있을 것이다.

# 참고문헌


고영근·구본관(2018), 《우리말 문법론》, 집문당.

구본관·박재연·이선웅·이진호·황선엽(2015), 《한국어 문법 총론 1》, 집문당.

김성규(1987), 語彙素 설정과 音韻現象, 《국어연구》 77, 서울대학교 국어연구회.

김창섭(1984), 形容詞 派生 接尾辭들의 機能과 意味, 《진단학보》58, 145-161.

김창섭(1995), 국어 파생접사와 파생어의 사전적 기술, 《애산학보》 16, 165-201.

김창섭(1996), 《국어의 단어형성과 단어구조 연구》, 국어학총서 21, 태학사.

김한샘(2015), 교육용 어휘 선정을 위한 접두사의 생산성 연구, 《우리말글》 65, 47-73.

남기심·고영근·유현경·최형용(2019), 《표준 국어문법론(전면개정판)》, 한국문화사.

노명희(2005), 《현대국어 한자어 연구》, 국어학총서 49, 태학사.

송철의(1977), 派生語 形成과 音韻現象, 《국어연구》 38, 서울대학교 국어연구회.

송철의(1992), 《國語의 派生語形成 硏究》, 국어학총서 18, 태학사.

시정곤(2006), 국어 형태론에서의 '생산성' 문제에 대한 연구, 《형태론》 8-2, 257-276.

신효필(2005), 언어 자료의 통계 분석과 관련된 몇 가지 고려사항들, 《어학연구》 41-3, 655-682.

신효필(2009/2012), 《언어학과 통계 모델》, 서울대학교 출판문화원.

안예리(2008), 형용사 파생 접미사의 어기 공유 현상, 《형태론》 10-1, 63-83.

윤동원(1986), 形容詞 派生接尾辭(-스럽-), (-롭-), (-답-)의 硏究, 《국어국문학 논문집》 23.

이광호(2009), 《국어 파생 접사의 생산성과 저지에 대한 계량적 연구》, 국어학총서 65, 태학사.
李秉根(1986), 國語辭典과 派生語, 《어학연구》 22-3, .
鄭承喆(2007), 被動詞와 被動接尾辭, 《진단학보》 104, 127-146.
차준경(2011), 형태적 생산성과 저지 현상, 《형태론》 25, 125-145.
최정도(2020), 한자어 접미사의 생산성에 대한 소고 - '-당(當), -별(別), -식(式), -용(用)'을 중심으로, 《반교어문연구》 54, 13-45.
최형용(2016). 《한국어 형태론》, 역락.
하치근(2010), 《우리말 파생형태론》, 경진.
황화상(2008), 접사, 단어 형성 그리고 어휘부, 《형태론》 10-1, 157-167.
황화상(2013), 《현대국어 형태론》, 지식과교양.
Akbari, N.(2016), Word frequency and morphological family size effects on the accuracy and speed of lexical access in school-aged bilingual students, International Journal Of Applied Linguistics 26-3, 311-328.
Anshen, F. and Aronoff, M.(1988), Producing morphologically complex words, Linguistics 26-4, 641-656.
Aronoff, M.(1976), Word formation in generative grammar, Linguistic Inquiry Monographs Cambridge, Mass 1, 1-134.
Aronoff, M. and Lindsay, M.(2014), Productivity, blocking, and lexicalization, The Oxford Handbook Of Derivational Morphology, Oxford University Press.
Baayen, R. H.(1989). Corpusgebaseerd onderzoek naar morfologische produktiviteit. Spektator 19, 213-217.
Baayen, R. H. and Lieber, R.(1991), Productivity and English derivation: A corpus-based study, Linguistics 29-5, 801-844.
Baayen, R. H.(1992), Quantitative aspects of morphological productivity, Yearbook of Morphology 1991, 109-149.
Baayen, R. H.(1993), On frequency, transparency and productivity. Yearbook of Morphology 1992, 181-208.
Baayen, R. H.(1994), Productivity in language production, Language and

Cognitive Processes 9-3, 447-469.
Baayen, R. H. and Renouf, A.(1996), Chronicling the Times: Productive lexical innovations in an English newspaper, Language, 69-96.
Baayen, R. H. and Neijt, A.(1997), Productivity in context: A case study of a Dutch suffix, Linguistics 35-3, 565-587.
Baayen, R. H. and Schreuder, R.(1999), War and peace: Morphemes and full forms in a noninteractive activation parallel dual-route model, Brain and Language 68-1·2, 27-32.
Baayen, R. H. and Schreuder, R.(2000), Towards a psycholinguistic computational model for morphological parsing, Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 358-1769, 1281-1293.
Baayen, R. H.(2009), Corpus linguistics in morphology: Morphological productivity, Corpus Linguistics 2, 899-919.
Baayen, R. H, Schreuder R. and Sproat R.(2000a), Morphology in the mental lexicon: A computational model for visual word recognition, Lexicon development for speech and language processing, 267-293.
Bauer, L.(2001), Morphological productivity, Cambridge: Cambridge University Press.
Bertram, R, Laine, M, Baayen, R. H, Schreuder, R. and Hyönä, J.(2000), Affixal homonymy triggers full-form storage, even with inflected words, even in a morphologically rich language, Cognition 74-2, B13-B25.
Bybee, J.(1985), Morphology: A study of the relation between meaning and form, John Benjamins Publishing Company.
Bybee, J.(2010), Language, usage and cognition, Cambridge: Cambridge University Press.
Dalton-Puffer, C. and Plag, I.(2000), Categofywise, some compound-type morphemes seem to be rather suffix-like: On the status of -ful, -type, and -wise in Present Day English, Folia Linguistica XXXIV, 225-244.
Embick, D. and Marantz, A.(2008), Architecture and blocking, Linguistic

Inquiry 39-1, 1-53.
Frauenfelder, U. H. and Schreuder, R.(1992), Constraining psycholinguistic models of morphological processing and representation: The role of productivity, Yearbook of Morphology 1991, 165-183.
Gardani, F, Rainer, F. and Luscheutzky, H. C.(2019), Competition in morphology: A historical outline, Competition in Inflection and Word-Formation, 3-36.
Giegerich, H. J.(2001), Synonymy blocking and the elsewhere condition: Lexical morphology and the speaker, Transactions of the Philological Society 99-1, 65-98.
Goldberg, A. E.(2011), Corpus evidence of the viability of statistical preemption, Cognitive Linguistics 22-1, 131-153.
Goldberg, A. E.(2016), Partial productivity of linguistic constructions: Dynamic categorization and statistical preemption, Language and Cognition 8-3, 369-390.
Haspelmath, M. and Sims A. D.(2013), Understanding morphology, Routledge.
Hay, J.(2001), Lexical frequency in morphology: Is everything relative?, Linguistics 39-6, 1041-1070.
Hay, J. and Baayen R. H.(2002a), Parsing and productivity, Yearbook of Morphology 2001, 203-235.
Jensen, J. T.(1990), Morphology: Word structure in generative grammar, Amsterdam: John Benjamins Publishing Company.
Katamba, F.(1993/2006), Morphology, New York : Palgrave Macmillan.
Kiparsky, P.(1982), Lexical morphology and phonology, Linguistics in the Morning Calm, 3-91.
Kiparsky, P.(2005), Blocking and periphrasis in inflectional paradigms, Yearbook of Morphology 2004, 113-135.
Paul, K.(2010), Dvandvas, blocking, and the associative: The bumpy ride from phrase to word, Language (Baltimore) 86-2, 302-331.

Lindsay, M. and Aronoff, M.(2013), Natural selection in self-organizing morphological systems, Morphology in Toulouse: Selected Proceedings of Décembrettes 7, 133-153.
Mendaza, R. M.(2015), Matching productivity indexes and diachronic evolution: The Old English affixes ful-, -isc, -cund, and -ful, Canadian Journal of Linguistics 60-1, 1-24.
Meyer, A. S. and Bock, K.(1992), The tip-of-the-tongue phenomenon: Blocking or partial activation?, Memory and Cognition 20-6, 715-726.
Miyagawa, S.(1984), Blocking and Japanese causatives, Lingua 64-2, 177-207.
Ortmann, A.(2004), A factorial typology of number marking in noun phrases: The tension between economy and faithfulness, Explorations in Nominal Inflections, 229-267.
Pierrehumbert, J. and R. Granell(2018), On hapax legomena and morphological productivity, Proceedings of the Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology, 125-130.
Plag, I.(2003), Word-formation in english, Cambridge: Cambridge University Press.
Plag, I.(2012), Morphological productivity : Structural constraints in English derivation, Boston: De Gruyter Mouton.
Plag, I.(2020), Productivity, The Handbook of English Linguistics(second edition), 483-499.
Plag, I., Dalton-Puffer, C. and Baayen, R. H.(1999), Morphological productivity across speech and writing, English Language and Linguistics 3, 209-228.
Rainer, F.(1988), Towards a theory of blocking: The case of Italian and German quality nouns, Yearbook of Morphology 1988, 155-185.
Saul, L.(1972), Non-paradigmatic forms: Suppletion or preemption?, Foundations of Language 8-3, 346- 359.

Scalise, S.(1986), Generative morphology, Berlin/Boston: De Gruyter.
Schreuder, R. and Baayen R. H.(1995), Modeling morphological processing, Morphological Aspects of Language Processing 2, 257-294.
Van Marle, J.(1992), The relationship between morphological productivity and frequency: a comment on Baayen's performance-oriented conception of morphological productivity, Yearbook of Morphology 1991, 151-163.
Zipf, G. K.(1949), Human behavior and the principle of least effort: An introduction to human ecology, Addison-Wesley Press.

Abstract

# A Quantitative Linguistic Study on Blocking in Contemporary Korean

Cho Sungwon
Dept. of Korean Language and Literature
The Graduate School
Seoul National University

This paper is about capturing the pattern of blocking of affixal derivation in Korean by introducing quantitative linguistic methodologies, and about seeking for explanations behind such pattern. Blocking is a phenomenon in which a certain morphologically complex form being not able to appear due to the mere existence of a synonymous rival. Unfortunately, there has been tendency among both Western and Korean researchers in the field of morphology to shun dealing with this concept because of its complexity. Unlike more popular subjects such as morphological productivity, one has to deal with multiple affixes(or derivational rules) simultaneously in order to find any regularities from blocking. So the concept of blocking has long been treated like a dustbin which contains all the non-occurrence leftovers that linguistic restrictions could not handle. It was of course for most of morphologists, even those who are familiar with quantitative approaches, to concentrate working with productivity. As for the Korean linguistics, It would be safe to say that

there has not been any research in the field of quantitative linguistics that was exclusively on blocking.

However, as high-quality Korean corpus continue to be built and various statistical methods and software to process big data developed, dealing with intricate morphological topics is becoming unprecedentedly feasible. Since morphological phenomena such as derivation are closely related to frequency, it was considered golden opportunity to study the neglected subject of blocking in a quantitative perspective. Under this context, three general objectives were set for this paper. First, revise the existing quantitative methods, which resorted largely to intuitive analysis. Second, collect the latest frequency distribution data of derivatives for each affix and analyze them to obtain necessary indicators for exploring their linguistic implications. Third, identify the previously unknown type of blocking that occurs between affixes.

In chapter 2, new quantitative method for processing corpus into data table and visualizing them to be more suitable for the purpose of this discussion was devised. According to previous arguments, when a derivative is retrieved from the lexicon, it follows one of either paths: direct whole-access route or parsing route. This mechanism is related with frequency, since derivative words with high frequency tend to be stored as a whole, while infrequent ones are more likely to be composed when they are needed. It was natural that a lot of quantitative researches on productivity eventually endeavored to locate frequency threshold where the derived words start to be easier to be stored as a whole. The threshold was necessary for calculating productivity in the sense that words that did not follow the parsing route should be excluded from analysis beforehand. The new method for analyzing frequency distributions in chapter 2 was designed to locate this critical point of parsing.

In chapter 3, frequency distribution of derivatives of competing and non-competing affixes are analyzed, and then the results are used to

estimate productivity and parsing/storage threshold of each affix. Analyzed data of frequency distribution works as an basic data for identifying how blocking realizes between word tokens. As former researchers asserted, frequency takes crucial part in accounting for blocking. This is because blocking is basically a deterrence in morphological occurrence, of which deterring power is a function of frequency of the pre-stored word.

What this paper tried to assert the most in chapter 4, though, is that the blocking between suffixes - which have long been disproved of its existence - does occur, when neologisms are coined. Correlation between productivity and threshold frequency of multiple affixes suggest that there is a systematic and general gap in productivity between competing affixes. This gap is caused by 'semantic specialization' in producing new words between affixes. The specialized meanings were irrelevant to their typical usage. Then it could be said that blocking between affixes has occurred, since the mere presence of competing affix coerced each affix to specialize in certain meaning and usage, which in turn lead to the productivity gap between them.

Highly convoluted phenomenon like blocking reveal its true identity only if adequate means are introduced. By integrating linguistic theories and statistical achievements into more advanced quantitative linguistic methodology, this paper was able to capture the macroscopic structure inherent in natural language.

**keywords : blocking, derivation, competition, productivity, parsing, frequency**

*Student Number* : 2019-26018