

파생어의 생산성 측정

차 준 경, 강 범 모
고려대학교 언어학과

A Measure of Productivity in Derivational Morphology

Joon-kyung Cha, Beom-mo Kang
Department of Linguistics, Korea University

요 약

이 연구는 지금까지 국어 형태론에서 사용되지 않았던, 코퍼스를 이용한 계량적인 방법으로 파생어의 생산성 정도를 측정하고, 그 결과로 국어 파생 형태론에서의 생산성을 기술한 것이다. 각각의 접사들의 생산성 정도에 대한 수치를 제시함으로써 좀 더 정확하게 상대적인 생산성 비교를 할 수 있도록 하였다. 접사의 생산성 정도 측정방법은 Baayen(1989)에서 제시한 것으로, 특정 접사를 가지고 코퍼스에 단 한번 출현하는 단어의 수(n_i)와, 주어진 접사를 가지고 코퍼스에 나오는 단어의 총수(N)의 비율로 접사의 생산성 정도를 측정한다($P = n_i/N$). 200만 어절 및 1000만 어절 코퍼스를 기반으로 국어의 대표적인 파생접미사들 중 명사파생 접미사 '-이', '-음', '-기', 형용사파생 접미사 '-스럽-', '-롭-', '-답-', 동사파생 접미사 '-거리-', '-대-', '-이-'의 생산성 정도를 측정하였다. 본 연구에서 채택한 코퍼스를 이용한 언어 연구 방법은 기존의 사전을 이용하여 파생어의 생산성을 측정하는 것에 비해 앞선 것이라 할 수 있다.

1. 서 론

파생법에서 새로운 단어는 단어 어기(word base)에 일정한 접사가 붙어서 형성된다. 새로운 단어형성의 기제인 접사에 관한 연구는 국어 형태론에서 지금까지 많이 진행되어 왔다. 주로 파생어 형성 규칙을 정밀화하고 제약을 밝혀 내어 각각의 접사의 의미를 살피는 연구들이었다(송철의 1991). 연구자들은 연구 대상을 선정하면서 어떤 접사가 공식적으로 생산적인가 하는 문제, 또 생산적인 규칙들 사이의 생산성의 차이를 구별하는 문제 등에 직면하게 되었다. 지금까지 국어 형태론에서 생산성은 연구자의 직관에 의존하거나 사전 등을 이용하여 측정하였다. 그런데 사전에는 실제 언어화자들이 잘 사용하지 않는 단어들이 포함되어 있어서 정밀한 연구가 불가능하다. 사전을 대신하여 코퍼스를 이용하는 방법이 그 대안이며, 본 연구에서 코퍼스를 이용하여 파생어들의

생산성 정도를 측정하고 이를 비교하였다

2 생산성

파생 형태론에서 생산성이란 언어화자가 단어 형성규칙에 의해 무한한 단어를 무의도적으로 만들어 낼 수 있는 가능성이다(Schultink 1961). Lyons(1977)는 '화자가 전에 듣지 못했던 발화(단어)를 포함하여, 무한히 많은 발화(단어)를 만들어 내고 이해할 수 있는 시스템이 있고 이것은 언어의 속성으로 생산성이다'라고 한다. 또, 일반적으로 '많은 형태에 영향을 끼치고, 많은 단어를 생산해 내는 단어형성과정은 생산적이다'라고 간주한다(Katamba 1993). 단어형성과정에서 어떤 접사가 새로운 단어를 많이 생산할 수 있다면 그 접사는 생산적인 접사가 되고, 그 단어형성과정은 생산적인 단어형성규칙이 된다. 즉, 접사가 생산적이라면 자유롭게 접미되고, 일반 원리에 의하여

그 단어를 전에 듣지 못한 사람도 쉽게 알아들을 수 있는 단어를 만들어 내는 것이다(Bochner 1993).

생산성은 상대적인 개념으로 정도성의 문제이다. 이분법적으로 어떤 단어형성규칙은 생산적이고 다른 규칙은 그렇지 못하다고 말할 수 없을 것이다. 또, 생산성은 시간과 관련이 있다. 어떤 시기에 생산적인 규칙이 다른 시기에는 생산적이지 않을 수도 있고, 전에 없었던 새로운 규칙이 나타나기도 한다.

일반적으로 생산적인 규칙이라면 파생어형성 규칙에 의해 만들어진 단어의 수가 많다는 뜻으로 보았다. 즉 생산성 정도를 빈도수에 기초하여 측정하였다. Aronoff(1976)은 단어형성과정에서 어기와 접사간에 일정한 제약이 있음에도 불구하고 제약조건을 무시한 위와 같이 빈도수로 생산성을 측정하는 관점을 반대하였다. 그는 '단어형성규칙의 입력이 되는 단어의 총수에 대한 파생되어 나오는 단어의 수에 대한 비율'로 생산성을 측정하였다. 이를 수식으로 나타내면 다음과 같다.

$$(1) I = \frac{V}{S}$$

여기서 I는 생산성 지표를 나타내고, S는 규칙의 입력으로 가능한 단어의 총수이고, V는 출력형으로 나오는 실제 단어의 수를 말한다. 이 방법은 그 전까지의 빈도수에 근거하여 접사의 생산성 정도를 측정하는 방법보다는 앞선 것이었으나 여러 문제점을 내포하고 있다. 즉, 이 방법은 가능단어(possible word)와 실제 단어(actual word)의 수효를 정확히 측정하여야 한다는 것을 전제로 한다. 그러나, 현실적으로 가능단어와 실제단어의 수를 정확히 측정하는 것은 어려운 것이다. 사전을 이용하여 단어 목록을 얻는다고 해도 사전은 몇몇 사전편찬자에 의해 집필되므로 편찬자의 주관에 반영된 것이어서 정확한 단어목록을 얻기 힘들다.

Baayen(1989)은 코퍼스를 이용하여¹⁾ 파생접사의 생산성 정도를 측정하는 방법을 고안해 냈다. 그는 생산성이 있는 단어형성규칙에서 신조어가 많이 만들어지는 것에 착안하여, 어떠한 표본(코퍼스)에서 주어진 접사를 가지고 있는 모든 단

어의 수(the number of tokens)에 대한 주어진 접사를 가지고 코퍼스에서 오직 한번만 나온 단어(hapax legomena)의 수의 비율로서 생산성 정도를 측정하였다.

$$(2) P = \frac{n_1}{N}$$

여기서 P는 생산성 정도(degree of Productivity), n_1 은 오직 한번 나온 단어의 수(the number of hapax legomena), N은 코퍼스에서 주어진 접사를 가지고 있는 단어의 수이다.

이 수식 자체는 주어진 접사를 가지고 있으면서 한번 나온 단어의 비율을 측정하는 방법이지만, 코퍼스의 크기가 커지게 되면 한번 나온 단어 중에서 신조어가 나타날 가능성이 높아질 것이라는 점에 근거하여, 간접적으로 생산성 정도를 측정하는 것이다.

생산적인 접사들은 코퍼스에서 상대적으로 한번 나온 단어의 수가 많고, 생산적이지 못한 접사들은 상대적으로 적은 수의 한번 나온 단어들이 출현한다. 또, 여기서 측정된 생산성 정도를 가지고 같은 통사범주의 단어를 형성하는 규칙들 간의 상대적인 생산성 비교도 가능하다. 어떤 접사가 생산적인지를 알고 싶을 때는 접사의 파생범주와 동일한 범주의 단일어(simplex word)의 생산성 정도와 비교한다. 단일어는 배우기 어려운 단어가 많이 있고, 빈도수가 높거나 한번 출현한 단어가 적다. 이 부류의 단어들은 P값이 매우 낮기 때문에 같은 범주를 형성하는 파생접사의 생산성 정도와 비교하면 그 접사의 생산성에 대한 비교가 가능하다. 주어진 접사의 생산성 정도가 단일어의 생산성 정도보다 낮다면 그 접사에 의한 파생규칙은 생산적이지 않음을 알 수 있게 되고, 반대로 접사의 생산성정도가 단일어의 생산성정도보다 높다면 그 파생규칙은 생산적인 규칙이 된다.

3. 연구방법

고려대학교에서 구축한 200만 어절의 시험용 코퍼스(TESTDB2)와 1000만 어절 코퍼스(가칭 KOREA1 Corpus)를 사용하였다. TESTDB2는 200만 어절의 코퍼스로 1000만 어절 규모의 한국어 코퍼스를 구축하기 이전 단계의 시험용 코퍼스

1) 언어연구를 위한 코퍼스의 일반적인 사용 상황은 강범모(1995) 참조

이다. 먼저 200만 어절의 코퍼스에서 실험한 뒤 1000만 어절의 코퍼스로 이용해 보았다.

여기에 사용된 코퍼스는 태그된 코퍼스가 아니기 때문에 중의성 문제를 해결하기 위해 연구자가 일일이 자료를 보아가며 수작업으로 연구를 진행해야 했다. 예를 들어 '그의 명예로운 죽음을 잊지 말자'의 '죽음'은 '-음'이 붙은 파생명사이지만 '그가 일찍 죽음이 안타깝다'에서의 '죽음'은 동사의 명사형이다. 현재 상태에서 이들의 구분은 자동적으로 될 수 없고 예문을 가지고 연구자가 수작업으로 해야만 한다. 200만 어절에서는 어느 정도 수작업이 가능하다고 볼 수 있겠지만 1000만 어절이 넘는 경우에는 수작업으로 처리하기에는 시간이 너무 많이 걸린다. 따라서 여기서는 비교적 수작업이 덜 필요한 형용사파생 접미사의 경우만 200만 어절 코퍼스에서의 분석과 아울러 1000만 어절 코퍼스를 사용한 분석을 시도하였다.

코퍼스의 내용은 현대 국어 표준어를 중심으로, 시기는 1970년대부터 최근 것을 위주로 하며 문어와 구어를 적절히 안배하고, 일반적인 내용을 담은 인문, 사회, 자연, 문화, 예술, 과학 등의 분야를 망라한 것이다(대략, 구어/준구어 12%, 신문 20%, 잡지 10%, 책-정보 33%, 책-상상 30%).

자료는 용례추출프로그램인 KCAT²⁾을 이용하여 파생접미사가 포함되어 있는 단어를 추출하였다. 또 형태소 분석을 통한 빈도수 조사 통계 프로그램을 이용하기도 하였다. 이 코퍼스는 통사의 의미적으로 분석이 된 코퍼스가 아니므로 데이터를 추출할 때 중의성의 문제가 발생한다. 이러한 중의성의 문제를 해소하기 위해 개별 단어의 용례를 일일이 보아가며 분석하였다. KCAT는 단순 용례 탐색에서 형태소 분석에 의한 탐색과 어절에 의한 탐색을 지원한다. 본 연구에서는 어절 탐색을 통해서 파생어의 수효를 계산하였다. 다음 장에서는 이러한 방법을 가지고 실제 분석한 결과를 보기로 한다.

2) KCAT(Korean Corpus Analysis Tool)은 코퍼스색인, 용례추출, 통계정보 추출, 연어 추출 부분으로 구성되어 있다(이호 외 1994). 본 연구에서는 용례추출 부분을 이용하였다.

4. 명사파생 접미사 '-이', '-음', '-기'

국어의 명사파생 접미사들 중, 기존의 연구에서 생산성이 높다고 일컬어지는 접사 3종을 선정하여 200만 어절 코퍼스에서 각각의 생산성 정도를 측정해 보았다.

접사	N	V	n ₁	P
2,000,000 어절				
동사에서				
기	1151	103	25	0.00271
이	1441	40	3	0.00208
음	10258	158	19	0.00185
형용사에서				
기	152	5	2	0.01315
음	1531	65	16	0.01045
이	227	5	0	0.00000

【표 1】 명사파생접미사의 생산성 정도

위 표에서 N은 코퍼스에서 주어진 접사를 가지고 출현한 단어의 총 수이고, V는 주어진 접사를 가진 단어의 종류의 수이며 n₁은 주어진 접사를 가지면서 한 번 나온 단어의 수이다. P는 생산성 정도로서 n₁/N을 나타낸다.

'-이'는 동사어간이나 형용사 어간으로부터 명사를 파생한다(예: '놀이', '먹이', '깊이', '높이' 등). 중세국어 이래 가장 대표적이고 현대국어에서도 생산적인 파생능력을 가지고 있다고 여겨진다. 200만 어절 코퍼스에서 1441번 출현하였고, 종류는 40종, 그 중 한번 나온 단어는 3종이고 생산성 정도는 0.00208이다. 이것은 단일이 어기와 비단일이 어기를 구별하지 않고 계산한 것이다. 비단일이 동사 어간과 단일이 동사 어간을 구별하여 생산성 정도를 측정하면 비단일이 경우 생산성 정도는 n₁=50, P=0.03698로 단일이 어기일 때 보다 높아진다. 단일이 형태소를 어기로 하는 명사파생어는 12종, 563번 출현함으로써, '-이' 명사파생어의 약 39%를 차지하였다.

'-음'은 형용사가 어기일 때는 추상명사를 파생시키고(예: '슬픔', '기쁨'), 동사가 어기일 때는 행위, 사건명사를 파생시킨다(예: '죽음', '값음'). 어기가 동사일 때 '-음' 명사파생어의 출현 빈도(N)는 10258이고 모두 158종류이다. 한번 출현 단어(n₁)의 수는 19로 생산성 정도(P)는 0.001852이

다. 형용사 어간에 접미할 때의 생산성 정도는 0.01045 이다.

하치근(1989: 224)에서는 '-음' 명사의 파생력(생산력)이 같은 계열의 접미사 가운데 가장 높다고 하였다. 이것은 '-음'이 결합되어 명사를 파생한 수가 '-기'나 '-이'보다 많기 때문이라고 한다. 이것은 사전에 등재된 단어의 수만으로 생산성 비교를 한 것이다. 그러나 생산적인 단어형성 규칙들은 신조어를 계속 만들어 내므로 사전에 등재된 단어의 수만으로는 생산성의 비교를 할 수 없다.

'-음' 파생 명사의 종류 수는 다른 파생명사에 비해 많지만 빈도수가 높은 단어가 많이 출현했기 때문에('삶': 1021번 출현) 생산성 정도가 다른 접사('-기')에 비해 낮게 나왔다. 한편, 파생명사의 수(V)는 파생 규칙의 사용량을 나타낸다. 즉, '-음' 파생규칙이 사용된 양이 '-기' 파생규칙의 사용량보다 크다는 것을 나타낸다. 사용량의 차이는 '-기'가 '-음'보다 후대에 나타났고, 파생어형성규칙에서 일단 파생어가 한번 생성되면 계속 쓰여진다는 점에서 설명될 수 있다(송철의 1992: 155).

'-기'는 형용사어간과 결합할 때는 척도명사, 동사어간과 결합할 때는 행위, 사건명사를 파생한다. 형용사가 어기일 때, 코퍼스에 출현한 '-기' 파생 척도명사의 종류는 5가지이지만 전체 총 출현 횟수는 152이다. 특정명사가 빈도수의 대부분을 차지한다('크기' 132번 출현). 빈도수가 높은 단어는 생산성정도를 낮게 하는 요인이지만 한번 나온 단어가 2종이므로 생산성 정도는 0.01315이다. 척도명사를 파생하는 '-이' 접사의 생산성은 0.0000이므로 '-기'에 의한 척도 명사 파생이 '-이'에 의한 파생보다 상대적으로 생산적이라고 할 수 있다.

동사가 어기일 때 '-기' 파생접사의 생산성 정도는 0.02172로 같은 계열의 접사 중에서 가장 높다. 이것은 단일형태소 어간에서 파생된 명사와 비단일어 어간에서 파생된 명사도 포함하여 측정된 것이다. 단일 형태소 어간보다는 비단일 형태소 어간으로부터가 생산적이다(송철의 1992: 156). 비단일어 어기를 포함하고 측정하면 단어의 종류는 206종으로 전체 858번 출현하였고 이중 한번 나온 단어는 108로 생산성 정도는 0.1258이다.

이 접미사들 간의 상대적인 비교를 해보면 생산성 정도는 '-기' > '-음' > '-이'의 순서이고 파생어의 종류 수는 '-음' > '-기' > '-이'의 순서를 보였다. 이것은 단일어 어기나 비단일어 어기를 구별하지 않고 측정한 결과이다. '-음' 파생명사는 단어의 종류 수에 있어서 다른 파생어들 보다 높지만 빈도수가 높은 단어들 많이 있고 어휘화한 단어도 많이 있으므로 상대적인 생산성에 있어서 '-기'보다 낮게 나왔다. '-이' 또한 동사와 결합하여 준접미사³⁾의 역할을 할 정도로 어휘화된 경우가 많이 있으므로 생산성 정도가 낮게 나왔다. '-기'는 다른 접미사와 달리 출현 빈도수가 상대적으로 높은 단어가 적었고, 어휘화한 단어들도 적었으므로 생산성 정도가 가장 높게 나왔다.

5. 동사파생 접미사 '-거리-', '-대-', '-이'

접사	N	V	m	P
2,000,000 어절				
-거리-	1419	339	148	0.10429
-대-	204	87	52	0.25490
-이-	1130	60	19	0.01681
단일어동사	327151	416	0	0.00000

[표 2] 동사파생 접미사의 생산성 정도

'-거리-'는 의성어·의태어(부사)를 어기로 하여 동사로 만든다. 200만 어절의 코퍼스에서 1419개의 파생동사가 출현하였고, 모두 339종에서 한번 출현 단어는 148로 생산성정도(P)는 0.10429 이다.

'-거리-' 파생어 중에서 가장 높은 빈도를 보인 단어는 '중얼거리-'로 94번 출현하였다. 그 다음으로는 '두근거리-'(37), '비틀거리-'(34), '서성거리-'(30)의 순으로 나타났다. '츠즈즈거리다', '씩씩거리다' 등 의성어를 어기로 하는 신조어들이 눈에 뜨인다. '-거리-'의 어기가 될 수 있는 의성어·의태어는 자연의 소리를 본따거나 대상의 모습을 본따서 만든 것이다. 소리나 모습 등이 일단 분절음으로 바뀌면 의성어·의태어는 그 언어를 사용하는 사람들 사이에서 원래의 소리나 모습을 쉽게 예측할 수 있다고 한다. 이것은 의성어·의태어가

3) 교영근(1989)에서는 '-잡이', '-살이' 등을 준 접미사로 보기도 한다.

새롭게 만들어지면 언중들 사이에서 쉽게 이해되어 사회성을 얻는 데 기여한다(윤희원 1993).

의성어-의태어는 주로 구어, 대화체, 사적인 대화, 은어 등에서 많이 쓰이고 동시, 동요 등에서 많이 볼 수 있다(채완 1993: 68). 이러한 의성어-의태어를 어기로 한 동사들 또한 텍스트의 문체상의 제약을 받는다. 본 연구에서 살펴 본 '-거리-' 동사들은 대부분 소설 등에서 출현한 단어들이다. 코퍼스의 성격에 따라 '-거리-' 파생동사의 빈도수가 달라질 수 있다. 그러므로 자세한 '-거리-'의 파생규칙을 설정하려면 다양한 대화를 수집한 구어 코퍼스를 이용해야 한다.

'-대-'는 의성어-의태어 어기에 붙어, 같은 동작을 잇달아 되풀이하는 뜻을 가진 동사를 파생한다. 어기의 분포나 음운 형태적 제약은 '-거리-'와 동일하다. '-대-'에 의한 파생어는 '-거리-'에 의한 파생어의 방언형이라고 보기도 한다⁴⁾. 분포면에서 대부분 '-거리-'와 동일하나, '-대-'에 의한 파생어에서만 쓰이는 단어들이 있다. '무턱대-', '으스대-', '허둥대-', '뿔대-', '나부대-' 등이다. 이들은 빈도수가 다른 단어들에 비해 높은 편에 속하고 어휘화한 단어들이다('무턱대-'(21), '으스대-'(11), '허둥대-'(10)).

'-대-' 접사 파생어는 코퍼스에서 204회 등장하여 52개의 한번 출현 단어가 있었고 생산성 정도(P)는 0.25490이다. 다른 접사들에 비해 전체 코퍼스에서 출현 회수가 상대적으로 적다. '-거리-' 파생동사는 코퍼스 전체에서 339종의 동사가 출현하였음에 비해 '-대-' 파생동사는 87종의 동사만 출현하였다. 이 단어들의 빈도 또한 그리 높지 않다. 본 연구에서 사용된 코퍼스는 실제 출판된 텍스트가 많고 출판과정에서 교정작업을 거쳤다. 이 과정에서 교정자들이 '-대-'에 의한 파생어를 비표준어로 보고 '-거리-'로 바꾸었을 수도 있다. 또 연구에 사용된 텍스트의 구성상 구어보다는 문어 텍스트를 위주로 하므로 구어적 특성을 많이 갖는 접사의 연구로는 적절하지 않을 것이다.

'-이-'는 의성어-의태어 어기에 접미되어 반복 동작을 뜻하는 동사파생 접미사이다. 음운론적 제약으로 주로, 'ㄱ', 'ㅇ'으로 끝난 어기에 결합하고

소수의 'ㄹ' 말음 어근과도 결합한다. 형태론적 제약은, 어근에만 접미되며 반복형어근이 아니어야 한다는 것이다(김창섭 1992).

200만 어절의 코퍼스에서 '-이-' 파생동사는 1130회 출현하였고 종류는 60종, 한번 나온 단어는 19종으로 생산성 정도(P)는 0.01681이다. 같은 어기를 취하여 같은 통사범주를 형성하는 접사들 중에서 가장 낮은 생산성 정도를 보였다. 빈도 분포에 의하면 특정한 단어의 빈도가 높아서('움직이-' 604) 전체 출현 단어의 수(N)를 증가시켰기 때문이다. 특히 '-이-' 동사 파생어들 중에는 전체 출현 빈도에 비해 높은 빈도수를 가지는 단어들이 있다. '움직이-'(604), '망설이-'(84), '끄덕이-'(31), '지겹이-'(41), '반짝이-'(28)의 순서로, '움직이-'는 의미론적으로 어휘화가 되었고 '망설이-', '지겹이-', '설레이-' 등도 어휘화된 단어들이다. 어휘화된 단어들이 상대적으로 많고 음운론적, 형태적인 제약도 있으므로 이것은 규칙이 적용될 어기의 수를 제한하는 역할을 한다. 이러한 요인으로 인해 생산성정도가 상대적으로 다른 접사들에 비해 낮게 나온 것 같다.

6. 형용사파생 접미사 '-스럽-', '-롭-', '-답'

접사	N	V	n ₁	P
2,000,000 어절				
스럽	1763	283	112	0.06352
롭	2678	57	11	0.00410
답 ₁	559	6	0	0.00000
답 ₂	267	102	89	0.33333
단일형용사	145446	308	10	0.00095

[표 3] 200만 어절에서의 형용사파생 접미사

접사	N	V	n ₁	P
10,000,000 어절				
스럽	8036	565	223	0.02775
롭	10880	96	29	0.00266
답 ₁	3079	6	0	0.00000
답 ₂	983	395	292	0.29705

[표 4] 1000만 어절에서의 형용사파생 접미사

4) 신기철-용철의 새 우리말 큰 사전에는 '-대-' 파생어를 '-거리-' 파생어의 방언형으로 처리하고 있다.

위 결과는 200만 어절의 코퍼스와 1000만 어

절 코퍼스에서의 분석결과이다.

‘-스립-’은 명사와 어근을 어기로 하여 ‘어기의 속성의 일부를 가지고 있음’을 뜻하는 형용사를 파생한다. 현대 국어의 형용사파생 접미사 중에서 가장 생산적이라고 한다. 200만 어절의 코퍼스에서 ‘-스립-’ 형용사는 1763번 출현하였다. 단어의 종류는 283종이고 한번 나온 단어의 수는 112이다. 생산성 정도(P)는 0.06352이다. ‘-스립-’의 어기로서 빈도수가 가장 높은 단어는 ‘자연스립-’으로 152번 출현하였다. 다른 빈도수 높은 단어들과 비교해 볼 때 (‘새롭-’은 1298번 출현) 상대적으로 낮은 빈도수를 차지한다. ‘자연스립-’(152), ‘조심스립-’(110), ‘고통스립-’(99), ‘새삼스립-’(75), ‘자랑스립-’(49)의 순이다. 1000만 어절의 코퍼스에서는 가장 빈도수가 높은 단어는 ‘자연스립-’(841), ‘조심스립-’(397), ‘자랑스립-’(348), ‘새삼스립-’(333), ‘고통스립-’(274)의 순서로 200만 어절에서 높은 출현 빈도수를 보인 단어들이 1000만 어절에서도 높은 빈도수를 보인다.

‘-롭-’은 상대성 명사와 어근을 어기로 하여 ‘선행어기의 속성이 풍부히 있다’라는 뜻의 형용사를 파생한다. 여기는 모두 모음으로 끝난다는 음운적 제약 때문에 생산성이 ‘-스립-’보다 적다. 200만 어절의 코퍼스에서 ‘-롭-’파생어가 2678번 등장하였고, 가짓수는 57종이고 한번 나온 단어는 11종이다. 생산성 정도(P)는 0.00410이다. 1000만 어절의 코퍼스에서는 전체 10880번 출현하였고, 한번 나온 단어(n_1)는 29종으로 생산성 정도는 0.002666이다.

200만 어절에서 가장 빈도가 높은 단어는 ‘새롭다’로 1298번 출현하였다. 그 다음으로는 ‘자유롭-’(275), ‘괴롭-’(184), ‘외롭-’(146), ‘날카롭-’(110)의 순서이다. 1000만 어절에서는 ‘새롭-’(5622), ‘자유롭-’(961), ‘괴롭-’(645), ‘날카롭-’(447), ‘외롭-’(439), ‘흥미롭-’(221)의 순이다. 여기서 ‘자유롭-’, ‘흥미롭-’을 제외하면 모두 어휘화한 단어들이다. ‘새롭-’은 공시적으로 관형어 + 접사의 구성으로 ‘새’가 명사이던 시기에 접사에 의한 파생이 이루어졌다. 또 ‘외롭-’은 공시적으로 접두사 + 접미사의 구성으로 보아야 하고, ‘날카롭-’은 ‘날카’ + ‘-롭-’에서 ‘리’이 탈락된 것을 음운론적으로 어휘화된 단어이다.

이렇게 어휘화한 단어들이 높은 빈도수를 보임으로써 전체 출현 단어의 수를 증가시켰고 생산성 정도를 낮게 하였다. 즉, 생산성 정도가 낮은 규칙에는 빈도수가 높은 단어들이 많이 있고, 이 빈도수가 높은 단어들은 공시적인 규칙에 의해 형성되기보다는 어휘부에 저장되어서 반복 사용된다고 할 수 있다.

다음은 1000만 어절의 코퍼스에서 ‘-롭-’파생어의 어기 중 한번 나온 단어의 목록이다.

- (3) 가시, 감회, 공포, 광휘, 다스, 무위, 불화, 사치, 삼가, 생기, 순화, 신이, 양기, 여호와, 의미, 자조, 정기, 조마, 창의, 초조, 충의, 풍미, 허, 험거, 호사, 화기, 화미, 환희, 원조

이 단어들을 금성판 국어대사전(1991)에서 찾아 본 결과, ‘순화롭다, 험거롭다’만이 사전에 등재되어 있었다. 이것은 생산적인 규칙에서 한번 나온 단어 중에 신조어가 나타날 가능성이 있음을 보여주는 예이다.

‘-답-’은 명사, 어근에 접미하여 형용사를 만드는 접사로 알려져 왔다. 김창섭(1984) 이후 명사나 어근에 붙어 형용사를 만드는 ‘-답₁-’과 명사구에 붙어 형용사구를 만드는 ‘-답₂-’로 나누고 있다.

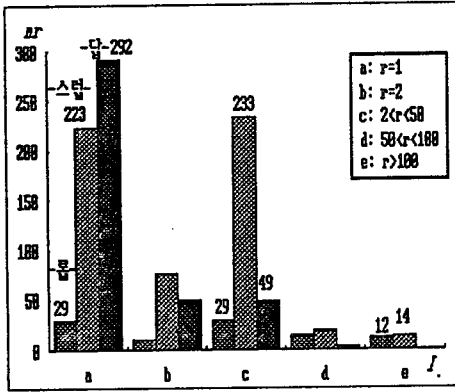
‘-답₁-’의 예로는 ‘아름답-’, ‘참답-’, ‘정답-’, ‘꽃답-’, ‘아릿답-’, ‘시(實)답-’ 등이 있으며 이들은 200만 어절 코퍼스에서 모두 6종류가 559번 출현하였다. 1000만 어절 코퍼스에서도 6종류만 출현했으며 한번 나온 단어는 없었다. 한번만 나온 단어가 없었으므로 생산성 정도(P)는 영(0)이다. 단일어의 생산성 정도보다 ‘-답₁-’의 생산성 정도가 더 낮음으로(0.00095 대 0.00000) ‘-답₁-’은 공시적으로 생산적이지 않다는 것을 알 수 있다.

명사구에 붙어서 형용사구를 만드는 ‘-답₂-’를 살펴보면, 200만 어절 코퍼스에서 전체 267번 출현하여 한번 나온 단어는 89종으로 생산성 정도는 0.3333이다. 1000만 어절 코퍼스에서는 $N=983$, $n_1=292$, $V=395$ $P=0.29705$ 로 나타났다.

1000만 어절에서 빈도수가 가장 높은 단어는 ‘사람답-’(111), ‘인간답-’(106), ‘여성답-’(58), ‘여자답-’(50), ‘남자답-’(22), ‘남성답-’(21) 등이 있다. 이 단어들은 사전에 표제어로 등재되어 있고, 용

법에 있어서 명사구보다 단어 자체에 '-답-'이 붙는 경우가 더 많으므로 의미적으로 어휘화했다고 보아야 할 것이다.

다음의 빈도분포표는 1000만 어절 코퍼스에 서의 각각의 형용사의 출현 빈도를 나타낸 것이다.



[그림 1] 1000만 어절에서의 형용사 파생어의 빈도분포

위의 표의 Y축, nr에서 r은 빈도수를 나타낸다. X축에서 a는 주어진 접사를 가지고 한번 나온 단어의 수를 나타내고, b는 두번 나온 단어의 수를 나타낸다. c는 2회 이상 50회 이하, d는 50회 이상 100회 이하, e는 100회 이상 나온 단어의 수를 표시한 것이다. '-스립-'형용사 파생어는 한번 나온 단어의 수(n_1)가 223이므로 위와 같이 표시했다.

'-스립-'과 '-롭-'을 비교하면 1000만 어절 코퍼스 전체에 출현한 모든 단어의 수는 '-롭-' 파생 형용사가 더 많다(8036 대 10880). 한번 나온 단어의 수와 생산성 정도는 '-스립-'이 더 높았지만, 빈도수가 500이상인 단어의 수는 '-롭-'이 더 많았다. 빈도수가 높은 단어가 많다는 것은 어휘화된 단어가 많다는 것을 의미하고 이것은 생산성 정도를 낮게 하였다. 그래서 '-롭-'의 생산성 정도는 '-스립-'보다 낮게 되었다. 통사적 파생인 '-답₂'를 제외하고 생산성 정도를 비교하면 '-스립-' > '-롭-' > '-답₁'의 순서이고, 코퍼스에 출현한 단어의 종류 수는 '-스립-' > '-롭-' > '-답₁'의 순서였다.

7. 결론

이 연구는 지금까지 국어 형태론에서 사용되지 않았던, 코퍼스를 이용한 계량적인 방법으로 생산성 정도를 측정하고, 그 결과를 기술하였다. 구체적으로 국어의 대표적인 파생접미사를 선정하여 각각의 생산성 정도를 측정해 보았다. 생산성 정도의 측정방법은 Baayen(1989), Baayen and Lieber(1991)에서 제시한 방법을 이용하였다. 이것은 생산성에 대한 언어화자의 직관을 잘 설명해 줄 수 있는 방법이고, 또 생산적인 규칙은 새로운 단어를 많이 형성할 수 있는 규칙이라는 생산성에 대한 개념도 잘 반영한 것이다.

먼저 동사에서 명사를 파생하는 접미사 '-이', '-음', '-기' 중 상대적으로 생산성 정도가 높은 접사는 '-기'였다. 빈도수에서는 '-음' 명사파생어의 수가 월등히 높지만 생산성 정도에 있어서는 '-기'보다 낮았다. 단어의 종류 수만으로 생산성 정도를 측정하는 이전 연구의 방법으로는 이러한 생산적인 규칙의 생성적인 측면을 반영하지 못하였다.

동사파생 접미사로는 의성이 의태어 어기에서 동사를 파생하는 '-거리-', '-대-', '-이-'의 생산성 정도를 측정하였다. 생산성 정도는 '-대-' > '-거리-' > '-이-'의 순서를 보이고 단어의 종류수는 '-거리-' > '-대-' > '-이-'의 순서였다. '-대-'는 '-거리-'의 방언형으로 보기도 하므로 '-거리-'와 '-대-'간의 생산성의 차이를 판별하기는 어려운 것 같다. 구어적 특성이 많이 반영된 코퍼스를 이용한다면 정확한 연구결과가 나올 수 있을 것이다.

형용사파생 접미사로는 '-스립-', '-답₁', '-답₂', '-롭-'의 생산성 정도를 측정하였다. 생산성 정도가 가장 높게 나온 접사는 '-답₂'로 이것은 '-답₂'가 통사적 파생이라는 점을 잘 반영한다. 어휘적 파생에서 생산성 정도는 '-스립-' > '-롭-' > '-답₁'의 순서를 보였다. '-롭-'은 다른 파생규칙에 비해 어휘화한 단어들이 많이 있기 때문에 상대적으로 생산성 정도가 낮게 나왔다. 어휘화한 단어들은 출현 빈도수가 높기 때문에 생산성 정도를 낮추고, 또, 생산성이 높은 규칙들은 상대적으로 어휘화한 단어의 수가 적음을 알 수 있었다. 1000만 어절의 코퍼스를 대상으로 형용사

파생 접미사의 생산성 정도를 측정된 결과, 200만 어절의 코퍼스에서의 측정결과와 비슷한 분포를 보였다. 상대적인 생산성 정도의 순서에서 일치하였고, 상위 빈도수의 단어들의 분포가 어느 정도 일치했다. 코퍼스의 크기가 증가함에 따라 한번 나온 단어의 수도 증가하였고 이들 중 신조어라고 여겨질 만한 단어들도 눈에 띄었다. 이런 결과에 비추어 200만 어절에서 측정된 명사파생 접미사나 동사파생 접미사의 생산성 정도도 어느 정도 타당하다고 볼 수 있을 것이다.

본 연구에서는 국어의 대표적인 파생접사의 일부를 선별하여 생산성 정도를 측정하였으나 여러 접사들에 대한 생산성 정도의 측정도 필요할 것이다. 또 대규모의 코퍼스를 이용하거나 특수한 코퍼스(구어 코퍼스등)를 사용하여 생산성 정도를 비교할 수 있을 것이다. 본 연구에서 채택한 계량적인 접근 방법, 즉 코퍼스를 이용하는 연구 방법은 기존의 연구 방법에 대한 새로운 대안으로서, 직관으로 설명하기 어려운 언어현상을 연구하는 데 많은 도움을 줄 것이다.

참고문헌

- 강범모 (1995) "언어 데이터베이스와 언어 연구", 정광의, 「한국어 데이터 베이스의 설계 및 응용을 위한 기초 연구」, 제1장, 13-44. 민음사.
- 고영근 (1989) 「국어형태론 연구」, 서울대 출판부.
- 김창섭 (1984) "형용사파생 접미사들의 기능과 의미: '-답-, -스럽-, -롭-, 하-'와 '-적'의 경우.", 진단학보 58, 145-161.
- 김창섭 (1992) "파생 접사의 뜻풀이", 새국어생활 2-1, 72-88.
- 민현식 (1984) "-스럽다, -롭다 접미사에 대하여". 국어학 13.
- 송철의 (1992) 「국어의 파생어형성 연구」, 국어학 총서 18, 태학사.
- 이호, 김진동, 임해창 (1994) "언어 정보 획득을 위한 한국어 코퍼스 분석도구", 제 6회 한글 및 한국어 정보처리 학술 발표 논문집, 297-304.
- 임홍빈 (1989) "통사적 파생에 대하여", 어학연구 25-1, 167-195.
- 윤희원 (1993) "의성어의태어의 개념과 정의", 새국어생활 3-2, 3-15.
- 정광, 이기용, 김홍규, 임해창, 강범모 (1995) 「한국어 데이터 베이스의 설계 및 응용을 위한 기초 연구」, 민음사.
- 조남호 (1988) "현대국어의 파생접미사연구", 국어연구 85.
- 채 완 (1993) "의성어의태어의 통사와 의미", 새국어생활 3-2, 54-72
- 하치근 (1989) 「국어파생형태론」, 남명문화사.
- 하치근 (1993) 「국어파생형태론」(증보판), 남명문화사.
- [국어 대사전] (1991) 김민수, 고영근, 임홍빈, 이승재. 금성출판사.
- [새 우리말 대사전] (1989) 신기철, 신용철. 삼성출판사.
- Aronoff, M. (1976) *Word Formation in Generative Grammar*, Cambridge, Mass.: The MIT Press.
- Baayen, R. (1989) *A Corpus Based Approach to Morphological Productivity. Statistical Analysis and Psycholinguistic Interpretation*. Ph.D. thesis, Vrije University, Amsterdam.
- Baayen, R. and Lieber, R. (1991) "Productivity and English Derivation: a corpus-based study", *Linguistics* 29, 801-844.
- Baayen, R. (1993) "On Frequency, Transparency, and Productivity.", In: Booji.G.E., Van Marle, J.(Eds.), *Yearbook of Morphology 1992*, 109-149. Dordrecht: Kluwer.
- Bauer, L. (1988) *Introducing Linguistic Morphology*, Edinburgh: Edinburgh University Press.
- Bauer, L. (1994) "Productivity", Asher, R. E, *The Encyclopedia of Language and Linguistics* Vol. 6, 3354-3357, Oxford: Pergamon Press.
- Bochner, H. (1993), *Simplicity in Generative Morphology*, The Hague: Mouton de Grueter.
- Culter, A. (1980) "Productivity in Word Formation", *Chicago Linguistics Society* 16, 45-51.
- Katamba, F. (1993) *Morphology*, London: The Macmillan Press.
- Lieber, R. (1992) *Deconstructing Morphology: Word Formation in Syntactic Theory*, Chicago: The University of Chicago Press.
- Lyons, J. (1977) *Semantics*, Cambridge: Cambridge University Press.
- Shultink, H. (1961) "Productiviteit als morfologisch fenomeen." *Forumder Letteren* 2: 110-125.