JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 20, No. 1, pp. 51-58, Jan. 2019

뉴스 기사에서 지속도와 편중도 기반의 신조어 사용 특징 분석

한 경 수
성결대학교 컴퓨터공학부

Analysis of Usage Characteristics of New Words Based on Persistency and Burstiness in News Articles

Kyoung-Soo Han
Division of Computer Engineering, Sungkyul University, Anyang-si 14097, Korea

[요 약]

지속적으로 변화해가는 인간의 언어에는 필연적으로 새롭게 생겨나는 단어들이 존재하기 마련이다. 기존의 언어 행태를 기반으로 개발되는 자연어처리 기술에서 이런 신조어의 처리는 매우 어려운 문제이다. 본 연구는 뉴스 기사들에서 신조어의 사용 통계를 수집하여 신조어의 사용 특징을 분석한다. 정량적인 분석을 위하여 사용량, 사용 지속성, 사용 편중도 등의 척도를 제안한다. 신조어가 처음 사용되기 시작한 이래로 사용량이 어떻게 바뀌어 가는지, 일반 신조어와 전문 분야 신조어의 사용량 변화에 차이가 있는지, 신조어의 사용 지속성이나 사용 편중도가 사용량에 어떤 영향을 미치는지, 신조어에 대한 설명 표현은 얼마나 자주 신조어와 같이 사용되는지 등을 여러 실험을 통하여 살펴본다.

[Abstract]

There are inevitably new words in human language that are constantly changing. The processing of these new words is a difficult problem for natural language processing technologies that are developed based on existing language behaviors. This study analyzes the use characteristics of the new words by collecting the usage statistics of the words in the news articles. To quantitatively analyze them, usage amount, persistency, and burstiness are proposed as analytical measures. Through a series of experiments, we examine how usage changes since the first use of the new word, whether there is a difference in the amount of use of the general new word and new terminology in special domain, how the persistency or use burstiness of the new word affects the usage amount, and how often the new word and its description are used together.

색인어 : 신조어, 사용 추이, 사용 지속성, 사용 편중도, 미등록어

Key word : New word, Usage trend, Persistency, Burstiness, Unknown word

http://dx.doi.org/10.9728/dcs.2019.20.1.51

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-CommercialLicense(http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 03 December 2018; **Revised** 13 December 2018
Accepted 20 January 2019

***Corresponding Author; Kyoung-Soo Han**

Tel: +82-31-467-8189
E-mail: kshan@sungkyul.ac.kr

Ⅰ. 서 론

인간의 언어는 시간이 흐름에 따라 변화해 나간다. 새로운 단어가 생겨나기도 하고, 기존 단어의 의미가 확장되거나 신규 의미가 추가되기도 하고, 어떤 단어는 더 이상 사용되지 않는 경우도 발생한다. 인간의 언어를 처리하는 자연어처리(natural language processing; NLP) 기술들은 이러한 언어의 동적인 특성을 반영하여야 한다. 말뭉치, 사전 등 기존 언어의 사용 행태를 기반으로 개발되는 자연어처리 기술에서 특히 새롭게 생겨나는 단어(신조어 또는 신어)의 처리는 매우 어려운 문제이다. 신조어는 사전에 등록되지 않은 단어로서 형태소 분석 과정부터 적절히 분석되지 않을 가능성이 높다. 이렇게 새롭게 생겨난 미등록어를 잘 처리하기 위해서는 신조어가 어떤 특징을 갖고 있는지, 인간이 신조어를 어떻게 사용하고 있는지 등에 관한 심도 있고 분석적인 연구가 필요할 것이다.

국립국어원에서는 1990년대부터 최근 2014년까지 신조어들을 조사하여 발표해오고 있다[1]-[3]. 이 신조어들을 기준으로 여러 관련 연구들이 수행되었다. [4]는 2002년 신조어에 대해 정착한 신조어의 의미 유형을 분류하였으며, 신조어가 정착하기 위해서는 구성 성분의 조어력, 의미의 명료성, 의미의 확장성이 필요하다고 분석하였다. [5]는 2002년~2004년 신조어 1690개에 대해 언론진흥재단의 뉴스 기사 검색을 이용하여 2015년까지의 연도별 신조어 사용 변화를 조사하였다. 신조어의 20% 정도가 10여 년 동안 지속적으로 사용되고 있었고, 약 1% 정도의 신조어만이 표준국어대사전에 등재되었다. [6]은 네이버 뉴스 기사 검색을 기반으로 신조어가 포함된 뉴스 기사 말뭉치를 구축하여 2005년, 2006년 신조어의 사용 추이를 분석하여 사전 등재의 기준으로 총빈도, 기사 건수, 연도별 분포 등을 제안하였다. [7]은 2005~2007년 생존 신조어의 사례 연구를 통해 신조어 사용 패턴을 탄생 및 성장, 쇠퇴 및 사멸, 재탄생 등으로 유형화하였다.

[8], [9]는 2000년~2012년 4대 일간지 신문 기사로 구성된 '물결21' 말뭉치를 이용하여 신조어 사용 추이를 분석하였다. [8]은 국립국어원의 신조어를 사용하지 않고 '물결21' 말뭉치로부터 신조어 명사 1243개를 추출하였고, 이들의 사용 양상을 증가 추이, 감소 추이, 일시적 사용, 간헐적 사용 등으로 유형화하였다. [10]은 신조어 조사를 위하여 신문 기사 말뭉치를 수집하고 신조어 후보, 용례 등을 추출하여 신조어 조사를 돕는 신조어 조사 지원 프로그램을 개발하였다.

기존 연구들은 주로 언어학적인 관점에서 장기간 연도별 사용 추이 등 거시적인 사용 변화에 초점을 맞추고 있다. 본 연구는 신조어 처리에 대한 시사점을 얻기 위해, 신조어의 첫 출현 시기부터 신조어 사용에 어떤 특징이 있으며 그 특징이 짧은 시간 구간별로 어떻게 변화해 가는지 보다 미시적인 분석을 수행하고자 한다. 또한 신조어의 특징을 분석적으로 파악하기 위해 사용량을 비롯하여 지속도, 편중도 등 정량적인 분석 척도를 제안한다. 본 연구는 다음과 같은 연구 질문들에 대한 답을 얻기 위해 데이터를 근거로 신조어에 대한 실험 및 분석을 진행하였다.

- 신조어는 처음 사용되기 시작하는 시점부터 미시적으로 사용량이 어떻게 변화해 가는가?
- 일반 신조어와 전문 분야 신조어의 사용량과 사용량 추이에 차이가 있는가?
- 신조어 사용량, 신조어의 지속적인 사용 경향, 일부 시기에 편중되어 사용되는 경향 등이 어떻게 관계되는가?
- 신조어에 대한 설명 표현은 얼마나 자주 신조어와 함께 사용되는가?

Ⅱ. 연구 방법

2-1 연구 대상 신조어

본 연구에서는 국립국어원에서 가장 최근에 발표한 2014년 신어 조사 보고서[3]에 등재된 신어와 미등재어를 사용하였다. 기존 연구들에서는 2000년대 초반의 신조어를 대상으로 주로 장기간 동안 생존 여부에 초점을 맞춘데 비해, 본 연구에서는 비교적 최근의 신조어를 대상으로 등장 초기의 사용 특징이 어떻게 변화해 가는지 살펴보고자 한다. 국립국어원에서는 조사 시점인 2014년에 최초로 등장한 단어를 신어, 최초 등장 시기가 그 이전인 단어를 미등재어로 구분하고 있다. 조사 도구나 조사 방법에 따라 최초 등장 시기는 달라질 수 있고, 본 연구는 새로 등장한 시기를 굳이 2014년으로 한정하여 신어와 미등재어를 구분할 필요가 없으므로 이 둘을 모두 신조어로 간주하여 분석하였다. 총 397개 단어이다.

2-2 신조어 사용 통계 수집

본 연구에서는 네이버 뉴스 기사 검색을 이용하여 신조어의 사용 통계를 수집하였다. 연구 수행 시기 기준, 네이버 뉴스 기사 검색은 일간지 13개, 방송/통신 54개, 경제/IT 71개, 인터넷 신문 152개, 스포츠/연예 128개, 지역지 81개, 매거진 86개, 전문지/기타 232개 등 총 817개 언론사 뉴스 기사를 검색한다.

네이버 뉴스 기사 검색 API[11]를 사용하면 한 검색어에 대해 최대 1000개까지만 검색이 가능하다. 검색 옵션에도 제한이 많다. 따라서 본 연구에서는 자체 제작한 수집기 프로그램을 통해 네이버 뉴스 기사 검색 페이지[12]에 접속하여 검색을 수행하고 검색 결과 페이지를 해석하여 검색 결과 및 통계를 자동 산출하였다. 네이버 뉴스 기사 검색은 해당 검색어가 포함된 총 뉴스 기사의 건수 정보는 제공하지만, 뉴스 기사에 대한 링크는 검색어당 4000개까지만 제공한다. 따라서 기간 제한 없이 검색하되 오래된 순으로 정렬되도록 검색하여 최초 사용일로부터 시간 순서대로 총 4000개 뉴스 기사를 검색하여 분석에 활용하였다.

표 1. 분석 대상 신조어
Table 1. New words to analyze

Category	Number of Words	Ratio
General Words	264	69.5%
Terminology	116	30.5%
Total	380	100.0%

검색어는 신조어 표현을 “ ”로 둘러싸서 구성함으로써 신조어 표현 그대로 등장하는 뉴스 기사를 검색한다. 다만 ‘강려크하다’처럼 ‘-하다’로 끝나는 용언의 경우는 ‘-하다’를 뗀 체언 형태(예: “강려크”)로 검색어를 구성하여 통계를 수집하였다.

2-3 분석 대상 선정 및 검색 후처리

397개 신조어 중 네이버 뉴스 기사 검색을 통해 사용 통계 수집에 어려움이 있는 단어 17개는 분석 대상에서 제외하여, 결국 총 380개 신조어에 대해 분석을 수행하였다. 제외 유형과 제외 사유, 제외 단어 예는 다음과 같다.

- 동형이의어 과다(12개): 신어 조사 보고서[3]에 등재된 의미와는 다른 의미로 사용되는 동형이의어가 많아 통계 값이 훼손되는 경우(예: 그린라이트, 익수, 브라운 칼라, 트윕, 잼스)
- 오타 과다(1개): 전혀 다른 단어가 오타로 인해 해당 신조어로 사용된 경우(예: 슈피 – 슈퍼의 오타)
- 부분 매칭(1개): 네이버 검색 기능상의 문제로서 신조어가 다른 단어의 일부분과 매칭 되어 검색되는 경우(예: 습제)
- 노이즈(1개): 일부 언론사에서 기사 내용과 상관없이 검색 가능성을 높이기 위해 모든 기사의 마지막 부분에 해당 단어를 삽입한 경우(예: 후방 주의)
- 비신조어(1개): 너무 오래 전부터 사용되고 있는 단어로서 신조어로 간주하기 어려운 경우(예: 커팅식)
- 용언-체언 중복(1개): ‘X하다’ 형태의 용언과 X 체언이 동시에 존재할 경우 체언만을 분석대상으로 사용하고 용언은 제외시킨 경우(예: 심쿵하다)

분석 대상으로 선정된 380개 단어라도 네이버 뉴스 기사 검색 결과에 신조어 의미에 해당되지 않는 검색 결과가 포함되는 경우가 있다. 이 검색결과들에 대하여 수작업으로 잘못된 혹은 의도치 않은 검색 결과를 삭제함으로써 분석 대상에서 제외되는 단어를 최소화하였다.

표 1은 분석 대상 신조어의 구성을 보여준다. 380개 단어 중 전문 분야 단어(Terminology)가 30.5%, 일반 단어(General Words)가 69.5%를 차지한다. 단어의 전문 분야는 [3]을 그대로 따랐다.

2-4 신조어 사용 특징 분석 척도

신조어 사용의 특징을 다각도로 분석하기 위하여 다음과 같은 다양한 척도를 이용하였다.

1) 사용량

신조어 사용량은 해당 단어가 등장한 뉴스 기사 건수로 정의한다. 총사용량은 단어가 최초로 등장한 시기부터 조사된 전체 뉴스 기사 건수이다. 기간별 사용량은 기존 연구들과는 달리 절대적인 시간 구간을 사용하지 않고 각 단어별 최초 등장 시기를 기준으로 상대적인 시간 구간으로 사용량을 집계한다. 각 단어마다 최초 등장 시기가 다르기 때문이다. 이런 방식으로 집계하여 단어의 최초 사용 시기부터 사용 추이가 어떻게 달라지는지를 알아볼 수 있다. 기간별 사용량은 기본적으로 월간 집계된 월 사용량을 바탕으로 다양한 기간별 사용량을 계산한다. 본 논문의 실험 결과에는 3개월 단위 사용량을 사용하였다. 예를 들어, 어떤 단어가 2018년 2월 5일에 처음 사용되었다면 3개월 단위 사용량은 <2월~4월, 5월~7월, ...> 방식으로 기간을 나누어 해당 기간 사용량을 합산하여 계산한다.

2) 사용 지속도

단순히 사용량만 고려하면 짧은 시기에만 반짝 많이 사용되는 단어와 꾸준히 사용되는 단어를 구분할 수 없다. 따라서 이들을 구분 짓기 위해 사용 지속도라는 척도를 이용한다. 단어의 사용 지속도는 (해당 단어의 사용량이 1 이상인 기간 수) / (총 기간 수)로 정의한다. 예를 들어, 어떤 단어가 2018년 1월에 처음 사용됐고 <1월, 2월, 6월>에 사용됐다면, 3개월 단위 사용량은 <1~3월, 4~6월, 7~9월> 등으로 집계될 것이다. 2018년 9월까지의 사용량으로 3개월 단위 사용 지속도를 계산한다면, 총 3개 기간 중 2개 기간에서 사용이 되었으므로 2/3로 계산된다.

3) 사용 편중도

신조어의 쓰임을 관찰해보면 대부분의 기간 동안 사용되지 않다가 특정 시기에만 매우 많은 양이 사용되는 경우가 많다. 이런 단어의 특징을 반영하기 위한 척도가 사용 편중도이다. 편중도는 (해당 단어의 최대 사용량) / (해당 단어의 평균 사용량)으로 계산한다. 예를 들어, 총사용량이 20으로 동일한 단어 A, B, C의 기간별 사용량이 각각 <1, 0, 0, 19, 0>, <10, 0, 0, 10, 0>, <10, 5, 0, 0, 5>이라면, 3개 단어 모두 평균 사용량은 4(=20/5)이므로 사용 편중도는 A가 19/4, B와 C는 10/4로 동일한 값을 갖는다.

2-5 신조어 설명 사용 통계 수집

텍스트에 새로운 단어가 처음 사용될 때 독자는 그 단어만 봐서는 의미를 이해하지 못할 수 있다. 따라서 독자의 이해를 돕기 위해 신조어를 사용할 때는 그 단어의 의미가 무엇인지에 관한 설명이 필요할 것이다. 그렇다면 신조어에 대한 설명은 얼

마나 자주 신조어와 같이 사용될까? 신조어의 첫 사용시점부터 경과한 시간에 따라 설명의 사용량에 차이가 있을까? 그 경향을 파악하기 위하여 380개 신조어 중에서 문서에 단어와 그 설명이 같이 등장했는지 파악하기 용이한 단어를 선정하여 사용 특징을 분석하였다. 네이버 뉴스 기사 검색을 이용하여 선정된 신조어와 설명이 동시에 등장한 뉴스를 검색하여 설명 사용 통계를 수집하였다.

검색어를 결정하기 위하여 선정된 각 신조어의 의미[3]와 신조어가 등장한 문서들을 사용 시기에 따라 샘플링하여 어떤 설명 표현들이 사용되었는지를 체크하여 해당 설명이 같이 검색될 수 있는 검색어를 생성한다. 예를 들어, 신조어 '일코'의 경우 다음과 같은 신조어의 의미와 '일코'가 등장한 문서 내용을 참고한다.

- '일코' 의미: '*일반인 코스튬 플레이*'를 줄여 이르는 말
- 샘플 기사 1: ...또 다른 2030팬들의 특징은 바로 '일코', 일명 '*일반인 코스프레*'다. 공연장이나 인터넷 팬페이지 상에서는...
- 샘플 기사 2: ...덕후에 대한 시선이 곱지 않았을 때 덕후는 '*일반인 코스튬플레이*(일코)'를 했다. 만화, 영화, 게임, 캐릭터, 연예인 등에 빠진...

이를 참고하여 '일코'와 설명이 동시에 등장한 뉴스를 검색하기 위해 "일코 일반인 코스프레", "일코 일반인 코스튬" 등으로 검색을 수행한다. 검색 결과에서 뉴스 기사 URL을 기준으로 중복을 제거하여 통계를 산출한다.

표 2. 총사용량 가장 많은 신조어 20개
Table 2. Top 20 frequent new words

Rank	Word	Amount
1	플친	338542
2	심쿵	209530
3	*해시 태그*	*111817*
4	핵꿀잼	34696
5	취향 저격	26092
6	셀피	22336
7	앵그리맘	20808
8	뇌섹남	20564
9	운빨	17901
10	*인구 절벽*	*17193*
11	혼밥	16992
12	남사친	13599
13	비글미	10496
14	여사친	9983
15	*혼밥족*	*7901*
16	로봇 연기	6555
17	호캉스	5222
18	세젤예	4625
19	*오포 세대*	*4362*
20	트랜짓몰	3875

Ⅲ. 실험 결과 및 분석

3-1 신조어 총사용량

표 2는 380개 분석 대상 신조어 중 총사용량이 가장 많은 단어 20개를 추출한 것이다. Amount는 해당 단어가 등장한 뉴스 기사 건수를 의미한다. 뉴스 기사 검색은 2018년 11월 2일부터 6일까지 수행되었다. 총사용량이 가장 많은 단어는 '플친'(카카오톡 '플러스 친구'를 줄여 이르는 말)이었다. 최근 몇 년간 관련 서비스의 괄목할 만한 성장을 이 단어 사용 통계가 보여주고 있다. 또한 '심쿵', '핵꿀잼', '취향 저격' 등 방송 프로그램이나 연예 기사 등에서 어렵지 않게 볼 수 있는 단어들의 총사용량이 높게 나타났다. 상위 20개 신조어의 총사용량이 전체 신조어 사용량의 93.9%를 차지하였다.

총사용량 상위 20개 단어 중 전문 분야 단어는 '해시 태그', '인구 절벽', '혼밥족', '오포 세대' 등 4개(25%)였고, 나머지 16개(75%)는 일반 단어들이었다. 표 3은 일반 단어와 전문 단어의 총사용량을 비교하고 있다. Sum of Amount는 해당 단어들이 등장한 뉴스 기사 건수를 모두 합한 값이다. 전체 신조어 사용량의 83.6%를 일반 단어가 채우고 있어 전문 단어보다는 일반 단어의 사용량이 훨씬 더 많다는 사실을 알 수 있다.

표 4는 총사용량 낮은 신조어들의 분포를 보인다. Number of Words는 해당 총사용량을 갖는 단어 개수이다. 총 380개 단어 중 127개(33.4%) 단어의 사용량이 10건 이하였다. [3]에서 조사된 신조어 중 2개 단어('갓수 시대', '탯물')의 검색 결과가 없었다. 언론사에서 해당 기사를 더 이상 제공하지 않는 것으로 보인다. 이후의 실험 결과는 이 두 단어를 제외한 결과이다.

3-2 사용량 변화

신조어의 사용량이 어떻게 변화하는지를 살펴보기 위해 네이버 뉴스 검색 한도보다 사용량이 많은 19개를 제외하고 총사

표 3. 일반 단어와 전문 단어 비교
Table 3. Comparison of general words and terminologies

Category	Sum of Amount	Ratio
General Words	804886	83.6%
Terminology	157360	16.4%
Total	962246	100.0%

표 4. 총사용량이 낮은 신조어 통계
Table 3. Statistics of low frequency words

Amount	Number of Words	Ratio
0	2	0.5%
1	22	5.8%
2	16	4.2%
3	19	5.0%
4	14	3.7%
5	16	4.2%
6	9	2.4%
7	9	2.4%
8	4	1.1%
9	9	2.4%
10	7	1.8%
Total	127	33.4%

표 5. 총사용량에 따른 단어 그룹
Table 5. Word groups according to usage

Group	Avg. Amount	Example Words
G1	1360.8	트랜짓몰, 박스피, 어깨 깡패, 대프리카, 심멎
G2	361.9	빠던, 셀기꾼, 할빠, 운발, 의느님
G3	146.2	조직압, 맛저, 필록싱, 이다루시주맙, 모루밍족
G4	87.8	시티 캠핑, 빛삭, 꼬돌남, 쿨몽둥이, 어그로꾼

용량이 많은 순으로 단어 30개씩 표 5와 같이 그룹을 지었다. 총사용량이 많지 않은 단어들에 대해서는 사용량 변화의 경향성을 관찰하기 어렵기 때문에 총사용량이 많은 단어들을 대상으로 사용량 변화를 분석하였다. Avg. Amount는 각 단어 그룹에 포함된 단어들의 총사용량 평균을 의미한다.

그림 1은 각 그룹별 사용량 변화를 보이고 있다. 단어마다 첫 사용 시기가 다르므로 각 단어마다 첫 사용 시기를 기준으로 3개월 단위의 사용량 통계를 산출하였다. Q0은 단어가 처음 사용된 월을 포함한 첫 3개월을 의미한다. 예를 들어, 어떤 단어가 2014년 3월 5일에 첫 사용됐다면 Q0은 2014년 3월 1일부터 2014년 5월 31일까지를 의미한다. Q1은 그 다음 3개월 즉 2014년 6월 1일부터 2014년 8월 31일까지이다. 따라서 Q4는 단어 첫 발생 후 만 1년이 지난 시점을, Q16은 만 4년이 지난 시점을 뜻한다. 각 기간의 사용량은 해당 단어 그룹에 포함된 각 단어의 해당 기간 사용량을 평균한 값이다. 이와 같은 방식으로 단어의 첫 발생 직후부터 사용량이 어떻게 바뀌어 가는지를 추적하였다.

사용량이 가장 많은 단어 그룹 G1은 증가형이다. 처음 사용될 때부터 사용량이 많았으며, 그 이후 4년 동안 대체적으로 줄곧 사용량이 증가하는 경향을 보이고 있다. 처음 사용된 지 4년이 지난 후 사용량(Q16)이 첫 사용량(Q0)의 3.7배에 이르러, 4

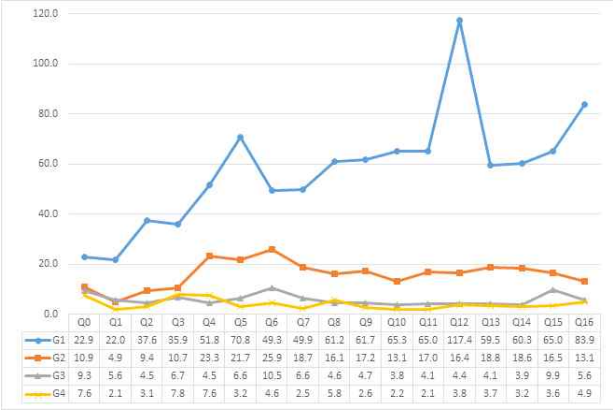

그림 1. 신조어 사용량 변화
Fig. 1. Usage change of new words

년 만에 사용량이 266% 증가하였다. 이에 비해 단어 그룹 G2는 증가 후 유지형이다. 첫 등장 후 1년 동안은 비슷한 사용량을 유지하다가 1년 이후 사용량이 2배 가까이 증가한 이후 크게 늘지 않고 비슷한 수준을 유지하였다. 4년 만에 사용량이 20% 증가한 수준에 머물렀다. 단어 그룹 G3과 G4는 점진 감소형이다. G3 소속 단어들이 때때로 G4보다 상대적으로 사용량이 많은 것을 제외하고는, 그룹 G3과 G4는 대체적으로 낮은 사용량을 유지하며 서서히 사용량이 줄고 있는 경향이 유사하다. 중간에 사용량이 높아지는 시기가 있더라도 첫 등장시기의 사용량을 크게 웃돌지 못하고 첫 사용량과 비슷한 수준에 머물렀다. 이 두 그룹 단어들의 사용량은 4년 만에 첫 등장 시기보다 오히려 감소하였다. 고빈도 신조어에 대한 이러한 유형 구분은 [5]에서 고빈도 신조어를 상승, 상승 후 점차 하강, 꾸준 등장 등으로 유형 구분한 것과 비슷한 결과이다.

그림 2는 일반 신조어와 전문 분야 신조어의 사용량 변화를 비교하고 있다. 총사용량이 많은 일반 신조어(Gen20)와 전문 신조어(Term20)를 각각 상위 20개씩 선정하여 비교에 사용하였다. 절대적인 사용량에 있어서 일반 단어가 전문 단어보다 훨씬 높게 나타난다. 일반 단어는 첫 등장 이후 대체적으로 꾸준히 사용량이 증가하는 경향을 보이는 반면, 전문 단어는 첫 등

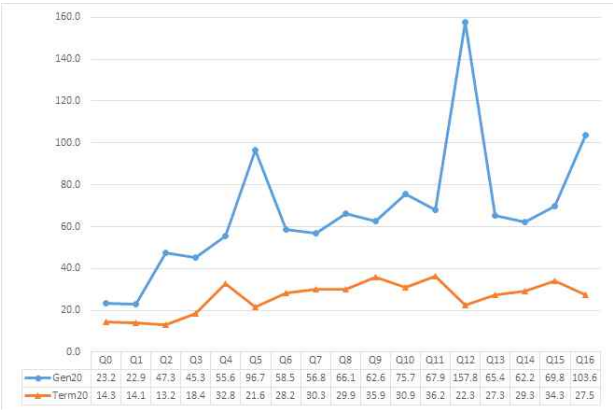

그림 2. 일반 단어와 전문 단어 사용량 변화
Fig. 2. Usage change of general words and terminology

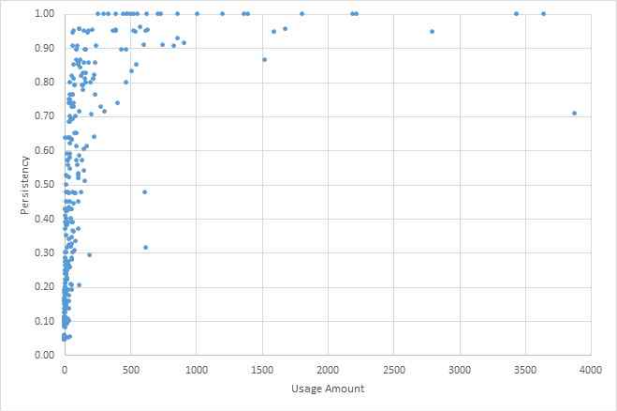

그림 3. 신조어 총사용량과 사용 지속도 사이의 관계
Fig. 3. Relationship between usage amount and persistency of new words

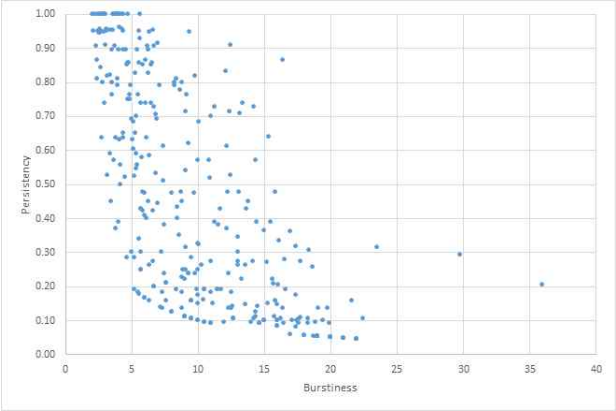

그림 4. 신조어 사용 편중도와 사용 지속도 사이의 관계
Fig. 4. Relationship between usage burstiness and persistence of new words

장 후 완만하게 사용량이 증가한 이후 그 사용량 수준을 유지하는 경향을 보인다. 첫 등장 시기 기준, 4년 만에 일반 단어는 346% 증가한 것에 비해 전문 단어는 92% 증가하였다. 신문 기사를 대상으로 분석한 것이므로 일반 신조어에 비해 전문 신조어의 사용은 제한적인 것으로 판단된다. 또한 전문 신조어에 비해 일반 신조어는 사용량이 증가할 때 상당히 큰 폭으로 증가하는 양상을 보였다.

3-3 신조어 사용 지속도

그림 3은 신조어의 총사용량과 사용 지속도의 관계를 보여준다. 가로축은 총사용량이고 세로축은 사용 지속도이다. 총사용량이 매우 높은 단어는 대부분 지속도도 매우 높았다. 대체적으로 총사용량이 높을수록 사용 지속도 또한 높은 경향을 보이고 있다. 총사용량이 높지 않은 단어들은 사용 지속도가 다양하게 분포하고 있다.

3-4 신조어 사용 편중도

그림 4는 신조어의 사용 편중도와 사용 지속도 사이의 관계를 보인다. 가로축이 편중도이고 세로축이 지속도이다. 편중도가 매우 높은 단어들은 지속도가 낮았다. 즉 일시적으로 많이 사용되는 신조어는 지속적으로 사용되지 않는 경향을 보인다. 또한 편중도가 낮은 신조어는 지속도가 높았다. 즉 사용량이 어느 한 시기에 치우치지 않게 사용되는 신조어가 지속적으로 사용되는 경향을 보인다. 중간 수준의 편중도에서는 지속도가 혼재되어 나타났다.

표 6은 표 5의 총사용량 높은 단어 그룹별 사용 지속도와 편중도 값을 보인다. Avg. Persistency는 각 그룹 소속 단어들의 지속도를 평균한 값이고, Avg. Burstiness는 편중도를 평균한 값이다. 총사용량이 가장 높은 그룹 G1에서 높은 지속도와 낮은 편중도를 보였다. 대체적으로 지속도가 높을수록 편중도가 낮을수록 총사용량이 높다고 볼 수 있다.

표 7은 그림 4의 편중도와 지속도의 분포를 기준으로 표 5의 단어들을 4가지 유형으로 분류하였다. BHPH는 편중도가 모든 단어의 편중도 평균(10.33) 이상이고 지속도가 모든 단어의 지속도 평균(0.43) 이상인 경우를 뜻한다. BLPH는 편중도는 평균 미만이지만 지속도는 평균 이상인 경우를, BLPL은 둘다 모두 평균 미만인 경우를, BHPL은 평균 이상의 편중도와 평균 미만의 지속도를 갖는 경우를 의미한다. 총사용량이 가장 높은 4개의 단어 그룹에서 모두 편중도는 낮고 지속도가 높은 단어들(BLPH)이 대부분을 차지하였으며, 총사용량 높은 상위 120개 단어 중 79.2%가 BLPH, 14.2%가 BHPH이었다.

그림 5는 총사용량 상위 120개 단어에 대해 각 유형의 사용량 변화를 보인다. 편중도와 지속도가 모두 높은 단어들(BHPH)은 편중되는 시기에 급격히 사용량이 높아지는 패턴을 꾸준히 보인다. 이에 비해 편중도가 낮고 지속도는 높은 단어들

표 6. 사용량 단어 그룹별 사용 지속도와 편중도
Table 6. Persistency and Burstiness according to word groups

Group	Avg. Persistency	Avg. Burstiness
G1	0.92	5.93
G2	0.89	6.03
G3	0.73	8.44
G4	0.64	8.24

표 7. 사용 편중도와 지속도에 따른 단어 그룹
Table 7. Word groups according to burstiness and persistency

Group	BHPH	BLPH	BLPL	BHPL
G1	4	25	0	1
G2	4	26	0	0
G3	3	25	0	2
G4	6	19	0	5
TotalNumWords	17(14.2%)	95(79.2%)	0	8(6.7%)

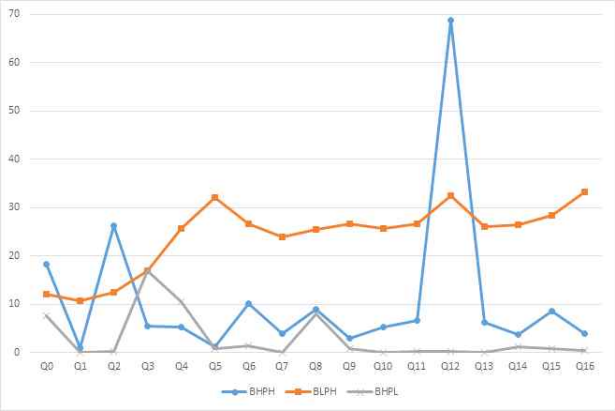

그림 5. 편중도 및 지속도에 따른 사용량 변화
Fig. 5. Usage change according to burstiness and persistency

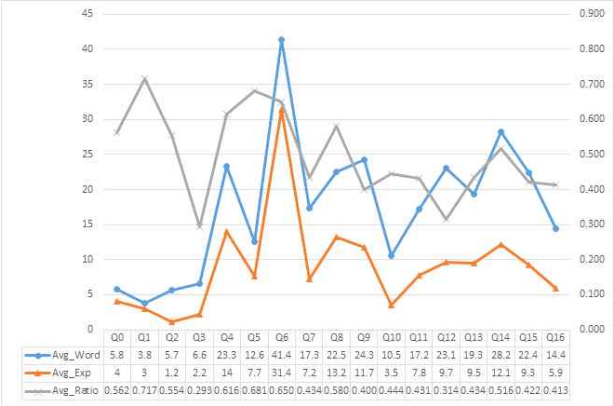

그림 6. 신조어와 신조어 설명의 사용 경향 비교
Fig. 6. Comparison of usages between new words and descriptions

(BLPH)은 첫 사용 후 줄곧 높은 사용량을 유지한다.

표 8은 전체 신조어를 편중도와 지속도를 기준으로 4가지 유형으로 그룹핑하여 단어 개수(Number of Words), 평균 사용량(Avg. Amount), 총사용량(Total Amount)를 보인 것이다. 신조어의 37.9%에 해당하는 BLPH 유형(예: '박스피', '어깨 깡패', '대프리카')이 총사용량의 79.2%를 차지하고 있고, 5.3%에 해당하는 BHPH 유형(예: '트랜짓몰', '두둠칫', '혼자녀')이 사용량의 14.6%를 차지하여 이 두 가지 유형의 신조어들이 전체 사용량의 93.8%를 차지하고 있었다. 이에 비해 신조어의 37.3%를 차지하고 있는 BHPL(예: '호캉스족', '꼬돌남', '땡큐 골') 유형은 사용량의 4.9% 정도만 차지하고 있다. 지속성이 총사용량에 큰 영향을 미치고 있었으며, 높은 편중성도 사용량 증가에 영향을 미칠 수 있으나 그러려면 지속성이 높아야 했다.

3-5 신조어 설명 사용 경향

380개 신조어 중에서 문서에 단어와 그 설명이 같이 등장했는지 파악하기 용이한 단어 10개를 표 9와 같이 선정하여 이 단어와 이 단어에 대한 설명이 어떻게 사용되는지 특징을 분석하였다. 그림 6은 표 9의 신조어와 그 신조어의 설명이 어떻게 사용되고 있는지 신조어의 첫 사용시점부터의 변화를 보이고 있다. Avg_Word는 단어 10개의 평균 사용량을, Avg_Exp는 각 단어와 그 단어의 설명이 동시에 사용된 평균 사용량을, Avg_Ratio는 각 단어 설명 사용 비율의 평균값이다. 단어 설명 사용 비율은 (단어와 단어 설명이 동시에 사용된 뉴스 기사 건수) / (단어가 사용된 뉴스 기사 건수)로 계산된다. 단어가 사용된 기사 건수가 0인 시기에는 단어 설명 사용 비율을 계산할 수 없으므로, 해당 시기의 해당 단어의 설명 사용 비율은 평균값 계산에 영향을 미치지 않는다.

신조어 사용 초반에는 약 72%까지 신조어와 설명이 같이 사용되었으나 점차적으로 설명이 같이 사용되는 비율이 줄어들고 있다. 4년 남짓이 지난 이후에는 약 41% 수준으로까지 줄어들어 설명 없이 신조어만 사용되는 경우가 더 많아지는 경향을 보인다. 점차 설명 사용 비율이 줄어들고는 있으나 아직 완전히 단어로 편입되기 전까지는 여전히 설명이 신조어와 함께 사용되고 있는 것으로 판단된다.

표 8. 사용 편중도와 지속도에 따른 단어 그룹
Table 8. Word groups according to burstiness and persistency

Group	Number of Words	Avg. Amount	Total Amount
BHPH	19(5.3%)	484.4	9204(14.6%)
BLPH	136(37.9%)	366.9	49902(79.2%)
BLPL	70(19.5%)	12.3	858(1.4%)
BHPL	134(37.3%)	22.9	3068(4.9%)

Ⅳ. 결 론

본 논문은 신조어의 분석적 연구를 위하여 사용량, 사용 지

표 9. 신조어 설명 사용 경향을 파악할 단어 10개
Table 9. Ten words to analyze the word description usage

Rank	Word	Amount
48	최애캐	556
50	빠던	540
51	셀기꾼	529
56	덕통사고	477
58	캠프닉	469
64	헌웃	389
76	일코	225
81	맛저	188
85	벨튀	167
89	베댓	159

속도, 사용 편중도 등의 척도를 제안하였고, 이를 바탕으로 뉴스 기사에서 신조어의 사용 통계를 수집하여 신조어의 사용 특징을 분석하였다. 총사용량 상위 20개 신조어가 전체 신조어 사용량의 약 94%를 차지하는 치우친 분포를 보였으며, 전문 분야의 신조어보다는 일반 신조어의 사용량이 훨씬 더 많았다. 반면 전체 신조어의 약 33% 정도의 단어가 10번 이하의 낮은 총사용량을 기록했다. 총사용량이 많은 단어들은 증가형, 증가 후 유지형, 점진 감소형 등으로 사용량 변화의 유형을 구분해 볼 수 있었다. 또한 신조어의 사용 지속도가 총사용량에 큰 영향을 미쳤는데 지속도가 높을수록 사용량이 높은 경향을 보였다. 일부 시기에만 집중적으로 많이 사용되는 편중도가 높은 단어는 지속도가 낮았으며, 편중도가 낮은 단어는 지속도가 높은 경향을 보였다. 총사용량이 높으려면 지속도가 높아야 했으며, 지속도가 높은 경우 편중도도 높으면 총사용량 또한 높았다. 신조어 총사용량의 대부분은 지속도가 높은 단어들이 차지하고 있었다. 신조어 사용 초반에는 신조어 사용 시 해당 신조어에 대한 설명이 같이 사용되었다가 점진적으로 설명 병기 비율이 줄어드는 경향을 보였으나, 첫 사용 후 4년여가 흐른 뒤에도 40% 정도는 신조어와 설명이 같이 사용되었다. 본 연구 결과로 도출된 신조어의 사용 특징들은 향후 신조어 자동 추출 시스템 개발이나 신조어 조사 연구 등에 응용될 수 있을 것이다. 향후 더욱 많은 신조어를 대상으로 분석을 확대하고, 뉴스 기사 범주에 따른 신조어의 종류나 빈도의 차이점을 분석할 필요가 있다.

참고문헌

[1] N. H. Cho, *Survey on New Words*, The National Research Institute of the Korean Language, 1994.

[2] H. S. Kim, *New Words of 2005*, The National Institute of the Korean Language, 2005.

[3] K. I. Nam, *New Words of 2014*, The National Institute of the Korean Language, 2014.

[4] S. Y. Lee, "A Study on Stabilization of New Words in Korean," *Korean Semantics*, Vol. 24, pp. 175-195, December 2007.

[5] H. D. Jung, "Investigation into 'Life of Neologisms' - Neologisms of 2002 ~ 2004 Year in KINDS," *Journal of Korea Linguistics*, Vol. 83, pp. 119-152, September 2017.

[6] K. I. Nam, "An Analysis on the Usage Change of Neologisms and the Criteria for Headwords in the Dictionary − Focused on 2005·2006 Neologisms," *Han-Geul*, Vol. 310, pp. 205-233, December 2015.

[7] K. I. Nam, S. J. Lee, and J. Choi, "Changes in the Use of Neologisms Based on a Largescale Webcrawled Corpus an Investigation of the Current State and Related Issues," *Journal of Korealex*, Vol. 29, pp. 72-106, May 2017.

[8] I. H. Kim, "Generations of Neologism and Its Settlement − Focusing on New Nouns in the Daily Newspaper," *Journal of Korealex*, Vol. 24, pp. 98-125, November 2014.

[9] H. D. Jung, "A Study on Formation and Frequency Change of Korean Neologisms: Focused on Neologisms in 2004," *Han-Geul*, Vlol. 310, pp. 171-204, December 2015.

[10] D. E. Kim and S. S. Lee, "Design and Implementation of Meaning Collecting Tool for New Words," *Journal of KISS: Software and Applications*, Vol. 40, No. 12, December 2013.

[11] Naver Developers, Introduction to Search API [Internet], Available: https://developers.naver.com/products/search/.

[12]Naver, Naver News Search [Internet], Available: https://search.naver.com/search.naver?where=news.

한경수(Kyoung-Soo Han)

1998년 : 고려대학교 컴퓨터학과 (학사)
2000년 : 고려대학교 대학원 (이학석사)
2006년 : 고려대학교 대학원 (이학박사-전산학)

2006년~2009년: SK텔레콤
2009년~현 재: 성결대학교 컴퓨터공학부 교수
※관심분야 : 정보검색, 질의응답시스템, 텍스트 마이닝 등