

20세기 초 신어의 현대 일본어에서의 정착에 관한 연구*

-사전, 코퍼스, 구글 검색엔진을 활용하여-

이 현 정**

object35@hotmail.com

〈 目 次 〉

- | | |
|------------|---------------------------------|
| 1. 머리말 | 3.1 사전과 코퍼스의 등재 여부 분석 |
| 2. 조사개요 | 3.2 구글 검색건수에 따른 조사대상어의 구간별 분포양상 |
| 2.1 조사자료 | 3.3 정착범주와 미정착범주의 대상어 |
| 2.2 조사방법 | 4. 맺음말 |
| 3. 결과 및 분석 | |

Key word : 20世紀初め(The early 20th century), 新語(New words), 現代日本語(Modern Japanese), 語種 (Types of words), 定着(Settlement)

1. 머리말

언어와 사회는 밀접한 관계에 있다. 사회에 변화가 생겨나면 그에 따른 새로운 단어가 유입되거나 만들어진다. 일본의 경우 19세기 개국과 함께 서양 문물과 사상이 쏟아졌고, 그와 함께 외국어가 들어오기 시작했는데 그 대부분은 영어였다. 일본에서는 이들 외국어를 새로운 한자어로 번역하거나 그대로 받아들이기도 하였다. 이렇게 생겨난 신어는 시간이 흐르면서 계속 사용되어 더 이상 신어가 아닌 일반어로 남기도 하고 사어(死語)화 되기도 한다. 본

* 이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (2015S1A5B5A07044434)

** 부경대학교 일어일문학부 시간강사, 일본어학

연구는 실제로 일본의 20세기 초 신어¹⁾가 현대에 이르러 얼마나 사용되고 있는지 혹은 사용되지 않고 있는지 파악하는데 목적을 두고 있다.

신어에 관한 기존의 연구는 대체로 신어의 유형 및 생성 이유, 만들어지는 방법 등이 주를 이루고 있다(加茂1944, 稲垣1982, 鈴木1988, 米川1989,1992등). 그러나 20세기 초 신어가 현대 일본어 어휘 속에 얼마나 남아서 사용되고 있는지 분석한 연구는 찾아보기 어렵다. 단어의 정착과 관련한 연구로는 양민호(2013)가 있는데 이 논문에서는 최근에 사회적으로 회자되거나 화제가 되는 외래어를 대상으로 한국어와 일본어 내에서의 정착 단계에 대해 대조하였다. 정착과 미정착의 구별 기준으로는 구글 검색건수를 토대로 상위 약 10%를 정착단계, 하위 약10%를 미정착 단계로 나누고 있다. 그러나 양민호(2013)에서도 언급하고 있듯이 어휘의 정착과 미정착을 단순히 검색건수로 판단하고 입증하는 것은 쉽지 않다는 점, 정착 정도를 살피기 위해서는 보다 다양한 조사가 필요하다는 점은 정착 연구에서 보완하여야 할 과제라고 판단된다.

이에 따라 본 연구에서는 좀 더 다양한 분석을 통해 신어의 정착을 살펴보고 한다. 20세기 초 신어를 대상으로 사전, 코퍼스, 구글 검색에서의 등재여부 및 검색건수를 조사하고 이를 바탕으로 현대 일본어에서의 정착여부를 파악하고자 한다. 동시에 신어의 어종에 따른 특징이 나타나는지에 대해서도 살펴보고자 한다.

2. 조사개요

2.1 조사자료

1) 조사자료 및 조사대상어의 선정

본 연구의 자료로는 『ポケット顧問 や、此は便利だ』²⁾를 선정했다. 『ポケット顧問 や、此は便利だ』는 일본의 첫 신어사전으로 제1편 신문어해설(新

-
- 1) 본 연구에서의 신어는 신조어와 유행어를 포함하는 넓은 의미의 신어를 뜻한다.
 - 2) 『ポケット顧問 や、此は便利だ』의 초판은 1914년 4월에 발행되었고, 1919년 7월 증보개정판이 발행되었다. 본 논문에서는 증보개정판을 사용하였다.

聞語解説), 제2편 실용숙어성구편람(實用熟語成句便覽), 제3편 실용문자편람(實用文字便覽)으로 나누어져 있다. 사전의 예언(例言)에서는 신문 잡지에서 이해하기 어려운 어(語)나 전문적인 용어는 채택하지 않고 현대인의 일상생활에 필요하다고 생각되는 어(語)·구(句)·문자(文字)를 채택하고 있다고 기술하고 있다. 제1편은 다시 세 부류로 나뉘는데 첫 번째가 최신의 술어(術語) 및 유행어, 두 번째가 통상의 번역어·외래어·신의어(新意語), 세 번째가 정면에서는 의미를 알 수 없는 현대식 전용어(轉用語)이다.

본 연구에서는 가장 신어의 경향을 보이는 것으로 판단되는 제1편의 최신의 술어(術語) 및 유행어에 실린 표제어를 조사대상어로 삼았다. 대상자료의 표제어 중 머리글자(頭文字)와 알파벳을 포함하는 것, 인명·지명·수사 및 이것을 포함하는 것, 『新しい女, 美的生活』와 같이 구의 형식을 갖는 것들도 모두 대상으로 하였다. 단지 『海洋の自由, 時間と空間』과 같이 「の」와 「と」를 포함하는 구는 제외하였다. 『予覚³⁾、予感』과 같이 한 표제어 안에 두 개의 단어⁴⁾가 있을 경우에는 두 개의 단어로 처리하였다. 이에 따른 본 연구의 조사대상어는 총 594개였다.

2) 어종의 분류

본 연구는 신어의 사용 양상에 있어서 어종에 따른 특징이 있는지를 파악하기 위해 표제어의 어종을 분류하였다. 이때 어종은 일본고유어(和語), 한자어(漢語), 외래어, 혼종어로 분류하였다. 외국어가 일본의 가타가나로 적혀 사전류에 실린 것은 이미 다른 외국어와는 다른 것으로 판단하여 정착여부와 관계 없이 외래어로 칭하였다.⁵⁾

대상으로 한 신어 중에는 『黒表, 軍国主義』와 같이 한자어를 표제어로 하고 후리가나는 이들의 외래어를 달아놓은 것이 있었다. 이 경우에는 외래어로

- 3) 『予覚』의 경우 조사자료에서는 『豫覚』로 표기되어 있다. 이처럼 조사자료인 『ポケット顧問』や、此は便利だ』에서는 한자의 약자를 사용하지 않고 있으나 본 논문에서는 이들 한자를 현대 일본어에서 사용하는 약자로 표기하고 검색하였다.
- 4) 조사대상어 중에는 구의 형태를 띠는 것도 있으나 본고에서의 단어는 사전의 표제어를 의미하므로 이를 포함하는 넓은 의미로 규정하도록 한다.
- 5) 양민호(2013)의 외래어 정의를 참고하였다.

처리하였다. 또 『青年獨逸 汎亜細亜主義』의 밑줄 부분과 같이 한자로 표기하였지만 외래어로 읽히는 고유명사를 포함하고 있는 경우가 있었다. 이들은 외래어로 처리하여 『青年獨逸 汎亜細亜主義』는 혼종어로 분류하였다. 인명의 경우 『石井ランシング協約』의 『石井』와 같이 일본인의 이름은 일본고유어로, 『モンロー主義』의 『モンロー』와 같이 외국인의 이름은 외래어로 분류하였다.

2.2 조사방법

1) 사전의 선정과 검색방법

조사대상어의 사전 등재여부를 검색하는 사전으로는 『大辭林』 제3판을 선택하였다. 『大辭林』은 일본의 대표적인 국어사전 중 하나로 제3판은 2006년에 발행되었으며 약 23만3천어를 수록하고 있다. 그 이후 종이발행은 없으나 Web 판에서의 증보(增補)가 이루어지고 있다. 인터넷상에서도 그 검색이 가능하므로 본고에서는 야후 재팬 사이트의 사전 검색을 통해 조사대상어가 『大辭林』 제3판에 실려 있는지 확인하였다.

2) 코퍼스의 선정과 검색방법

코퍼스는 BCCWJ(Balanced Corpus of Contemporary Written Japanese, 현대 일본어 글말 균형 코퍼스)를 활용하여 조사대상어의 등재여부를 검색하였다. BCCWJ는 현대 일본어의 글말인 모집단에서 생산실태(출판사)와 유통성(도서관)의 2가지 측면에서 표본을 추출한 2개의 서브코퍼스와 이것만으로는 추출하기 어려운 자료를 중심으로 비모집단 추출인 특수목적 서브코퍼스로 구성되어 있다(장원재 2014). 전체 구축량은 약 1억 400만 단어(짧은 단위)로 출판 서브코퍼스가 약 3400만, 도서관 서브코퍼스가 약 3000만, 특수목적 코퍼스가 약 4000만 단어로 구성되어 있다. 구축대상연도는 출판 서브코퍼스가 2001년부터 2005년까지이고 도서관 서브코퍼스는 1986년부터 2005년까지이며 특수목적 서브코퍼스는 대상 연도가 각각 다르나, 백서, 베스트셀러, 법률, 국회의원 의록은 1970년대 후반부터이다.(장원재 2014). 따라서 BCCWJ는 조사대상어

6) 이하 『大辭林』 제3판을 사전으로 칭하기로 한다.

인 1910년대의 신어가 현대 일본어 안에서 사용되고 있는지 검색하기에 적합한 코퍼스라고 판단하였다. BCCWJ에서 조사대상어의 검색 시에는 「神經質, 自由意志, 正当防衛」 등과 같이 3-4자의 한자어가 많고, 구형식의 긴 단어도 있으므로 장단위로 검색하였다.

3) 구글(Google)검색엔진에서의 검색방법

구글 재팬에서 검색 시, 조사대상어 검색의 정확도를 높이기 위해서 「어순도 포함하여 완전일치」라는 조건검색을 실시하였다. 예를 들면 조사대상어 「客觀描写」의 검색결과, 「内面描写, 心理描写, 客觀的」 등과 같은 다른 의미의 합성어도 함께 검색되었다. 그러나 「어순도 포함하여 완전일치」라는 조건검색을 추가하면 다른 의미의 합성어들이 상당수 배제되어 실제 검색건수는 405,000에서 5,080으로 줄어들었다. 구글 검색건수의 조사는 2018년 4월 초에 실시하였다.

3. 결과 및 분석

3.1 사전과 코퍼스의 등재여부 분석

사전과 코퍼스의 등재여부 분석에 앞서 조사대상어 594개를 어종별로 살펴보면 다음과 같다.

<표1> 조사대상어의 어종구성

일본고유어	한자어	외래어	혼종어	합계
3 (0.5%)	442 (74.4%)	82 (13.8%)	67 (11.3%)	594 (100%)

조사대상어의 어종비율은 「한자어>외래어>혼종어>일본고유어」의 순으로 나타났다. 한자어의 비율이 74.4%로 가장 높았으며 그 대부분은 서양외래어를 한자어로 번역한 것이었다. 일본고유어는 단 3개로, 20세기 초 신어에는 거의

없다고 할 수 있다.

이들 조사대상어가 사전과 코퍼스에는 얼마나 실려 있는지 분석한 결과는 <표2>와 같다.

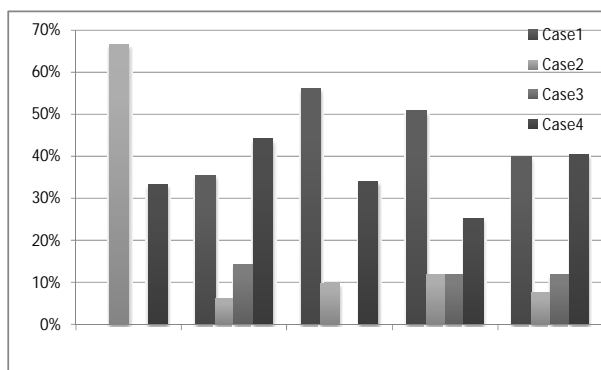
<표2> 조사대상어의 사전과 코퍼스의 등재 비율

		일본 고유어	한자어	외래어	혼종어	합계
大辞林	○	3 (100%)	222 (50.2%)	36 (43.9%)	25 (37.3%)	286 (48.1%)
	×	0 (0%)	220 (49.8%)	46 (56.1%)	42 (62.7%)	308 (51.9%)
BCCWJ	○	1 (33.3%)	258 (58.4%)	28 (34.1%)	25 (37.3%)	312 (52.5%)
	×	2 (66.7%)	184 (41.6%)	54 (65.9%)	42 (62.7%)	282 (47.5%)

먼저, 사전에서는 594개의 조사대상어 중 286개가 실려 있었고, 코퍼스에는 312개가 실려 있었다. 즉, 대상어의 사전과 코퍼스의 등재비율은 약 50% 전후임을 알 수 있다. 어종별로 보면 일본고유어는 그 수가 3개에 불과하기 때문에 등재비율을 비교하는 것은 어려워 보인다. 사전과 코퍼스에 등재되어 있는 한자어의 비율은 각각 50.2%, 58.4%로 높게 나타났고, 외래어와 혼종어의 비율은 30~40%대였다. 사전에는 외래어의 등재비율이 높았으나 코퍼스는 혼종어의 등재비율이 상대적으로 높음을 확인할 수 있었다.

다음은 조사대상어를 사전과 코퍼스 모두 등재되지 않은 대상어(이하 Case1), 사전에만 등재된 대상어(이하 Case2), 코퍼스에만 등재된 대상어(이하 Case3), 사전과 코퍼스 모두 등재된 대상어(이하 Case4)로 분류하여 어종별로 분석하였다. 분석결과를 그림과 표로 나타내면 아래와 같다.

<그림1>과 표<3>를 보면 합계인 「計」는 「Case1」과 「Case4」의 비율이 각각 약 40%이다. 한자어는 「Case4」의 비율이 44.1%로 가장 높은 반면, 외래어와 혼종어는 「Case1」의 비율이 각각 56.1%, 50.8%로 가장 높게 나타났다. 즉, 20세기 초 신어에 해당하는 조사대상어는 현대 일본어의 사전과 코퍼스에 모두 등재되어 있지 않거나 모두 등재되어 있는 경향이 높음을 알 수 있다. 또한, 모두 등재되어 있는 경우는 외래어와 혼종어보다는 한자어의 비율이 높았다.



<그림1> Case별 어종분포7)

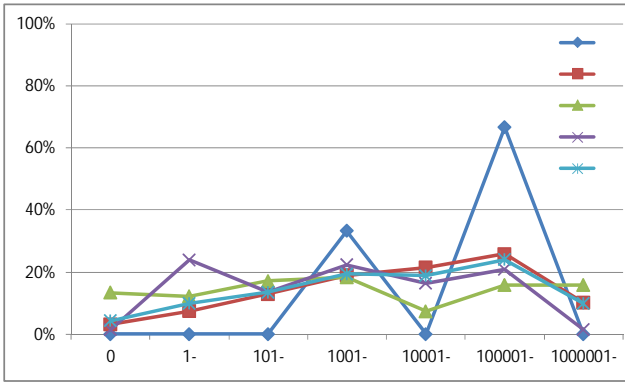
<표3> Case별 어종분포-단어수와 비율

	和	漢	外	混	計
Case1	0 0.0%	157 35.5%	46 56.1%	34 50.8%	237 39.9%
Case2	2 66.7%	27 6.1%	8 9.8%	8 11.9%	45 7.6%
Case3	0 0.0%	63 14.3%	0 0.0%	8 11.9%	71 11.9%
Case4	1 33.3%	195 44.1%	28 34.1%	17 25.4%	241 40.6%
計	3 100.0%	442 100.0%	82 100.0%	67 100.0%	594 100.0%

3.2 구글 검색건수에 따른 조사대상어의 구간별 분포양상

구글에 실려 있는 조사대상어의 검색건수는 0건부터 73,700,000건까지로 상당히 폭넓었다. 이 검색건수를 구간별로 나누어 어종별로 분석한 결과는 다음과 같다. 구글에서 검색되지 않는 조사대상어는 검색건수 0건으로 처리하였다.

7) 이하 그래프와 표에서 일본고유어는 和, 한자어는 漢, 외래어는 外, 혼종어는 混, 합계는 計로 표시함.



<그림2> 어종별 구글 검색건수의 구간별 분포

<표4> 어종별 구글 검색건수의 구간별 분포 - 단어수와 비율-

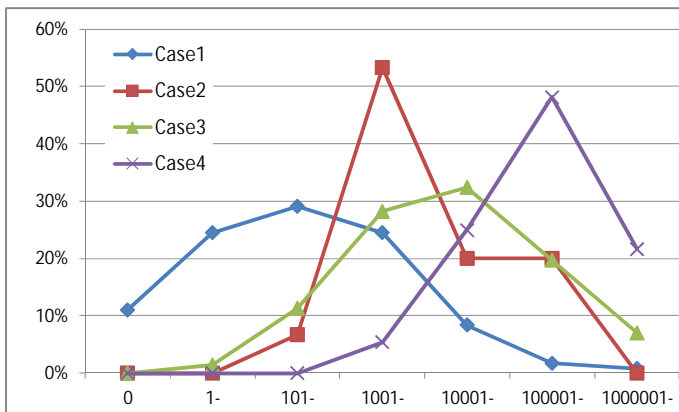
	0	1-	101-	1001-	10001-	100001-	1000001-	計
和	0 0.0%	0 0.0%	0 0.0%	1 33.3%	0 0.0%	2 66.7%	0 0.0%	3 100%
漢	14 3.2%	33 7.5%	57 12.9%	84 19.0%	95 21.5%	114 25.8%	45 10.2%	442 100%
外	11 13.4%	10 12.2%	14 17.1%	15 18.3%	6 7.3%	13 15.9%	13 15.9%	82 100%
混	1 1.5%	16 23.9%	9 13.4%	15 22.4%	11 16.4%	14 20.9%	1 1.5%	67 100%
計	26 4.4%	59 9.9%	80 13.5%	115 19.4%	112 18.9%	143 24.1%	59 9.9%	594 100%

구글 검색건수의 분석 결과, 검색건수가 0건에 해당하는 수는 조사대상어 594개 중 26개로 극히 적었다. 「計」는 구글 검색건수가 높아짐에 따라 대상어의 수가 꾸준히 증가하다가 「10001-」구간에 24.1%로 비율이 가장 높았으며 「100001-」구간에서는 9.9%로 낮아졌다. 단, 「10001-」구간에서는 「101-」구간에 비해 비율이 소폭 감소하였다. 한자어 또한 「計」와 유사한 경향을 나타내며 「100001-」구간에서 25.8%로 가장 높은 비율을 차지하고 있다. 반면, 외래어는 모든 구간에서 7.3%~18.3%의 범위로 큰 차이를 보이지 않았으며, 혼종어는 「1-」, 「1001-」, 「100001-」구간에서 각각 20% 이상의 높은 비율을 나타내었다.

구글 검색건수에 따른 조사대상어의 양상을 좀 더 구체적으로 살펴보기

위해 사전과 코퍼스의 등재여부로 구분한 4개의 Case별 검색건수 구간비율을 분석하였다.

아래의 <그림3>과 <표5> 를 보면 구글 검색건수의 분포는 각 Case별로 뚜렷한 차이를 나타내고 있다. 먼저 구글 검색건수 「0」에 해당하는 26개의 조사대상어는 모두 「Case1」에 속하는 것을 알 수 있다. 「Case1」에는 구글 검색건수 「0」에서 「1001-」사이의 구간에 대상어의 약 90% 가량이 분포하고 있는 반면, 「Case4」의 경우는 「10001-」에서 「1000001-」구간에서 대상어의 비율이 94.6%를 차지하고 있다. 한편 「Case2」에 해당하는 대상어는 45개로 구글 검색건수 「1001-」에서 「100001-」구간에 대상어의 95% 이상이 존재하며 특히 「1001-」구간은 전체의 53.3%의 비율을 차지하고 있었다. 「Case3」의 대상어는 71개로 「Case2」와 유사하게 「1001-」에서 「100001-」구간의 비중이 높았으며 「10001-」구간에서 32.4%로 가장 비중이 높았다. 즉, 사전과 코퍼스에 모두 실려 있지 않을 때는 검색건수가 낮고, 사전과 코퍼스에 모두 실려 있을 때는 검색건수가 높으며, 어느 한쪽에만 실려 있는 경우 검색건수는 중간구간에 분포하는 경향을 나타내고 있다.



<그림3> Case별 구글 검색건수의 구간별 분포

<표5> Case별 구글 검색건수 구간별 분포 - 단어수와 비율

Case	0	1-	101-	1001-	10001-	100001-	1000001-	計
1	26 11.0%	58 24.5%	69 29.1%	58 24.5%	20 8.4%	4 1.7%	2 0.8%	237 100.0%
2	0 0.0%	0 0.0%	3 6.7%	24 53.3%	9 20.0%	9 20.0%	0 0.0%	45 100.0%
3	0 0.0%	1 1.4%	8 11.3%	20 28.2%	23 32.4%	14 19.7%	5 7.0%	71 100.0%
4	0 0.0%	0 0.0%	0 0.0%	13 5.4%	60 24.9%	116 48.1%	52 21.6%	241 100.0%
計	26 4.4%	59 9.9%	80 13.5%	115 19.4%	112 18.9%	143 24.0%	59 9.9%	594 100%

이상의 결과로부터 정착과 미정착을 명확히 단정 짓기는 어렵지만 그 경향을 살펴볼 수는 있다. 우선 사전과 코퍼스에 실려 있지 않을 때는 구글 검색건수가 낮고, 사전과 코퍼스에 모두 실려 있을 때는 구글 검색건수가 높은 점을 감안할 때 「Case4」 중에서도 비율이 높은 「100001-」구간은 정착범주에 있는 단어일 가능성이 높다고 할 수 있다. 또한, 사전과 코퍼스에 모두 실려 있지 않은 경우 대상어의 대부분은 구글 검색건수 「10001-」구간 이하에 존재했고, 사전과 코퍼스에 모두 실려 있을 경우 「1001-」이하 구간의 비율이 5.3%에 불과하다는 것은 구글 검색건수 10,000 이하는 정착범주에 있지 않는 단어임을 의미한다고 볼 수 있다. 그러나 사전에만 실려 있는 「Case2」와 코퍼스에만 실려 있는 「Case3」의 「1001-」이하 구간의 비율을 고려한다면 정착하지 못한 단어는 구글 검색건수 1,000이하로 구분 지을 수 있는 것으로 판단된다.

3.3 정착범주와 미정착범주의 대상어

1) 정착범주의 대상어

「3.1」과 「3.2」의 결과를 토대로 20세기 초 신어에 해당하는 조사대상어의 정착범주는 사전과 코퍼스에 실려 있고 구글 검색건수가 「100001-」구간 이상에 속하는 것으로 구분 지었다. 이에 해당하는 조사대상어는 「100001-」구간 116개와 「1000001-」구간 52개로 총 168개이다. 구간별로 대상어를 제시하면 아래와 같다.

【「Case4」의 구글 검색건수 「100001-」구간 116어 -오름차순】

知識階級,センチメンタリズム(sentimentalism),デカダン(Décadent),二科会,産業組合,自由劇場,緩衝地帯,優先株,美顔術,プロレタリアート(Proletariat),重商主義,制海權,事大主義,三国同盟,象徴主義,物質主義,普通選挙,人本主義,危険思想,パナマ運河,水上飛行機,海上保険,法定利息,展望車,簡易保険,無政府主義,人間嫌い,立体派,早教育,時代精神,娘子軍,国家総動員,信用状,再保険,官僚政治,粗製濫造,白兵戦,国家主義,散文詩,功利主義,平和運動,近代思想,戒嚴令,生存競争,悲喜劇,不当利得,抗毒素,戒告,街路樹,副甲状腺,意志薄弱,治外法權,級数,精神分析,産業革命,正当防衛,待命,美的生活,自由港,二重生活,優性,キュービズム(Cubism),潛在意識,強迫観念,執行猶予,社会主義,意思表示,禁輸,帝国主義,進化論,共產主義,エキゾチック(Exotic),資本主義,無尽,精神療法,機會均等,神秘主義,水先案内,南京事件,買占,軍国主義,ジャーナリズム(Journalism),共和党,地方分權,社会政策,神經質,失語症,世紀末,個人主義,世界經濟,領海,プロテスタント(Protestant),特別會計,中央集權,制空權,^{ブラックリスト}黒表(Blacklist),自由意志,信用組合,危険人物,陰謀,自然淘汰,保護色,エレクトロン(Electron)⁸⁾,自然主義,閥,クーデター(Coup d'État),二重人格,シンプルライフ(Simple life),外国為替相場,労働組合,無意識,古典主義,劣性,染色体,會計士,デモクラシー(Democracy)

「100001-」구간에서는 대상어 116개 중 93개가 한자어로 이 구간 대상어의 대부분을 차지하고 있는 가운데 「會計士」는 구글 검색건수 949,000건으로 한자어 중에서 가장 높은 검색건수를 기록하였다. 「會計士」는 현대 일본어 안에서 일반적으로 사용되는 단어이다.

이 구간에서는 「知識階級, 二科会, 自由劇場, 緩衝地帯」 등과 같은 구글 검색건수 100,000건에 가까운 한자어보다는 검색건수 1,000,000건에 가까운 「危険人物, 二重人格, 保護色, 無意識, 染色体」 등의 한자어가 좀 더 일반적으로 사용되고 있는 단어임을 알 수 있다. 즉, 검색건수가 높아짐에 따라 정착도가 높은 단어들이 분포하는 경향이 나타나는 것으로 여겨진다. 또한 한자어 중에서는 「-主義」의 합성어가 17개로 구간 전반에 걸쳐 나타나고 있다.

외래어를 살펴보면 「ジャーナリズム, クーデター, シンプルライフ」 등은

8) 「エレクトロン」은 신어사전에서 알파벳표기가 「Electron」로 되어 있는데 표기오류일 가능성이 있으며 현재는 「Elektron」을 사용하고 있다.

대체할 수 있는 다른 단어가 없다고 할 수 있다. 반면 「デカダン, プロレタリアート, キュービズム, プロテスタント, エレクトロン」 등은 외래어 외에 이와 같은 의미의 한자어도 존재한다. 외래어와 한자어의 구글 검색건수를 비교해보면 「デカダン」의 경우 구글 검색건수가 103,000건인데 반해 한자어인 「退廢的」의 구글 검색건수는 407,000건으로 외래어보다 한자어가 많았다. 「프로테스타ント」와 「에レクト론」의 경우도 한자어인 「新教」와 「電子」의 구글 검색건수가 많았다. 반면에 한자어보다 외래어의 구글 검색건수가 많은 대상어도 있었다. 「프로레タリアート」의 한자어는 「無産階級」로 「프로레タリアート」의 구글 검색건수는 114,000건이었으나 「無産階級」는 39,400건에 그쳤다. 또 「キュービズム」는 한자어 「立体派」보다 구글 검색건수가 많았다. 이들 단어는 외래어와 한자어 중 어느 한쪽의 사용이 더 많거나 적을 수는 있으나 현대 일본어 안에서 양자 모두 사용되는 것으로, 어느 한쪽이 온전한 단어가 될 수는 없는 외래어와 한자어의 공존체계에 있다고 할 수 있다.

또한 이 구간에서 고유어는 단 1개로 「買占」가 이에 해당했다. 「買占」는 「Comer」의 번역어로 「買い占め」로도 표기되며 증권, 금융용어로 사채기를 의미한다. 이 구간 혼종어는 10개로 대부분 「優先株, 水先案内, 信用組合, 外国為替相場, 労働組合」 등의 한자어와 고유어의 조합을 이룬 합성어였다.

다음으로 「Case4」의 구글 검색건수 「1000001-」구간에 해당하는 단어를 제시하면 아래와 같다.

【「Case4」의 구글 검색건수 「1000001-」구간 52어 -오름차순-】

時代錯誤, 보이콧(boycott), 變質者, ハンディキャップ(Handicap), 第三帝国, 甲状腺, 予備知識, 副腎, ダンピング(Dumping), 人間味, 安全地帯, 社会問題, 代用品, 林間学校, 姓名判断, プロパガンダ(Propaganda), 教書, 万有引力, 人口統計, 腹式呼吸, 千里眼, 無線通信, 排水量, エスペ란트(Esperanto), 共和, 色盲, 内分泌, トラスト(trust), 社会教育, 植民, ルネッサンス #9), 恐慌, 三人称, 借款, 三位一体, 予感, 都市計画, 白晝, 副作用, 健康診断, オペラ(Opera), 紫外線, 通信販売, 禪, マラソン(Marason), ホルモン(Hormon), 予告, 財団法人, エキストラ(Extra), 著作権, モード(Mode), 統計

- 9) 외래어의 경우는 원어를 함께 표기하였는데 신어사전에 원어가 없는 경우는 #을 표기하였다. 또한 신어사전에는 한자어도 원어가 표기되어 있었으나 본고에서는 지면관계상 외래어만 제시하였다.

52개 단어 중 한자어가 40개를 차지했다. 2자 한자어가 11개, 3자 한자어가 11개로 2, 3자 한자어가 많았고, 그 의미를 쉽게 알 수 있는 단어가 대다수이다. 「予感」은 「2.1.1」에서 언급한 바와 같이 「予覺」와 한 표제어 안에 있는 단어이다. 「予覺」의 경우 사전과 코퍼스에 모두 등재는 되어있으나 구글 검색건수가 6,320건으로 상당히 낮은 편에 속한다. 이에 반해 「予感」은 구글 검색건수가 무려 9,330,000건으로 「予覺」보다 「予感」의 사용이 두드러진다.

이 구간의 한자어 중에서 조금은 생소하게 느껴질 수 있는 것은 「林間學校」와 「姓名判斷」 정도로 여겨진다. 「林間學校」는 시내에서 조금 떨어진 산림 속에서 아이들에게 자연을 느끼게 해주는 학교로, 현재는 방학동안 산림 속에서 심신을 단련하고 단체 활동을 배우는 목적의 교육활동 및 그 시설을 일컫는다. 「姓名判斷」은 사람의 성명으로 그 사람의 운명을 판단하는 것을 말한다. 이처럼 한자어만 봤을 때는 의미를 대략적으로 떠올릴 수 있으나 확인을 위해 사전의 의미를 살펴보면 쉽게 이해할 수 있는 단어이다.

한자어 40개를 제외한 단어 12개는 모두 외래어로 친숙한 단어들이 대부분이다. 「モード」는 구글 검색건수 34,900,000건으로 외래어 중에서 검색건수가 가장 높았으며 다른 대체어가 없다. 이 외에도 「ボイコット, オペラ, マラソン¹⁰⁾, ホルモン¹¹⁾」 등 또한 대체할 수 있는 다른 단어가 없는 외래어이다. 이처럼 Case4의 「100001-」구간의 대상어는 현대 일본어 안에서 일반적으로 사용되고 있는 것으로 판단된다.

2) 미정착범주의 대상어

「3.1」과 「3.2」의 결과를 토대로 20세기 초 신어에 해당하는 조사대상어 중 정착되지 못한 단어는 구글 검색건수가 1000건 이하인 것으로 구분 지었다. 이에 해당하는 조사대상어는 총 165개이다. 이 중 「0」구간에 해당하는 26어는 다음과 같다.

【「0」구간 26어】

經濟的白禍, 國際擁護連動, 死刑廢止同盟會, 自筆記錄器, 紙幣勘定器, 曙光利用

10) 「マラソン」은 조사자료의 철자가 「Marason」으로 되어 있었는데 해설을 보면 현재의 「Marathon」의 의미로 표기의 오류일 가능성이 있다.

11) 「ホルモン」의 철자는 「Hormon」외에 「Hormone」으로도 사용되고 있다.

法, 生物測定率, 絶糧策, 腺増殖, 中継内閣, 蓄勢電話, 徴兵撤廃運動, 全露主義, 無償無合併, ジャック・ジョンソン(Jack Johnson), スツールム・ウンド・ドラング(Sturm und Drang), 세칸드·핸드·나レッジ(Second hand knowledge), 다이내마이즘(Dynamitism), 데ッド·타입(Dead-type), 하우스#, 플랜셋(Planshet), 프레츨리슴(Fretulism), 보지슴# , 포ッシ슴(Possimism), 모더너·린#, 陸上드레드노트

「0」구간 26어는 사전과 코퍼스에도 실리지 않고 구글 검색건수가 0건이었다. 한자어의 경우 5자 이상으로 이루어진 것이 많고 외래어도 합성어나 구의 형태가 있음을 볼 수 있다. 이 중에는 실제로 구글 검색건수가 0건이 아니었으나 0건으로 처리한 것이 있는데 「腺増殖, 中継内閣, ジャック・ジョンソン, 데ッド·타입, 하우스, 플랜셋(Planshet)」 등이 이에 해당한다. 「腺増殖」는 3,700건이 검색되었지만 「脂腺増殖症」이 대다수였고 「腺増殖」만으로 사용되는 사례는 없었다. 「中継内閣」역시 검색건수가 1,540건이었으나 「中継内閣」만으로는 사용되지 않고 있었다. 외래어 중에서 「ジャック・ジョンソン」은 조사자료에서의 의미를 보면 독일의 포탄명이나 구글 검색결과에서는 유명무지선 뿐이었다. 또한 「데ッド·타입」은 고공에서 지상을 향해 급각도로 하강하는 것을 뜻하나 구글 검색결과는 「포올데드타입」로 전부 3M사의 제품이였다. 이러한 경우의 외래어도 모두 0건으로 처리하였다. 이처럼 표제어 단독으로 사용되지 않거나 조사자료의 해설과는 다른 의미로 검색되는 자체가 정착되지 못했음을 의미한다고 할 수 있겠다.

「스투름·운드·드랑」의 경우는 구글 검색건수 0건이었으나 표기 문제로 「슈투름·운드·드랑」로 검색 시에는 8,190건이었다. 현대 일본어 안에서는 「스투름·운드·드랑」대신 「疾風怒濤」라는 한자어가 정착하였다고 볼 수 있다. 이 외의 「經濟的白禍, 自筆記録器, 蓄勢電話, 보지슴, 陸上드레드노트」 등은 검색건수 0건으로 이 구간의 대상어는 현대 일본어 안에서 정착하지 못했다고 할 수 있겠다.

구글 검색건수 1건에서 1000건 범위의 대상어는 아래와 같다.

【구글 검색건수1~1000 139어 - 케이스별 오름차순-】

独逸十戒, 헤르다條約, 臨海移住,ウイルソン十四個條, 茶食主義, 独逸東進, ネオ·

로맨티즘(Neo-romanticism), 八時間労働運動, यू-세닉스(Euthenics), 華想主義, 아·로·오·주·문·제, 覇制主義, エフィシエンシー(Efficiency), 乙種興行, 生命人造說, マルコニー事件, 出獄人保護協會, 査問会, ヘックレリズム(Hackelism), 小口商品券, 万国仲裁裁判, 國際的デモクラシー, 休日植民, 死骸会社, チェック・スロバック#, 援段方針, 国民保險協會, レントゲン・レー#, 南進乎北進乎, 半官新聞, ビザンティン式建築, ブリューストッキング(Blue stocking), リヒテル事件, 輸入防濁, トンチン法, 哲学復古, 土地錯綜, マンナリズム(Mannerism), 航空機射撃砲, 三益主義, サンチカリズム(Syndicalism), 事務簡潔, カールテル(Carutell), 実証說, 石井ランシング協約, 広軌問題, 犯罪人定型, 聯想診断, 乃木問題, プレストリトウスク条約, 青年獨逸, 減食主義, 絶食同盟, 空雷機, 新東洋主義, ルスタニア事件, 母性中心說¹²⁾ \ 船腹問題, 潜在活動, 示意運動, バハイズム(Bahaism), コロナ説, 児童虐待防止会, 無痛安産法, ワイルス氏病, 一国一票主義, 潜航艇戰, 戦争赤ん坊, 常識宗, 間送電報, 一人一黨主義, ヴォルシェヴィキ(Bolsheviki), 青白政府, 救貧問題, 側面結婚, 回読会, デシマルシステム(Decimal system), 哲人主義, ウバニシャッド#, 手形引受業者, ガール・スカウト(Girl-scouts), 空中郵便, 過度文化, 本能満足主義, ゲーリー・システム(Gealy-System), 一人一票主義, 教養過度, 汎亜細亞主義, 女性中心說, 白禍論, 宙帰り, ジャコビニズム(Jacobinism), マンモニズム(Mammonism), 感情移入說, ユーゼニクス(Eugenics), 販売要点, 唯心說, 不老長寿論, 自働書記, 空軍局, ノースクリップ#, 善政主義, 懷疑思想, 營養不良, 全獨逸主義, 勤勞主義, 商用略語, 遊蕩文学, ノーベル賞金, 金貨流出, 物価調節, 原価計算組織, モダーニズム#, 人口増殖率, スペース・ブローカー(Space broker), 国防充実, 赤衛兵, ボーイ・スカウト(Boy scouts), 排貨運動, 米券倉庫, 光筆版, 食餌療養, 少年義勇団, 田園都市運動, 種族發生, 主婦同盟//主知說, ディアテルミー(Diathermy), 制限外発行//本的生活 \ 自我實現說, 唯物說, 今日主義, 宿命說, 暫定条約, 産児制限論, 非募債主義, 自給自足主義

139개의 단어 중 무려 127개가 「Case1」에 해당하였고, 「Case4」의 단어는 없었다. 이 구간의 대상어는 의미를 찾아보아야 그 뜻을 알 수 있는 것들이 대다수이다. 139개 중 한자어는 90개에 해당하였는데 그 중 「-主義」가 18개, 「-說」가 10개나 되었다. 외래어는 총 24개로 그 중 「-ism」을 포함하는 단어가 10개였

12) 139어를 각 케이스별 구간 순으로 제시하고 있는데 구간이 바뀔 때는 \ 로 표시하고 Case가 바뀔 때는 //로 표시하였다. 구글 검색건수 1~1000의 Case1의 「1-」구간 58어, 「101-」구간 69어, Case2의 「101-」구간 3어, Case3「1-」구간 1어, Case3「101-」구간 8어 순이다.

다. 「-主義」의 경우는 「3.3.1」의 정착범주의 한자어에서도 상당수 있음을 언급한 바 있다. 조사자료의 편찬 시기에 서양으로부터 학술, 사상, 종교, 예술, 도덕 등의 문화를 많이 받아들이면서 「-主義」, 「-說」, 「-ism」과 같은 단어도 함께 수용하고 있었음을 알 수 있다.

한자어 「營養不良」은 「Case1」에 해당하며 구글 검색건수 444건에 그치고 있으나 「營」의 한자 표기변환으로 「榮養不良」로 상용화 되었다. 「榮養不良」은 사전과 코퍼스에 등재된 것은 물론 구글 검색건수도 3,900,000건에 이르고 있다. 미정착범주의 외래어 중에는 「ガール・スカウト, ボーイ・スカウト, モダンニズム」가 있는데 이들 대상어는 현대 일본어에서 「ガール・スカウト, ボーイ・スカウト, モダンニズム」로 사용되고 있으며 구글 검색건수는 각각 672,000건, 574,000건, 1,960,000건으로 정착범주에 속했다.

혼종어의 경우는 「3.3.1」의 혼종어와는 달리 외래어와 한자어의 조합이 대부분이었고, 단어를 이루고 있는 외래어는 「獨逸十戒, ア・ロ二州問題, 汎亜細亞主義」 등과 같은 나라 및 주(州), 대륙명과 「ウイルソン十四箇条, 石井ラシニング協約」 등과 같은 인명이 주를 이뤘다. 나라명 중에서도 특히 독일을 포함하는 「獨逸東進, 青年獨逸, 全獨逸主義」 등의 단어가 다수 있음을 볼 수 있었다. 이처럼 고유명사를 포함하는 혼종어의 경우는 미정착범주에서 나타나고 있었다.

3) 예외적 대상어

위의 1)과 2)에서는 사전과 코퍼스의 등재여부를 고려하여 구글 검색건수가 100,000건 이상인 것을 정착범주로, 구글 검색건수 1,000이하를 미정착범주로 구분지어 그 단어를 살펴보았다. 이를 통해 정착범주의 단어들은 정착어, 미정착범주의 단어들은 정착되지 못한 단어의 경향이 나타나고 있음을 알 수 있었다. 여기에서는 예외적인 대상어 즉, 사전과 코퍼스에 등재되어 있으나 구글 검색건수가 낮은 것과 사전과 코퍼스에 등재되어 있지 않으나 구글 검색건수가 높은 것이 어느 정도 있고 어떠한 것이 있는지 알아보고자 한다.

먼저, 사전과 코퍼스에 등재되어 있으면서 구글 검색건수가 낮은 대상어(구글 검색건수 10,000건 이하)는 다음과 같다.

軟文学(1,610), 慈善鍋(3,640), 雌雄淘汰(3,810), 超然内閣(4,120), 客觀描写(4,620), 社会劇(4,670), 異国情調(Exotic mood, 5610), プラグマチズム(Pragmatism, 6050), バーバリズム(Barbarism, 6070), 胆汁質(6260), 郷土芸術(6280), 予覚(6320), 人為淘汰(7830)

괄호안의 숫자는 구글 검색건수를 나타내며 1,000이하의 검색건수는 없었다. 이 구간의 단어들은 대체로 더 많이 사용되는 다른 단어들이 존재하였다. 예를 들어 『軟文学』은 남녀 간의 애정을 다룬 문학으로 지금은 『恋愛小説』라는 단어가 대신하고 있다. 『異国情調』는 이것과 같은 의미인 『異国情緒』의 구글 검색건수가 2,060,000 건에 달했다. 『プラグマチズム』는 『實用主義』로 『プラグマチズム』보다는 『實用主義』의 구글 검색건수가 훨씬 높았다. 『バーバリズム』는 『野蠻』 또는 그런 행위를 뜻하는 단어로 한자어가 사용되고 있음을 알 수 있다. 이 밖에 『雌雄淘汰, 超然内閣, 人為淘汰』 등은 다른 유사어가 있는 것은 아니나 단어의 사용 자체가 많지 않은 것으로 여겨진다.

다음은 사전과 코퍼스에 등재되어 있지 않지만 구글 검색건수가 높은 단어(구글 검색건수 100,000건 이상)들을 살펴보겠다.

液火(130,000), 民主的傾向(264,000), 安全第一(Safety First, 7,800,000)

『液火』는 고대의 병기(兵器)로 공기와 접촉하면 착화하여 타는 액체를 뜻한다. 『民主的傾向』와 『安全第一』는 의미를 찾아보지 않아도 알 수 있는 것으로 『安全第一』의 구글 검색건수는 7,800,000건에 달한다. 이 단어들은 사전과 코퍼스에 실리지 않았음에도 구글 검색건수가 높았으나, 이러한 예는 극소수에 불과했다.

4. 맺음말

본 연구에서는 20세기 초 신어를 대상으로 사전(『大辭林』제3판)과 코퍼스(BCCWJ)의 등재여부와 구글 검색건수의 분석을 통하여 현대 일본어에서의

정착여부를 파악하였다.

사전과 코퍼스의 등재여부를 분석한 결과, 조사대상어는 사전과 코퍼스에 각각 약 50% 전후로 실려 있었으며 한자어가 외래어와 혼종어에 비해 등재 비율이 높았다. 이들 조사대상어를 사전과 코퍼스 모두 등재되지 않은 대상어(Case1), 사전에만 등재된 대상어(Case2), 코퍼스에만 등재된 대상어(Case3), 사전과 코퍼스 모두 등재된 대상어(Case4)로 분류하여 분석한 결과, 「Case1」과 「Case4」의 비율이 각각 약 40%에 달했다. 어종별로 보면 한자어는 「Case4」에서의 비율이 가장 높고, 외래어와 혼종어는 「Case1」에서의 비율이 가장 높았다. 즉, 20세기 초 신어에 해당하는 조사대상어 중 사전과 코퍼스에 모두 등재되어 있는 단어는 약 40%이고, 다른 어종에 비해 한자어가 많다고 할 수 있다.

조사대상어의 구글 검색건수를 구간별로 나누어 분석한 결과는 Case별로 뚜렷한 차이를 나타내고 있었다. 「Case1」의 경우는 구글 검색건수가 낮고, 「Case4」의 경우는 구글 검색건수가 높으며 「Case2」와 「Case3」의 경우는 검색건수가 낮지도 높지도 않은 중간구간에 많이 분포하는 경향이 나타났다.

위의 결과를 토대로 본 연구에서는 사전과 코퍼스에 등재되어 있으면서 구글 검색건수가 100,001건 이상인 것을 정착범주로, 사전과 코퍼스의 등재여부와 관계없이 구글 검색건수가 1,000건 이하인 것을 미정착범주로 구분지어 조사대상어를 살펴보았다. 그 결과, 대부분 정착범주의 단어들은 친숙한 단어인 반면에 미정착범주의 단어들은 그 뜻을 찾아봐야 비로소 의미를 알 수 있었다. 어종별로 살펴보면 정착범주에 속하는 대상어는 한자어>외래어>혼종어의 순으로 많았다. 한자어는 다른 어종에 비해 그 비율이 압도적으로 높았으며, 구글 검색건수가 높을수록 좀 더 일반적으로 사용되는 한자어임을 알 수 있었다. 외래어의 경우 미정착범주의 단어는 생소할 뿐만 아니라 구글 검색에서 전혀 다른 의미로 검색되는 단어들이나 반면에 정착범주의 단어들은 다른 한자어로 대체가능한 외래어와 그렇지 않은 외래어로 분류할 수 있었다. 대체할 수 있는 단어가 없는 외래어는 정착범주 안에서도 구글 검색건수가 높게 나타났다. 혼종어의 경우는 미정착범주의 단어들은 한자어와 외래어의 조합이 많았던 반면, 정착범주의 단어들은 한자어와 고유어의 조합이 많았다.

본 연구는 20세기 초 신어의 현대 일본어에서의 정착여부를 파악하였다는 것, 이를 위해 사전과 코퍼스의 등재여부 및 구글 검색건수의 분석을 시도한

자체로 의의가 있다고 할 수 있다. 그 결과, 단어의 정착과 미정착을 구글 검색건수만으로 판단하기에는 어려운 면이 있으나 사전과 코퍼스의 각 Case 별 분석을 토대로 한 구글 검색건수는 단어의 정착을 파악하는 분석자료로 활용 가능하다고 할 수 있다. 그러나 본 연구는 조사자료의 일부만을 대상으로 하고 있으므로 20세기 초 신어를 대표하기에는 부족함이 있다. 향후 조사자료의 『통상의 번역어·외래어·신의어(新意語)』와 『정면에서는 의미를 알 수 없는 현대식 전용어(轉用語)』에 실려 있는 단어의 분석을 통하여 20세기 초 신어의 현대 일본어에서의 정착여부에 대한 연구를 좀 더 완고히 하고자 한다.

<參考文獻>

- 양민호(2013) 『외래어 정착 과정에 관한 한일 대조연구—구글 검색엔진을 통하여』 『日本語学』 第59輯 pp.199-218
- 장원재(2014) 『코퍼스를 활용한 일본어연구와 일본어 교육연구』 한국문화사 pp.32-34
- 稲垣吉彦(1982) 『現代の新語・流行語』 佐藤喜代治編 『講座日本語の語彙第7巻現代の語彙』 明治書院 pp.153-174
- 加茂正一(1944) 『新語の考察』 三省堂
- 鈴木英夫(1988) 『私の新語観』 『日本語学』7-6 明治書院 pp.4-7
- 松井栄一他(1996) 『新語辞典の研究と解題』 大空社
- 米川明彦(1989) 『新語と流行語』 南雲堂
- _____ (1992) 『新語と造語力』 『日本語学』11-5 明治書院 pp.50-57

<자료>

- 下中芳岳(1919) 『ポケット顧問 や、此は便利だ増補改版第二版』 平凡社 『大辞林』 제3판(<https://dic.yahoo.co.jp/>)
- 코퍼스(BCCWJ)(<https://chunagon.ninjal.ac.jp/>)
- <https://www.google.co.jp/>

<Abstract>

A study on settlement of new words accepted in the early 20th century in modern Japanese

—using dictionary, corpus and Google—

In this study, in order to analyze settlement of new words accepted in the early 20th century in modern Japanese, we investigated registration of the new words in dictionary (Daizirin 3rd edition) and Balanced Corpus of Contemporary Written Japanese(BCCWJ). And we compared it with number of Google Internet searches.

Registration ratio of new words in the dictionary and the BCCWJ is approximately 50%. For the analysis of new words, we divided it into 4 groups; new words unregistered in both the dictionary and the BCCWJ, new words registered only in the dictionary, new words registered only in the BCCWJ, new words registered in both the dictionary and BCCWJ. About 40% of the new words registered in both the dictionary and BCCWJ and Chinese words are dominant in this group.

The results of this study shows high correlation between the registration (in the dictionary and the BCCWJ) and the number of Google Internet searches. In other words, new words registered in both the dictionary and the BCCWJ tend to have higher number of Google Internet searches than other groups.