

[특집] 우리의 삶, 우리말에 담다

핵심어로 본 시대상의 변화

1946~2014년 《동아일보》 기사를 중심으로

최재웅 · 김일환 · 홍정하 · 이도길

고려대학교 언어학과 · 고려대학교 민족문화연구원

1. 도입

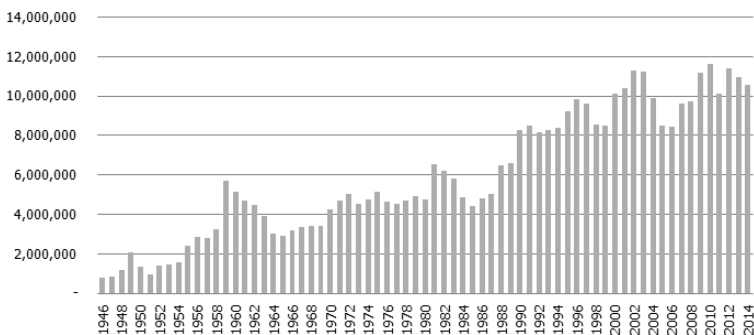
세상의 흐름을 읽고 이해하는 것은 예전에는 현자들만이 할 수 있는 일이었다. 정보화 사회에 들어선 지금은 상황이 많이 바뀌었다. 시대의 변화를 살펴보거나 요약할 수 있도록 도와주는 방법이 많이 개발되었고 또 무수히 개발되고 있다. 이는 대규모로 축적된 정보 자원과 비약적으로 발전하는 정보 처리 및 소통 도구들로 가능하게 되었다.

2015년은 광복 70주년을 기념하는 해다. 지난 70년을 되짚어 보는 방법이 여러 가지가 있겠지만, 신문을 통해 살펴보는 것도 한 가지 좋은 방법이다. 신문만큼 사회를 종합적으로 잘 반영하는 도구가 별로 없다. 이 글에서는 국내 대표 일간지 《동아일보》에 실린 기사를 기초로 핵심어를 통해 한국 사회가 변화해 온 모습을 살펴보기로 한다.

2. 연구 대상 자료와 전처리

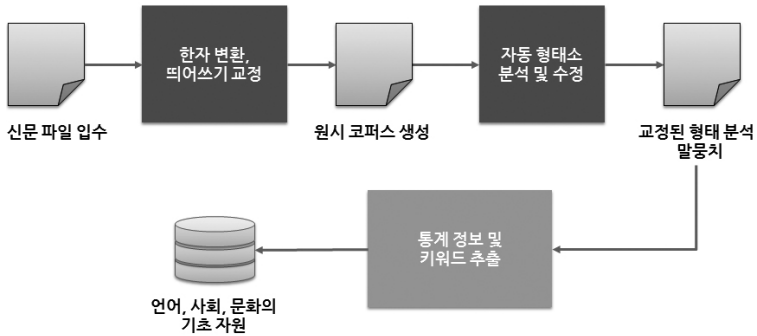
이 연구의 연구 대상 자료는 1946년부터 2014년까지 발간된 《동아일보》 기사로 전체 약 260만 기사(연간 평균 약 3만 7,000여 건), 전체 약 4억 1,000만(연간 평균 약 600만) 어절로 구성되어 있다. [그림 1]은 어절 단위의 연도별 《동아일보》 텍스트 규모이다.

[그림 1] 연도별 《동아일보》 텍스트 규모(단위: 어절)



신문 기사를 입수하는 단계부터 핵심어를 추출하는 과정은 [그림 2]와 같다. [그림 2]의 처리 과정에 대해 약간의 부연 설명이 필요하다. 이는 초기 현대 국어를 전산화하기 위해 선행되어야 하는 작업이라는 점에서 향후 유사한 연구를 수행할 때에도 참고가 될 수 있기 때문이다.

[그림 2] 연구 진행 과정 개요



2.1. 신문 기사 입수

《동아일보》 원본 기사는 동아일보사의 허락을 받고 사용했다.

《동아일보》 전체 기사에 대한 원문 이미지는 네이버의 뉴스 라이브러리에서도 확인할 수 있다(그림 3 참조).

[그림 3] 《동아일보》 원문 이미지 예시



2.2. 한자 변환 및 띄어쓰기 교정

우선 이 당시의 신문 기사는 한자가 그대로 노출되어 있고 띄어쓰기도 되어 있지 않은 상태이므로 한자 변환 도구와 띄어쓰기 교정 도구를 이용하여 전처리를 하는 과정이 필요하였다. 그 예시를 보이면 [그림 4]와 같다.

[그림 4] 원문 텍스트에서 한자를 한글로 변환한 파일 예시(띄어쓰기 포함)

```

<news>
<date>
19500101
</date>
<title>
犧牲精神固持 國防建設에 全力
</title>
<body>
犧牲精神固持
國防建設에 全力
國移總理談
오늘은 우리나라가獨立된後  두번째맞이하는 新年元旦임에따라  우리의偉大하고도險難한 建國聖業은第三年度로 進入하는첫날입니다. 時代
過去는모든 混亂을整頓하고 國家基礎를樹立하는데적극하였지만  오늘맞이하는新年은 國家의基礎를鞏固히하여  모든部面에있어 國力を積極
新年을맞이하는우리三千萬愛國同胞는  그思想과行動을오로지 建國에만 集注할覺悟를가져야할것이며建國에는 國家一切의建設을包含하
</body>
</news>
<news>
<date>
19500101
</date>
<title>
희생정신고지(犧牲精神固持) 국방건설(國防建設)에 전력(全力)
</title>
<body>
희생정신고지(犧牲精神固持)
국방건설(國防建設)에 전력(全力)
국무총리담(國移總理談)
오늘은 우리나라가 독립(獨立)된 후(後) 두번째맞이하는  신년원단(新年元旦)임에따라  우리의 위대(偉大)하고도  험난(險難)한  건국성업(建國)
과거(過去)는모든  혼란(混亂)을  정돈(整頓)하고  국가기초(國家基礎)를  수립(樹立)하는데  급(汲)적(汲)하였지만  오늘맞이하는  신년(新年)은  국가(國)
신년(新年)을맞이하는우리 삼천만에국동포(三千萬愛國同胞)는  그  사상(思想)과  행동(行動)을오로지  건국(建國)에만  집중(集注)할  각오(覺悟)
</body>
</news>

```

2.3. 기사 정제

당시의 사건, 사고와 비기사 텍스트는 핵심어 기반의 분석에서 고려 사항이 아니므로 이들을 제거하였다. 즉, ‘부음, 인사, 연재소설, 바둑’ 등의 기사는 모두 분석 대상에서 제외하였다.

2.4. 자동 형태소 분석 및 수정

KMAT를 이용하여 자동으로 형태소 분석을 수행하고,¹⁾ 그 결과를 전처리 지침에 따라 수정, 보완하였다. 한국어를 대상으로 개발된 형태 분석 도구는 현대 국어를 대상으로 할 뿐 아니라 비교적 현대 국어의 정서법에 맞게 표기된 자료를 전제로 한다. 1946년부터 1955년까지의 신문 기사는 물론 현대 국어라 할 만하지만 표기에서는 일관된 모습을 보이지 않는다. 연철 표기가 종종 나타나기도 하며 종성부용초성이 아닌, 대표음으로 표기한 경우도 심심찮게 발견되기 때문에 형태소 분석 도구를 사용할 경우 분석 오류가 상당수 포함된다. 따라서 신뢰할 만한 분석 결과를 확보하기 위해서는 적절한 수정, 보완 과정이 필수이다.

구체적인 수정, 보완 내용 중 일부를 제시하면 다음과 같다.

- 만히 만히/MAG → 많이/MAG
- 조타 조타/NA → 좡/VA+다/EM
- 잇어 잇/VV+어/EM → 있/VV+어/EM
- 가치 가치/NNG → 같이/MAG

이러한 과정을 거치고 나면 [그림 5]와 같은 형태 분석 코퍼스를 최종적으로 확보하게 되며, 이 코퍼스를 기반으로 하여 다양한 언어 사용 양상을 계량적으로 분석할 수 있게 된다.

1) KMAT(Korean Morphological Analyzer & Part-of-speech Tagger)는 고려대학교 민족문화연구원(이도길 교수)에서 개발한 자동 형태소 분석 및 품사 태거로서 확률 모델을 이용하여 최적의 형태소 분석 결과를 출력, 주석해 주는 도구이다(Lee & Rim 2009). 이 분석 도구는 현대 국어를 대상으로 할 때 분석 성공률이 96%를 상회하는 것으로 보고되었다(김일환 외 2013).

[그림 5] 한글로 변환된 파일에 형태소 정보를 부착한 결과

<title>	신년을>	신년/NNG+을/JKO
희생정신고지 국방건설에 전력	맞이하는>	맞이하/VV+는/ETM
</title>	우리>	우리/NP
<body>	삼천만>	삼천만/NR
<p>	애국동포는>	애국/NNG+동포/NNG+는/JX
</p>	그>	그/MM
<p>	사상과>	사상/NNG+과/JKB
</p>	행동을>	행동/NNG+을/JKO
희생정신고지	오로지>	오로지/MAG
</p>	건국에만>	건국/NNG+에/JKB+만/JX
<p>	집주할>	집주/NNG+하/XSV+르/ETM
</p>	각오를>	각오/NNG+를/JKO
국방건설에 전력	가져야>	가지/VV+어야/EM
</p>	할>	하/VX+르/ETM
<p>	것이며>	것/NNB+이/VCP+며/EM
</p>	이>	이/MM
국무총리담	건국에는>	건국/NNG+에/JKB+는/JX
</p>	국가>	국가/NNG
<p>	일체의>	일체/NNG+의/JKB
</p>	건설을>	건설/NNG+을/JKO
오늘은 우리나라가 독립된 후 두 번째	포함하고>	포함/NNG+하/XSV+고/EM
</p>	있으니>	있/VX+으니/EM
<p>		
과거는 모든 혼란을 경돈하고 국가 기초를 수립하는데 급하였지만 오늘 맞이하는 신년은 국가		
</p>		

3. 핵심어와 시대

3절에서는 1946년부터 2014년까지 1년 단위로 구축되어 있는 《동아일보》 텍스트에 기반해 시대를 구분하고, 각 시대의 특성을 살펴볼 수 있는 핵심어를 제시한다. 이를 위해 1946년부터 2014년까지 69년간의 《동아일보》 텍스트에서 총 빈도 1,000 이상의 1만 7,878개 일반 명사를 관찰 대상으로 하며, 각 어휘의 연도별 특징적 출현 정도를 나타내는 핵심도(keyness)를 계산하였다. 즉, 각 어휘에 대해 특정 연도에서 분포하는 경향이 다른 모든 연도의 분포 값에 비해 높은 것들을 핵심

어로 선정하였다. 측정 통계로는 카이스퀘어 잔차 값을 활용하였다.

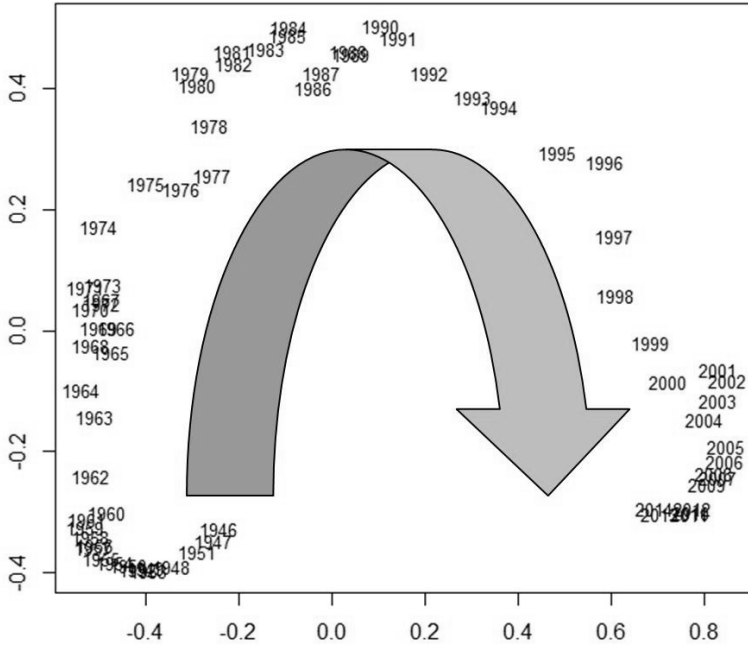
3.1. 시대 구분

먼저 총 빈도 1,000 이상의 1만 7,878개 일반 명사의 연도별 카이스퀘어 잔차 값, 즉 핵심도를 토대로 1946년부터 2014년까지 1년 단위로 구축되어 있는 《동아일보》 텍스트를 몇 개의 시대로 구분한다. 물론 편의상 10년 단위로 시대를 구분하거나, 기존에 알려진 광복 70년간의 시대를 구분하는 것도 방법이겠으나, 이 연구에서는 각 시대를 관통하는 유사성을 탐색하기 위해 1만 7,878개 일반 명사의 핵심도가 유사한 연도별 《동아일보》 텍스트를 분류하여 시대를 구분한다. 이러한 시대는 데이터에 기반을 두고 유형을 탐색하고 숨어 있는 지식을 발굴한다는 점에서 의의가 있다.

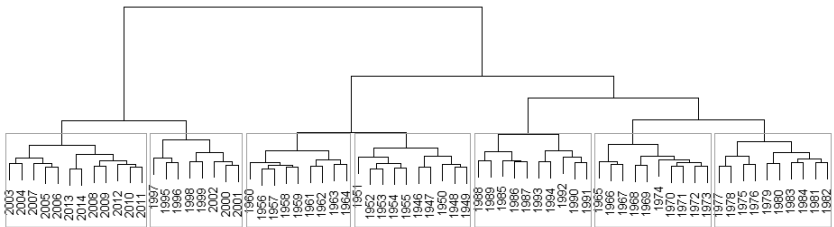
[그림 6]은 연도별 1만 7,878개 일반 명사의 핵심도에 기반을 두고 다차원 척도법을 통해 연도별 유사도를 분석한 것이다. [그림 6]에서 공간적으로 인접한 연도의 텍스트 사이에는 1만 7,878개 일반 명사의 핵심도가 유사함을 의미한다. 대체로 시간의 흐름에 따라 왼쪽 아래에서 시작하여 화살표에 따라 오른쪽 아래로 이동하고 있다. 이처럼 인접 연도의 텍스트 사이에는 각 텍스트를 구성하는 어휘의 핵심도가 유사함을 알 수 있다. 이러한 특성을 토대로 시대를 구분하기 위해 계층적 군집 분석을 통해 연도별 유사도에 대한 군집을 분석하면 [그림 7]과 같이 총 7개의 시대로 구분할 수 있다.

[그림 7]과 유사하게 인접 연도는 인접 군집으로 분류되어 있으며, 다음과 같이 각 시대는 10년 내외 연도로 구분되어 있다. 이 연구에서는 이러한 시대 구분에 따라 각 시대를 관통하는 핵심어가 무엇이며, 그 특성이 무엇인지 제시하고자 한다. 3.2부터는 각 시대에서 높은 핵심도를

[그림 6] 다차원척도법으로 본 연도별 유사성



[그림 7] 계층적 군집 분석을 통한 시대 구분



보이는 상위 2,000개 일반 명사 중 증가 추세를 보이는 각 100개를 토대로 각 시대의 특성을 논의하겠다.

- 1946~1955년
- 1956~1964년
- 1965~1974년
- 1975~1984년
- 1985~1994년
- 1995~2002년
- 2003~2014년

3.2. 광복 후 대한민국의 모습: 1946~1955년

[그림 8]은 1946년부터 1955년까지 핵심도가 높은 2,000개 어휘 중 증가 추세를 보이는 100개 어휘를 나타낸 워드 클라우드(word cloud)이다. 글자 크기가 클수록 증가 추세의 정도가 높은 핵심어이다. 이 핵심어들을 통해 관찰할 수 있는 시대적 특성을 살펴보면 다음과 같다.

- 소식통, 탐문: 신문은 “소식통에 의하면”과 같은 인용을 통한 보도가 증가. 취재라는 표현 대신 “탐문” 사용
- 협약, 조인, 가맹국, 서방측, 중립국: 외국과의 관계
- 재무부, 징수, 재무, 계정, 계획안, 계상, 공매, 인지, 상공부, 지불, 수요량, 회계, 불화: 국가 경제 초석 다지기. 세금 징수와 인지 판매, 국가 계정의 확충, 불화(미국 달러화)
- 재무부, 국방부, 농림부, 치안국, 교통부, 상공부, 문교: 주요 행정 기관

[그림 8] 1946~1955년의 증가 추세 핵심어



- 착복, 분실, 사치품, 대부, 도주, 기피자, 흑막: 당시 사회상
- 국방부, 육군, 군인, 삼군, 비행사, 군사비, 중령, 참모: 군 관련 핵심어
- 비료, 자재, 염전: 주요 산업 자재
- 포로, 동란, 대공: 한국전쟁
- 감사장, 표창장, 훈장, 수여: 공훈에 대한 표창
- 확약, 탐문, 향발(목적지를 향해 출발), 당년, 상기, 환(돌아오다), 추력(트럭), 작주(昨週), 노정(거쳐 지나가는 길 또는 과정), 피체(체포): 시대적 특성 어휘

3.3. 어려운 경제 상황과 새로운 밑거름: 1956~1964년

[그림 9]는 1956년부터 1964년까지 핵심도가 높은 2,000개 어휘 중에서 서 증가 추세를 보이는 100개 핵심어이다. 이 핵심어들에 반영되어 있는 시대적 특성은 다음과 같다.

- 미터, 톤, 밀리: 미터법의 사용
- 시합, 미들, 급, 대전, 데뷔(테뷔), 기권, 헤비, 권투, 득승(승리): 권투의 인기
- 국민교, 중학, 국교, 기성회비, 국민학교, 학부형: 교육에 대한 관심
- 합승, 내핍, 유괴, 식모, 빈곤, 판잣집: 합승 택시의 등장과 궁핍한 시대상
- 전동화, 어선, 어로, 어업, 개간, 양회, 보세, 외화, 외자: 경제 및 산업 현실

[그림 9] 1956~1964년 증가 추세 핵심어



- 쿠데타, 이양, 방첩, 군정, 정세: 정치 현실
- 라디오, 방송국, 방송국: 라디오 방송의 등장 및 인기

3.4. 경제 발전과 그림자: 1965~1974년

[그림 10]은 1965년부터 1974년까지 증가 추세의 핵심어 100개이며, 이를 통해 살펴볼 수 있는 시대적 특성은 다음과 같다.

- 공산품, 일반미, 석탄, 섬유, 중화학, 원유, 석유, 기능공, 도매, 택시, 프로판, 중업원, 대륙붕, 오일, 노임, 섬유류, 정부미, 합판, 설탕: 경제적 주요 물품 및 상황
- 자급, 개발도상국, 국력, 강대국, 수출, 개량, 증산, 조림: 국가적 경제 정책
- 무허가, 불량, 단속, 공해, 다방, 여관, 분노, 물가, 품귀: 시대상

[그림 10] 1965~1974년 증가 추세 핵심어



3.5. 문화의 시대: 1975~1984년

[그림 11]은 1975년부터 1984년까지 증가 추세의 핵심어이며, 이를 통해 살펴볼 수 있는 시대적 특성은 다음과 같다.

- 간염, 후원회, 지하상가, 운전기사, 자울화, 포니, 장애인, 여성계: 사회상의 변화
- 소강당, 창단, 디스코, 출판, 서클, 카세트, 연주회, 코피(커피), 올림픽, 문예, 연극제, 레코드, 해금, 도큐멘터리(다큐멘터리), 교회, 서양화가, 미술, 카톨릭(가톨릭), 교황, 꽃꽂이, 다과, 검도, 봉황기, 서예: 문화 예술 및 종교의 핵심어 등장 및 풍성
- 대북한, 훈련, 배치, 핵미사일, 동구권, 중거리, 방위비: 정세
- 국민학생, 합격선, 학장, 대입, 학력고사, 교육세, 눈치작전, 졸업, 탈락: 교육 문제

[그림 11] 1975~1984년 증가 추세 핵심어



- 이기주의, 도덕성: 사회적 풍토
- 실명제: 정책

3.7. IT, BT 등장과 새로운 산업: 1995~2002년

1995년부터 2002년까지의 [그림 13]을 통해 알 수 있는 시대적 특성은 다음과 같다.

- 휴대, 캠코더, 배아, 디지털, 사이트, 홈페이지, 메일, 클릭, 램, 복제, 게놈, 메모리: 기술에 대한 관심
- 팀장, 마케팅, 경영자, 고객, 회계사, 카드사, 연구원, 카드, 스톡, 애널리스트, 신용, 수익성, 쇼핑몰, 영업, 항공권, 회계, 네티즌, 텔레콤, 애니메이션, 순이익, 쇼핑, 컨설팅: 경제 환경의 변화

[그림 13] 1995~2002년 증가 추세 핵심어



- 근육, 몸, 정신과, 처방전, 내과: 육체적 · 정신적 건강
- 미드필더, 패스: 축구에 대한 관심

3.8. 혁신과 인간: 2003~2014년

마지막으로 2003년부터 2014년까지의 [그림 14]를 통해 알 수 있는 시대적 특성은 다음과 같다.

- 출시, 입점: 새로운 소식
- 특성화, 혁신적, 특화, 이노베이션, 특허, 신개념, 창의: 혁신과 창의성의 핵심어
- 지인, 협업, 소통, 융합, 다양, 공유, 파트너, 토크, 공감, 접목: 소통의 핵심어

[그림 14] 2003~2014년 증가 추세 핵심어



- 모바일, 탑재, 성능, 브랜드, 기술, 콘셉트, 역량, 기술력, 데이터, 외관, 모양새: 기술의 경향
- 스펙, 인재, 재능: 인재의 핵심어
- 진정성, 존재감, 편안, 안정성, 지속적, 소외: 심리적 핵심어
- 아웃도어, 식재료, 셰프, 마트: 새로운 산업

4. 단어 사용 추이와 그 함의

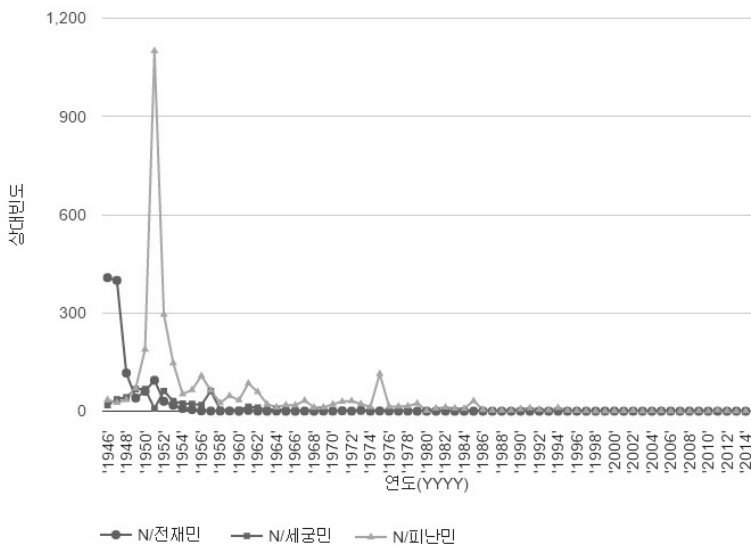
시기별로 단어 사용 추이를 관찰하고 의미 있는 패턴을 포착하는 과정은 단순하지 않다. 이 연구에서는 3절에서 제시된 핵심어와 빈도별 차이에서 의미 있는 상관성을 보이는 단어들을 주제별로 분류하여 이들의 단어 사용 추이를 살펴보고 이러한 추이가 의미하는 바가 무엇인지 분석해 보고자 한다.

4.1. 언어

4.1.1. '전재민'과 '세궁민'을 아십니까?

'전재민(戰災民)'은 태평양전쟁에 끌려갔다가 다시 조선으로 귀국한 사람들을 지칭하는 것으로, 이들이 사회적으로 다양한 문제(실업 문제, 주거 안정 등)와 결부되어 있음을 보여 준다. '세궁민(細窮民)'은 극도로 가난한 사람들을 일컫는 말이다. 이 단어들은 1950년 이후 그 쓰임이 거의 제약되어 사전에서나 찾아볼 수 있는 단어가 되었지만 해당 시기의 사회적 모습을 잘 보여 준다는 점에서 흥미로운 단어들이다.

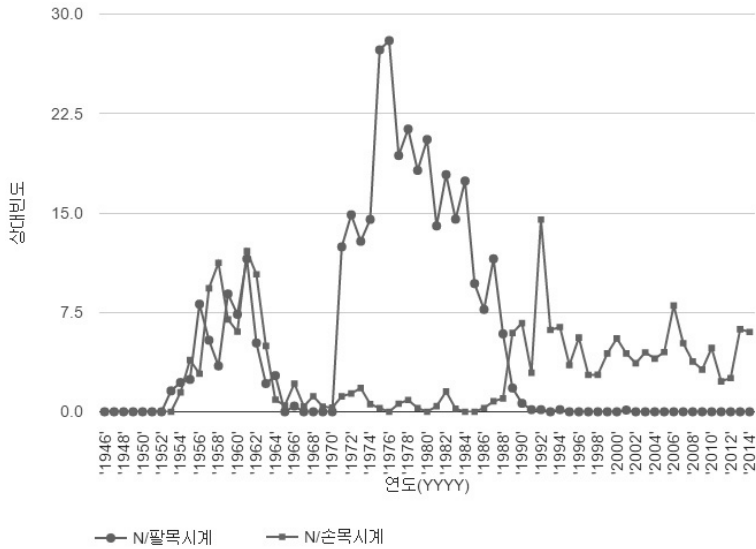
[그림 15] '전재민, 세궁민, 피난민'의 연도별 사용 빈도



4.1.2. '팔목시계'와 '손목시계': 팔목에 차는 시계냐, 손목에 차는 시계냐

'팔목시계'는 '손목시계'로 순화되었다(1989년). 흥미로운 것은 이전까지만 하더라도 '팔목시계'의 사용 빈도가 압도적으로 높았다는 점이다. 특히 초창기의 경쟁 단계에서 승리한 것으로 보이는 '팔목시계'가 국어 순화의 결과로 순식간에 그 빈도가 역전되는 것은 놀라운 결과로 보인다. 언중의 언어 사용조차 국어 순화 정책의 강력한 영향력을 넘어 서지 못하는 것일까?

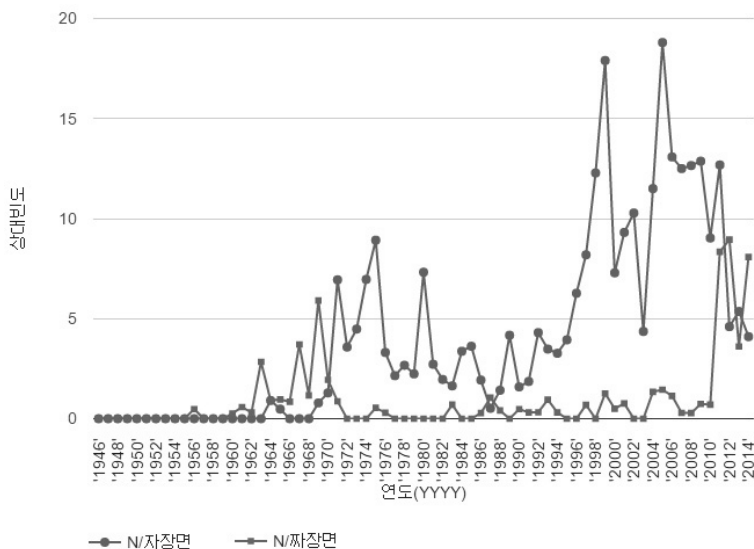
[그림 16] '팔목시계, 손목시계'의 연도별 사용 빈도



4.1.3. 짜장면→자장면→다시 ‘짜장면’으로?

이와는 반대로 ‘짜장면’은 2011년에 ‘자장면’과 함께 복수 표준어로 인정되었다. 이는 ‘짜장면’에 대한 언중의 높은 사용 빈도를 반영한 결과라고 하지만, 비록 신문 텍스트라는 한정된 자료일망정 ‘자장면’은 의외로 높은 빈도로 사용되었다. 오히려 1960년대에는 ‘짜장면’의 빈도가 더 높은 것으로 보아 ‘짜장면’의 복수 표준어 획득은 이전 시기의 언어 사용 양상을 반영한 것인지도 모르겠다. 이러한 추세라면 ‘빠스(bus)’도 복수 표준어가 되지 말라는 법은 없을 것이다.

[그림 17] ‘자장면’과 ‘짜장면’의 연도별 사용 빈도

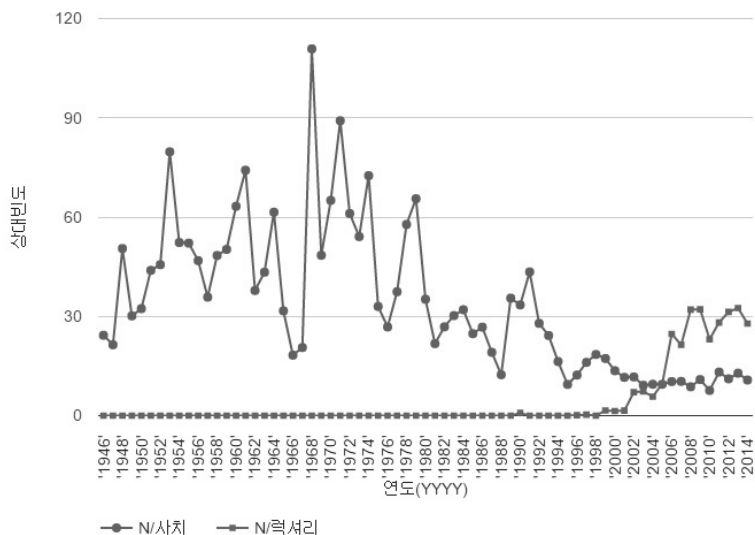


4.2. 문화

4.2.1. '사치'는 더 이상 허물이 아니다: '사치'에서 '럭셔리'로

'사치'는 부정적인 맥락에서 주로 쓰인다. '사치스럽다'도 부정적인 양상과 관련되어 있다. 그러나 외모를 중시하고 자신을 가꾸는 것에 대한 사람들의 인식이 바뀌면서 2000년대 중반부터 '사치' 대신 '럭셔리'라는 영어의 사용 빈도가 급격히 높아진다. 이는 '사치'가 가진 부정적 뉘앙스를 대체하기 위한 수단으로 영어를 선택한 결과로 보인다. 이러한 양상은 '주방장'과 '셰프' 등에서도 발견된다.

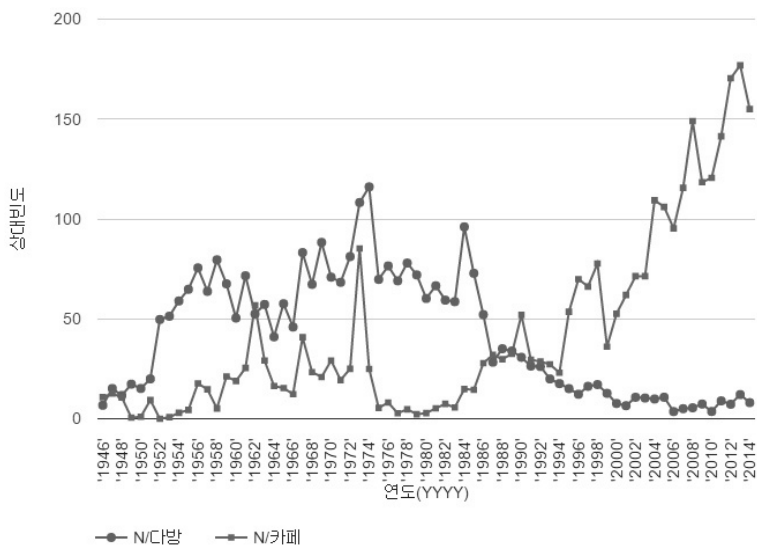
[그림 18] '사치'와 '럭셔리'의 연도별 사용 빈도



4.2.2. 커피는 ‘다방’에서도 마실 수 있지만

요즘 도시에서 ‘다방’은 찾아보기 어렵다. 다방의 대부분은 도산하거나 명칭을 바꾸기도 하였고, 새로 생기는 가게는 ‘다방’이 아니라 ‘카페’다. ‘카페’라는 외래어는 1946년부터 등장하였다. 그러나 이후 ‘다방’보다 빈도가 크게 낮았다가 1987년 무렵부터 크게 격차를 벌리기 시작하였다. 왜 ‘카페’는 뜨고 ‘다방’은 졌을까?

[그림 19] ‘다방’과 ‘카페’의 연도별 사용 빈도

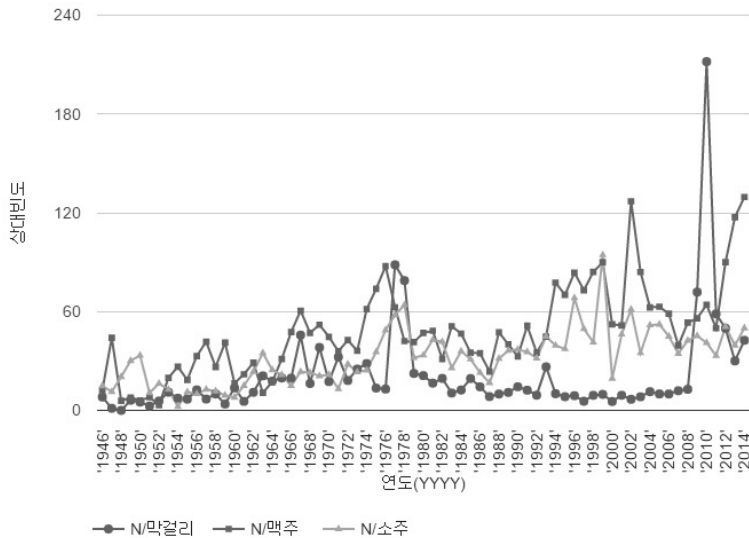


4.2.3. 우리의 술은 '막걸리'?

막걸리가 우리의 전통술임은 부인하기 어렵다. 그러나 신문의 언어 사용만 놓고 볼 때에는 양상이 다르다.

1978년 무렵과 2010년 무렵을 제외하고는 '맥주'의 빈도가 '소주', '막걸리'를 압도한다. '소주'의 사용 빈도도 꾸준한 편이지만 빈도상으로는 '맥주'를 능가하지 못한다. '맥주'는 적어도 우리에게 가장 많이 언급된 술이다.

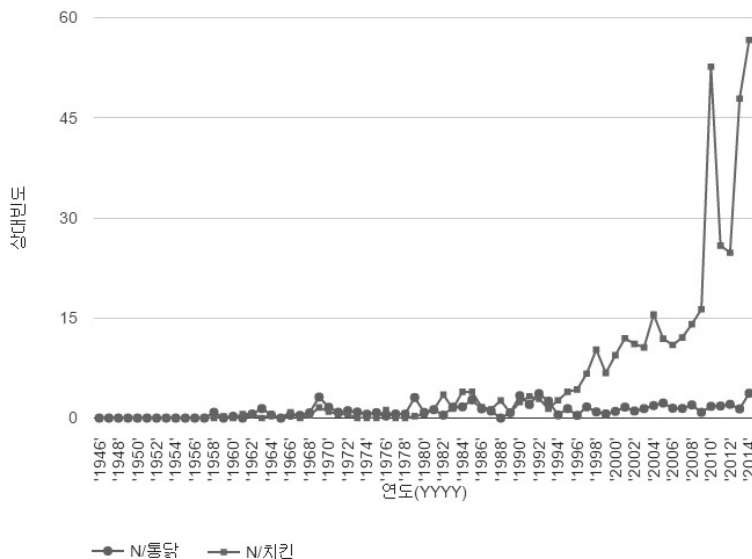
[그림 20] '막걸리, 맥주, 소주'의 연도별 사용 빈도



4.2.4. 응답하라 1994: '치킨'의 역전

'치킨'에 대한 대한민국의 관심은 놀라울 정도다. 2015년을 기준으로 치킨 브랜드만 해도 수십 가지가 넘고 동네마다 치킨 점포가 길에 넘친다. 최근에는 치킨과 맥주를 합성한 '치맥'이 큰 인기를 끌기도 하였다. 그러나 '치킨'이란 말이 신문에서 쓰이기 시작한 것은 1980년대 들어서부터다. 그 이전에는 근소한 차이긴 하지만 '통닭'의 사용 빈도가 높았고 출현한 시기도 훨씬 앞선다. '통닭'은 1958년 기사에 '통닭구이'와 같이 쓰이다가 이후 '통닭'이 하나의 음식으로 자리를 잡은 것으로 보인다. '치킨'이 '통닭'의 빈도를 압도하기 시작한 것은 1994년 무렵이다. 이는 1990년대에 '프라이드치킨', '양념 치킨' 등과 같은 메뉴가 도입되면서 나타나기 시작한 것으로 보인다.

[그림 21] '통닭'과 '치킨'의 연도별 사용 빈도

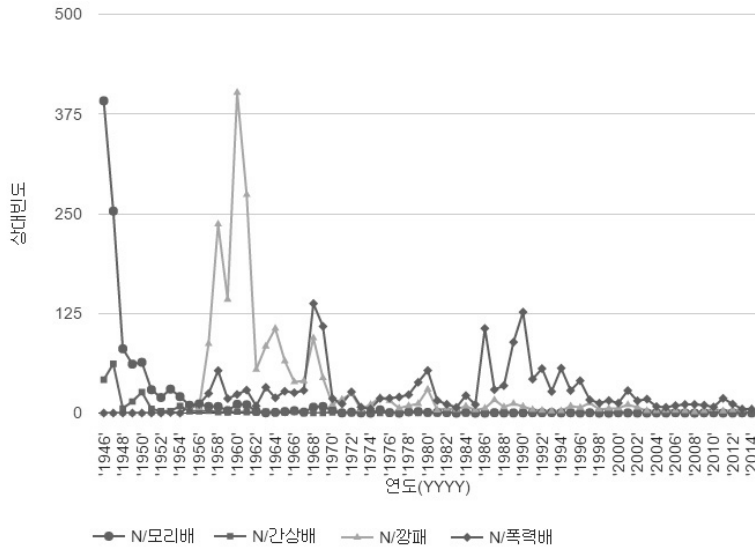


4.3. 사회

4.3.1. 나쁜 놈들 전성시대: '모리배, 간상배, 깡패, 폭력배'

시대마다 그 시대를 대표하는 '나쁜 놈'이 있다. 해방 이후에는 약 10년간은 '모리배(謀利輩)'가 전성기를 구가했다. 해방 이후 일본인 재산(적산)을 처분하는 과정, 남북의 분단과 정부 수립 등의 혼란기를 틈타 자신의 잇속을 챙기는 나쁜 인간들의 무리를 지칭하는 단어가 '모리배'였다. 이 과정에서 장사를 통해 폭리를 취한 '간상배(奸商輩)'도 빼놓을 수 없다. 1960년대부터는 '정치 깡패'들이 극성을 부렸다. 각종 부정부패, 부정 축재 등으로 얼룩진 이 시기는 '깡패'를 동원해서라도 정권을 유지해야 했던 것이다. 그리고 보면 나쁜 놈의 전성기는 사회적 혼란을 자양분으로 삼아 성립되는 것으로 보인다. 이후에는 '폭력배'가 오랜 시

[그림 22] '모리배, 간상배, 깡패, 폭력배'의 연도별 사용 빈도



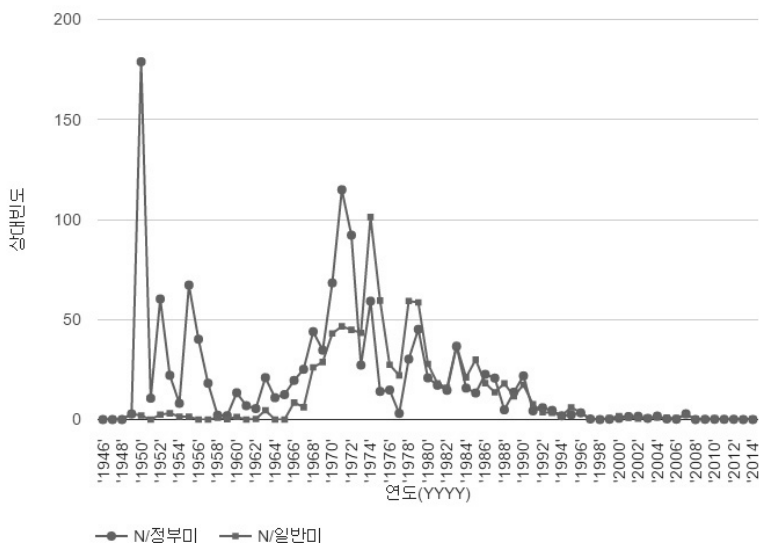
간 동안 생명을 유지하고 있다. 1990년대 초반에는 특히 조직 폭력배 소탕 작전과 관련하여 이들의 빈도가 높게 나타난 것으로 해석된다.

4.3.2. '정부미' 조차 귀했던 시절

'쌀'과 관련한 다양한 이름은 '쌀'에 대한 우리 사회의 높은 관심을 잘 보여 준다. '정부미, 방출미, 일반미, 배급미, 군량미, 비축미, 구호미, 보유미, 백미, 양질미, 공양미, 안남미' 등이 그것이다.

해방 이후와 한국전쟁 전후에는 '정부미'의 빈도가 급증한다. 정부가 나서지 않으면 안 될 정도로 국민의 식량 사정이 어려웠던 시기라는 점은 쉽게 추측된다. 1970년대와 1980년대 초반까지는 정부미와 일반미가 함께 높은 빈도로 사용되었다. 지금이야 정부미와 일반미라는 단어

[그림 23] '정부미'와 '일반미'의 연도별 사용 빈도

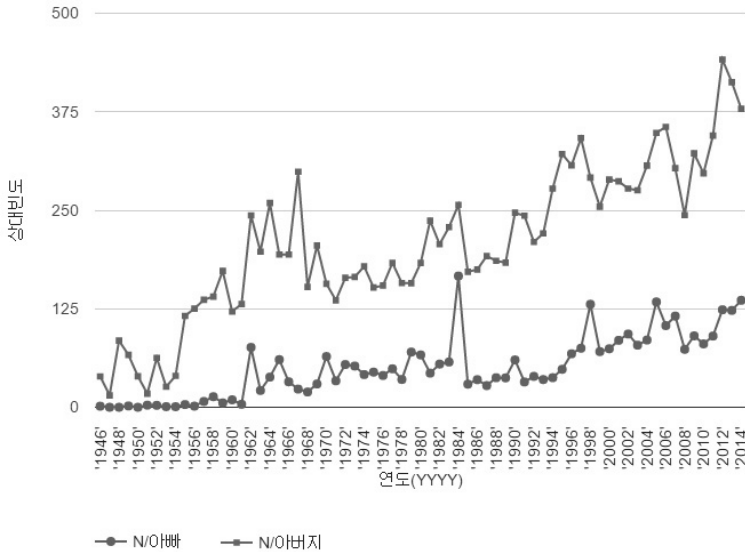


는 거의 사어화될 정도로 쓰이지 않지만 당시에는 쌀이 가장 중요한 식량 자원이었다. 사정이 좀 좋은 사람들은 ‘일반미’를 먹었을 것이다.

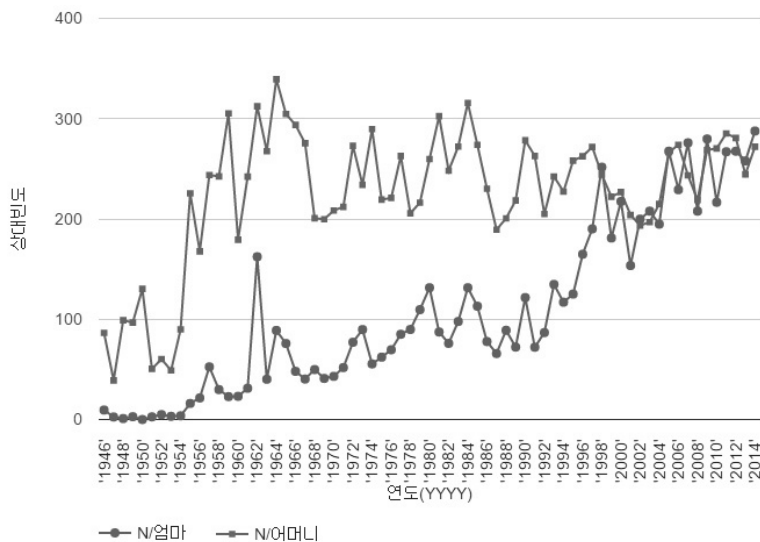
4.3.3. 가까워진 ‘엄마’와 ‘어머니’, 여전히 먼 ‘아빠’와 ‘아버지’

70년간의 가족 호칭에 대한 빈도 변화 중 눈에 띄는 것이 있다. ‘엄마’와 ‘아빠’는 다소 구어적이고 비격식적이어서 ‘어머니’, ‘아버지’에 비해 신문에서 그리 자주 사용되는 단어는 아니었다. 그러나 이들 단어의 빈도 변화는 묘한 대조를 보인다. 1990년대 후반에 이르러서는 ‘엄마’는 ‘어머니’와 대등한 빈도로 쓰이는 반면, ‘아빠’는 여전히 ‘아버지’와 간격을 좁히지 못한다. ‘아버지’라는 존재가 가지는 막중한 가치는 쉽사리 ‘아빠’에게 그 영역을 허락해 주지 않는 것으로 보인다.

[그림 24] ‘아빠’와 ‘아버지’의 연도별 사용 빈도



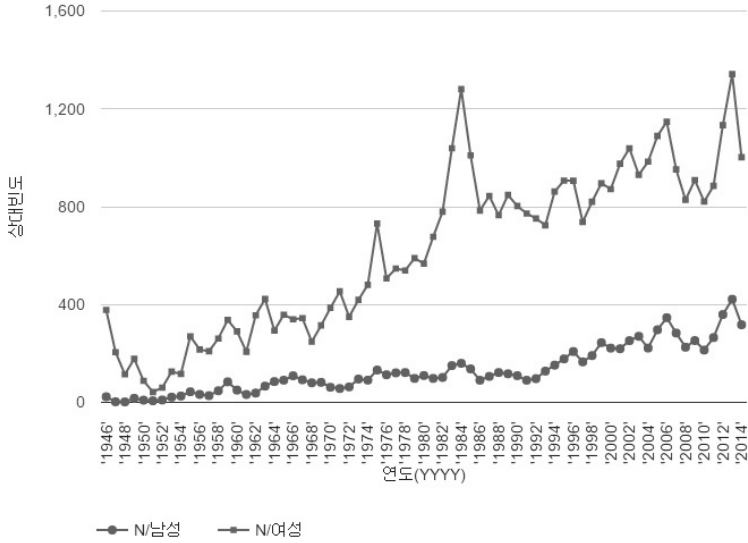
[그림 25] '엄마'와 '어머니'의 연도별 사용 빈도.



4.3.4. '여성'은 '여자'보다 위대하다

여성의 사회적 지위가 높아지면서 '여성'에 대한 빈도는 꾸준히, 때로는 극적으로 증가하는 추세를 보였다. 그러나 유사한 단어라 하더라도 '여자'는 그렇지 않다. 즉, '여자'는 '남자'와 사용 패턴이 대체로 일치하는 것으로 나타나고 있는 반면 '여성'은 남성에 비해 훨씬 높은 빈도로 사용되어 왔으며 그 추세도 꾸준히 증가하고 있다. '여자가 아닌 '여성' 상위 시대라는 말이 있는 것처럼 말이다.

[그림 26] '남성'과 '여성'의 연도별 사용 빈도



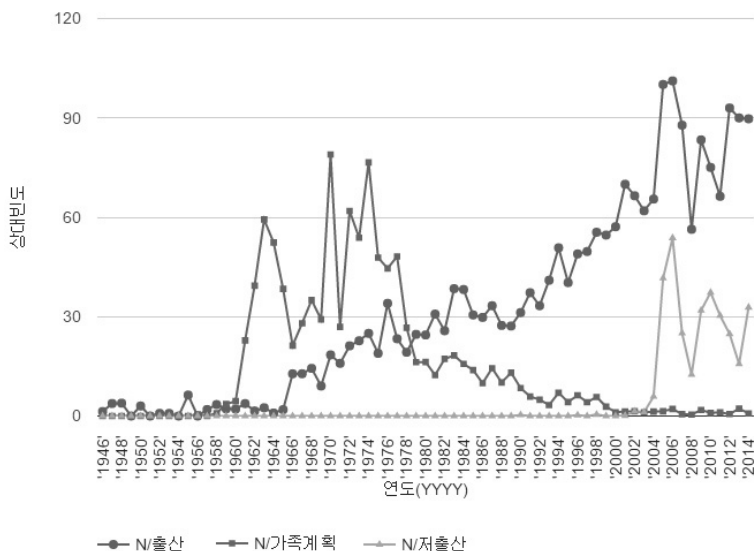
[그림 27] '남자'와 '여자'의 연도별 사용 빈도



4.3.5. 너무 길었던 ‘가족계획’

2015년, 대한민국은 저출산과 고령화로 인구 감소의 위기에 직면해 있다. 정부는 출산율 제고를 위해 다양한 정책을 쏟아내고 있으며 매스컴에서는 인구 감소가 예상되는 시점에서 심각한 위기 상황에 대해 다양한 분석 결과를 보도하느라 분주하다. 그러나 불과 20년 전만 하더라도 ‘가족계획’에 의한 인구 억제 정책이 논의되었다는 점을 생각하면 이러한 결과는 매우 놀라운 것이다. 인구 억제를 위한 ‘가족계획’은 1960년부터 논의되기 시작하여 1990년대 후반까지 사용 빈도가 관찰된다. 이는 ‘출산’에 대한 빈도가 크게 증가하기 시작하였던 시점과 비교하여 살펴볼 필요가 있다. 출산이 높은 빈도를 보인 것은 출산율이 높은 것보다 출산율이 낮기 때문일 가능성이 높다는 점을 고려해야 한다. 그렇

[그림 28] ‘출산, 가족계획, 저출산’의 연도별 사용 빈도

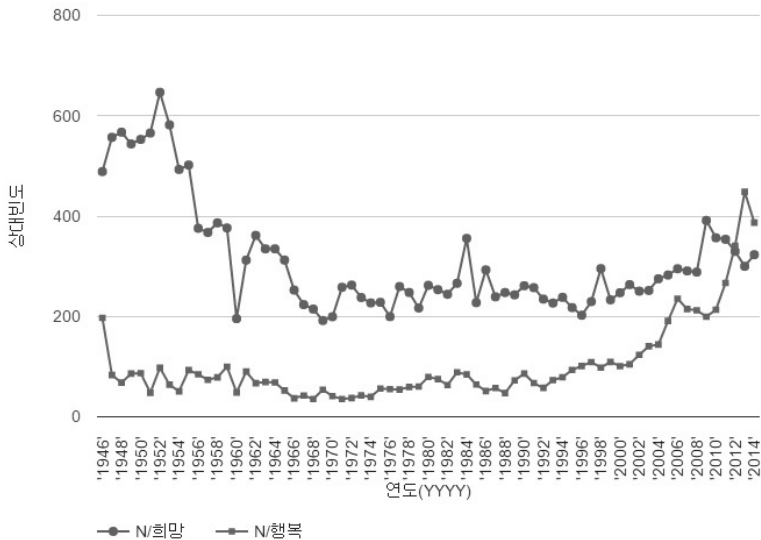


다면 가족계획은 너무 오래 살아남았다. 그리고 정부의 출산 장려 정책을 통한 고령화 사회 대비 정책은 이미 시기를 놓친 감이 없지 않다.

4.3.6. '희망'과 '행복'의 쌍곡선: 우리는 왜 '행복'을 얘기하는가

사람은 어려운 처지일수록 희망을 노래한다. 희망마저 없다면 우리는 힘든 세상 속에서 버텨 나가기 쉽지 않을 것이다. 즉, 언젠가는 이 어려운 상황을 극복하고 밝은 내일을 맞이할 수 있다는 희망은 인간의 생존에 필수 요소다. 희망에 대한 이러한 상황은 '희망'이 높은 빈도를 보인 시점과 비교된다. 즉, 해방 이후 새 정부에 대한 희망도 있었겠지만 한국전쟁 이후의 폐허 속에서도 우리는 '희망'을 이야기하였다. '희망'의 빈도 추이는 이 시점을 제외하면 대체로 고르게 분포한다. 반면 2000년대 후반에 오면서 '행복'의 빈도가 크게 증가 추세를 보이다가 급기야 2012년에는 희망의

[그림 29] '희망'과 '행복'의 연도별 사용 빈도

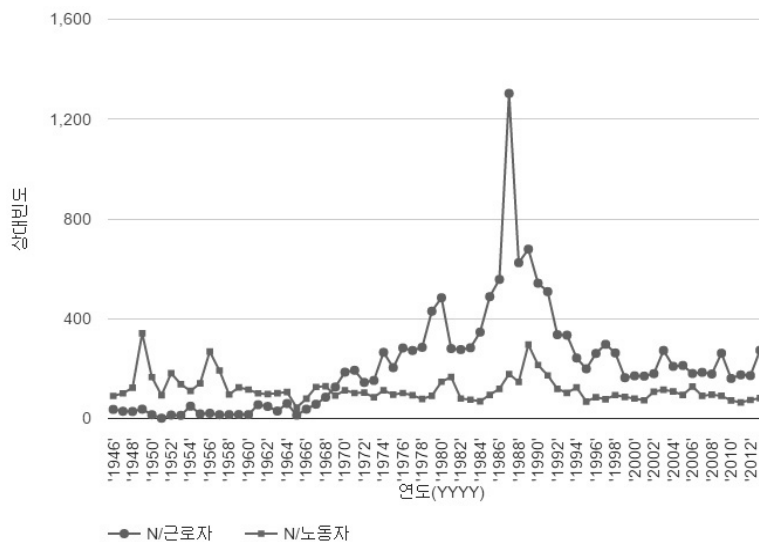


빈도를 추월하기에 이른다. 왜 우리는 ‘희망’보다 ‘행복’을 더 자주 이야기 하는 것일까? 경제적으로 부유해진 상황에서 이제 ‘행복’을 노래할 만한 시점에 이른 것인가? 여전히 ‘행복’은 추구되어야 할, 즉 아직 이르지 못한 것인가? 2012년, 왜 우리는 ‘희망’보다 ‘행복’을 더 우선시켰을까? 건강한 사람이 건강에 관심이 없고, 공기가 풍족한 세상에서 공기에 대해 논의하지 않듯이, 행복한 사람은 ‘행복’에 대해 관심을 가질 필요가 없을 것이다. 우리는 여전히 행복과는 거리가 있는 세상에 살고 있다.

4.3.7. 1969년: ‘노동자’에서 ‘근로자’로

노동자와 근로자는 사전적으로는 유사한 의미를 가지지만 그 뉘앙스에서 느껴지는 차이는 생각보다 큰 것으로 보인다. ‘노동자’는 1969년 이전까지는 ‘근로자’보다 높은 빈도로 사용되다가 이 해를 기점으로 ‘근로

[그림 30] ‘근로자’와 ‘노동자’의 연도별 사용 빈도



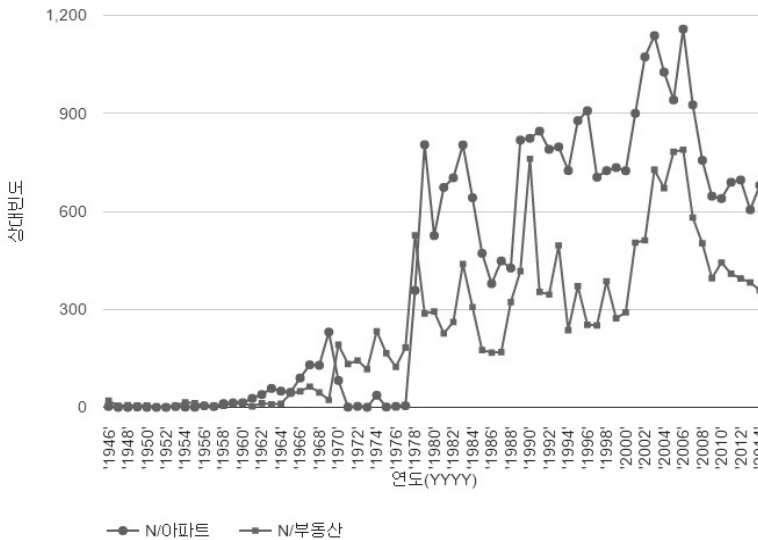
자'의 빈도가 '노동자'의 빈도를 넘어서 이후에는 줄곧 더 자주 언급되는 단어가 되었다. 특히 1987년도에는 '근로자'의 빈도가 폭발적으로 증가하였다. '근로자'의 빈도가 1960년대에 이르러 '노동자'보다 높아지게 된 데에는 1963년에 제정된 '근로자의 날'과도 관련이 있는 것으로 보인다. 정부나 기업에서는 노동 운동과의 연관성을 의식해서인지 '노동자'보다는 '근로자'를 선호하였고, 그 결과의 하나로 '노동자의 날'이 아닌 '근로자의 날'이 탄생한 것으로 보인다. 근로자는 그 이후 지속적인 호응을 받았고 1987년 6월 민주화 항쟁과 노조 운동과 관련해 관심을 받게 되었다.

4.4. 경제

4.4.1. 1979년: '아파트', 다시 '부동산'을 넘다

2015년 아파트는 한국인이 가장 많이 거주하는 주택 유형의 하나로

[그림 31] '아파트'와 '부동산'의 연도별 사용 빈도

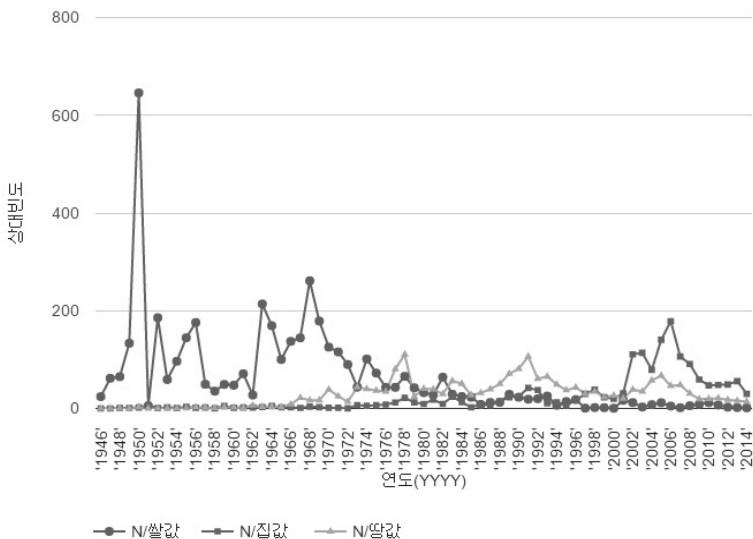


자리 잡았다. ‘아파트’는 한때 ‘어파아트, 아파아트’ 등으로 쓰이다가 1970년대 중후반부터 ‘아파트’로 일관되게 나타나기 시작한 것으로 보인다. ‘아파트’는 1960년대 초반 등장해 그 빈도가 증가하다가 1970년대에는 크게 감소하고 대신 ‘부동산’이 높은 빈도로 사용되었다. 그러다가 1980년대부터는 지속적으로 ‘아파트’가 ‘부동산’의 빈도를 넘어선다. 이는 부동산 투자가 주로 아파트를 중심으로 이루어졌다는 점, 그리고 아파트는 투자 대상이자 거주 공간으로서 그 중요성이 더욱 높다는 점과 관련된 것으로 보인다.

4.4.2. ‘쌀값’→‘땅값’→‘집값’

대한민국이 지난 70년간 눈부신 경제 성장을 이루었다는 점은 자명하다. 경제 성장과 함께 우리의 생활 경제도 관심 분야가 변화해 왔다.

[그림 32] ‘쌀값, 집값, 땅값’의 연도별 사용 빈도

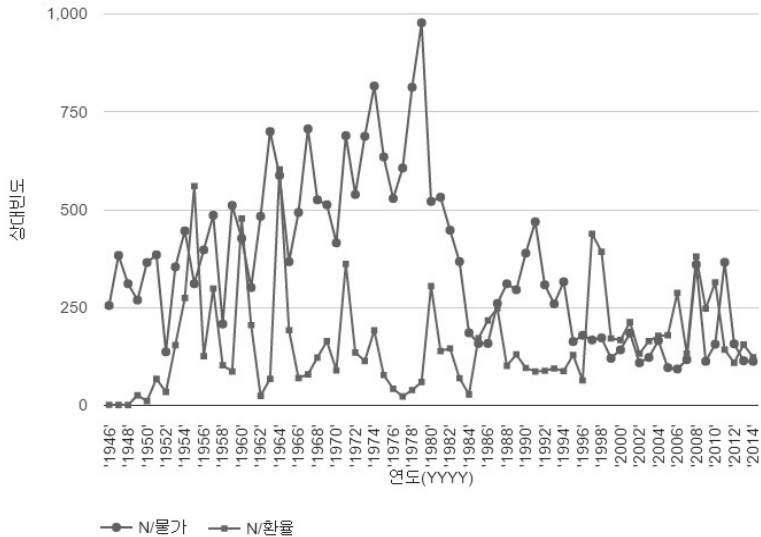


1970년대 중반까지는 ‘쌀값’의 빈도가 가장 높게 나타난다. 서민 생활의 생존과 직결되는 ‘쌀’의 물가에 대한 관심이 가정 경제에서 가장 큰 비중을 차지한 결과가 아닌가 추측된다. 즉, 주거 환경이나 투자 등에 대한 관심보다는 우선 먹고 사는 일이 가장 시급한 문제였을 것이다. 이후에는 ‘땅값’이 2000년대 초반까지 높은 빈도를 보였다. 땅값은 주로 개발이나 투자와 관련된 것이다. 땅값이 오르면 덩달아 집값이 치솟게 마련이다. 2000년대 들어서는 ‘집값’의 빈도가 가장 높게 유지되고 있는데, 이는 주거 환경과 크게 관련된다는 점에서 주목할 만한 변화로 보인다.

4.4.3. ‘물가’와 ‘환율’의 상관성

최근에는 ‘물가’와 ‘환율’의 관련성이 쉽게 납득되지만 이전에는 그렇

[그림 33] ‘물가’와 ‘환율’의 연도별 사용 빈도



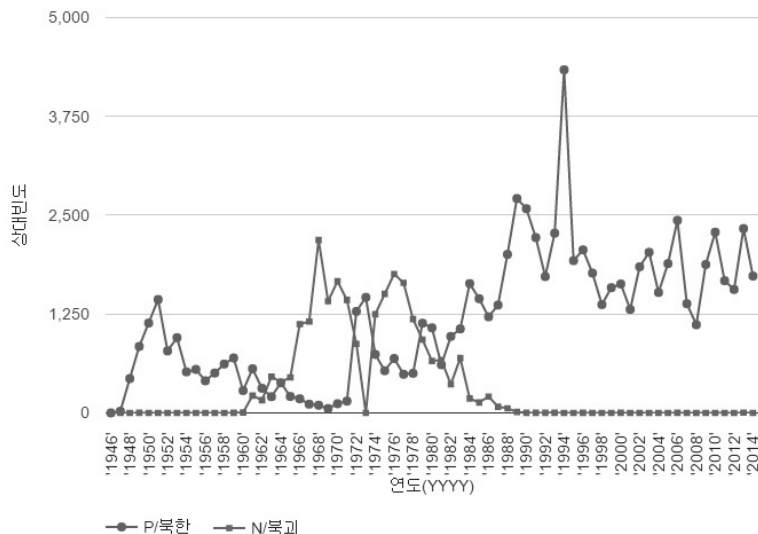
지 않았던 것으로 보인다. 즉, '물가'와 '환율'의 빈도에서 보이는 상관성은 2000년대 이전에는 특정한 몇 해를 제외하고는 그리 높지 않았다. 그러다가 2000년대로 넘어오면서 '환율'과 '물가'가 갖는 빈도상의 관련성이 점차 높은 것으로 포착되었다. 이는 국내의 물가가 환율과 더욱 밀접한 관련성을 갖게 되었음을 간접적으로 보여 준다.

4.5. 정치

4.5.1. 냉전의 종식: '북괴'와 '북한'

냉전의 종식과 함께 '북괴'라는 다소 적대적인 용어의 사용도 크게 감소하였다. 특이한 것은 한국전쟁 후 1960년대 초반까지만 해도 '북괴'보다 '북한'의 빈도가 압도적일 뿐 아니라 실제 '북괴'가 등장하기 시작한 것은 1960년 이후라는 점이다. 즉, '북괴'의 사용 연한은 1960년대부

[그림 34] '북괴'와 '북한'의 연도별 사용 빈도

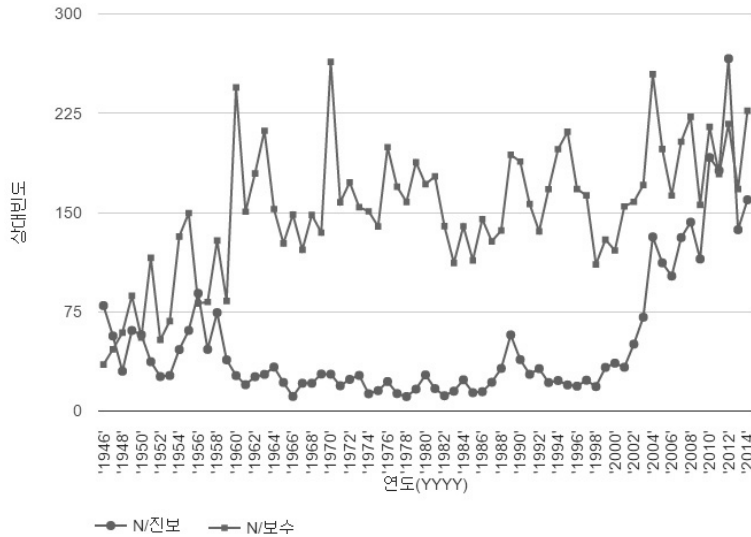


터 1980년대 말로 추정할 수 있다. 물론 남북 관계가 한때 좋았던 1973년도에는 ‘북한’이 주로 쓰였다는 점도 흥미롭다. 결국 ‘북괴’는 통치 수단
 의 일환으로 도입된 용어라고도 볼 수 있을 것이다.

4.5.2. ‘진보’의 약진

‘진보’라는 어휘가 정치적으로 사용되기 시작한 것은 그 연원이 깊지 않다. 진보는 오랜 기간에 걸쳐 정치적인 ‘보수’와 대립되는 의미로 사용되기보다는 ‘기술의 진보’, ‘역사의 진보’ 등과 같이 일반적인 용어로 더 많이 사용되어 왔다. 그러다가 1980년대를 거치면서 ‘보수’에 대립되는 정치적 의미로 사용되기 시작하면서 2000년대 들어서는 그 빈도가 폭발적으로 높아진다. 특히 2010년 이후부터는 정치적 ‘진보’가 거의 ‘보수’를 넘어설 정도의 빈도를 보이고 있다는 점에서 ‘진보’에 대한

[그림 35] ‘진보’와 ‘보수’의 연도별 사용 빈도

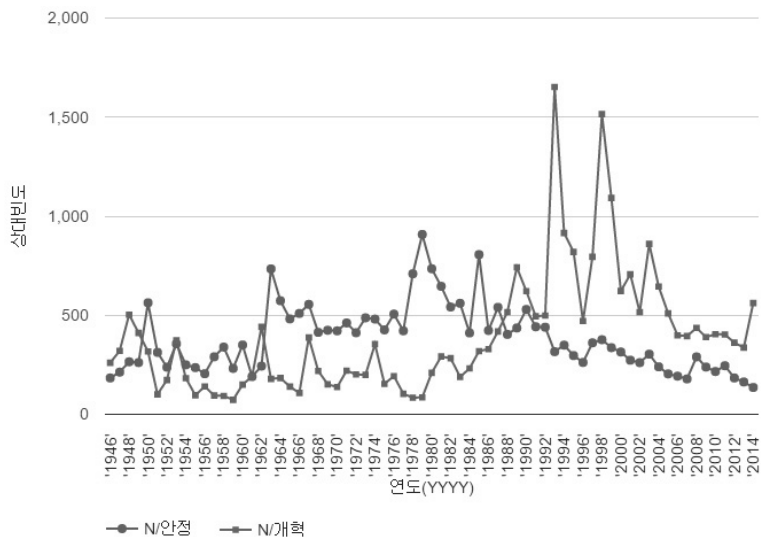


정치적 관심이 크게 증가한 것으로 해석된다.

4.5.3. 누구나 ‘개혁’을 외친다

일반적으로 사람들의 심리에는 변화보다는 안정을 선호하는 경향이 있다. 특히 극심한 사회적 혼란을 경험한 한국 사회에서는 변화보다 안정을 선호하는 경향이 두드러졌으며, 이는 북한의 위협과 결부되어 ‘안정’의 추구가 곧 우리의 생존과 직결되는 것으로 여겨지기도 하였다. 여당에서 선거 때마다 ‘안정’론을 내세워 정권 유지에 성공했던 기억이 오래지 않다. 그러나 이러한 경향은 1990년대 들어 크게 달라진다. 우선 ‘개혁’의 빈도가 ‘안정’을 크게 앞지르기 시작했고 그러한 경향은 2000년대 들어 더욱 심화되었다. 개혁은 이제 야권에서 부르짖던 구호가 아니라 여야 누구나 할 것 없이 ‘개혁’을 거론하게 되었다. 우리 사회는 여

[그림 36] ‘안정’과 ‘개혁’의 연도별 사용 빈도



전히 개혁해야 할 것이 많다고 해석해야 할까?

5. 결론

한국 사회 변화를 기록한 대표적인 자원 중 하나인 신문 자원을 활용할 경우 시대의 흐름을 여러 가지 측면에서 손쉽게 살펴볼 수 있다. 단순한 호기심 차원에서 관심이 갈 수도 있고, 또는 본격적인 연구를 위해서도 매우 흥미로운 자원의 보고라 할 수 있다. 지금까지 살펴본 내용들은 그야말로 빙산의 일각으로 여기에서 제기된 점들만으로도 무궁무진한 방향으로 추가 연구가 가능할 것이다. 특히 ‘경이적’이란 말로 표현되는 광복 이후 한국 사회의 탈바꿈은 앞으로도 오랫동안 탐구해야 할 대상이라 할 것이다.

앞으로 70년은 또 어떤 흐름을 이어갈 것인가? 이러한 자원을 활용하면 미래에 대한 예측도 가능할 것인가? 이는 매우 방대한 주제로 여기에서 단순하게 답을 할 수 없는 문제다. 이에 대하여 원론적인 차원에서 두 가지만 추가하자면, 우선 인문과학적 주제나 사회과학적 주제에서 예측 행위가 그동안 별로 성공적이지 못했다. 국내 저명한 사회과학자 중 한 사람은 7~8년 전의 경제 위기 시 사회과학이 그러한 위기를 예측해 내지 못한 점에 대하여 자괴감을 토로한 적이 있다. 공학이나 자연과학과는 달리 인문사회학적 주제에서는 ‘과학적’ 예측이 별로 활성화되지 못했다.

미래 예측과 관련해서 둘째로 제기하고 싶은 문제는 신문이 사회를 반영하기도 하지만 반면 ‘여론 선도’라는 측면도 있다는 점이다. 즉, 흐름을 만들어 가려는 시도를 끊임없이 하고 있다. 언어가 사회의 반영이

면서도 사회를 구축하는 도구라는 점은 이미 잘 알려져 있다. 그렇게 본다면 시대의 흐름이란 주어진 것이 아니라 만들어 가는 것이라고도 규정될 수 있다. 앞으로의 70년도 저절로 만들어지는 것이 아니라, 세계사적 환경 속에서 한국민이 어떻게 만들어 가느냐에 달려 있다고 볼 수 있다.

참고 문헌

- 김일환(2015), 한국 사회의 키워드: 1990년대와 2000년대의 언어 사용, 《아시아 문화연구》 37권, 5~33.
- 김일환· 이도길(2011), 대구모 신문기사의 자동 키워드 추출과 분석, 《한국어학》 53권, 145~193.
- 김일환· 정유진(2013), ‘대선’ 관련어의 추이 분석과 전망, 《민족문화연구》 58, 791~819.
- 김일환· 정유진· 강범모· 김홍규(2013), 《‘물결21’ 코퍼스의 구축과 활용》, 소명출판.
- 김일환(2015), 광복 이후부터 한국전쟁 직후까지의 핵심어 분석-동아일보 1946~1955년 기사를 대상으로, 《국제어문》 66집, 9~36.
- 네이트 실버, 이경식 옮김(2014), 《신호와 소음》, 더퀘스트
- 송길영(2012), 《여기에 당신의 욕망이 보인다》, 쌤앤파커스
- 에리즈 에이든· 장바티스트 미셸, 김재중 옮김(2015), 《빅데이터 인문학: 진격의 서막》, 사계절.
- 최재용· 이도길(2014), 물결21 코퍼스: 공개 웹 자원 및 활용 도구, 《민족문화연구》 64호, 3~23쪽.
- Feldman R. & J. Sanger(2007), *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press.
- Lee, Do-Gil & Hae-Chang Rim(July 2009), Probabilistic Modeling of Korean Morphology, *IEEE Transactions on Audio, Speech, and Language Processing*, 17(5), pp.945~955.
- Williams R.(1976), *Keywords: a Vocabulary of Culture and Society*, Flamingo.