

# 멀리서 읽는 “저희”\*

## — ELECTRA 기반 의미 분류 모델과 N-gram을 활용한 두 가지 ‘저희’의 통시적 변화 양상 연구

徐在玄\*\* · 河나\*\*\* · 金民雨\*\*\*\* · 韓智敏\*\*\*\*\* · 金昞俊\*\*\*\*\* · 朴素晶\*\*\*\*\*

I. 서론: “저희” 멀리서 읽기의 필요성  
II. 초기 근대 소설에 나타나는 ‘저희’의 두 가지 유형과 분류 기준  
III. 1954년~2024년의 신문데이터에 나타나는 ‘저희’ 용례의 통시적 변화  
IV. 결론

### • 국문초록

본 논문에서는 멀리서 읽기(distant reading)로 통칭되는 정량적 연구방법론을 채택하여 근대 소설 텍스트에 나타나는 두 가지 ‘저희’의 특징을 분석하고, 이를 오늘날에 이르는 확장된 데이터에 적용하여 ‘저희’의 사용 추이를 통시적으로 고찰함으로써 전체적인 변화 양상을 파악하고자 한다. 이를 위해 ELECTRA 기반 의미 분류 모델과 N-gram 등의 디지털방법론을 활용하여 약 1954년부터 2024년까지의 조선·동아 신

\* 본 연구는 선행연구인 “멀리서 읽는 “우리”(서재현 외, 2021)”의 후속연구로, 본 연구의 2장에서 활용한 근대 소설 데이터는 선행연구에서 활용하였던 데이터를 기반으로 하였음을 밝힌다. 활용한 데이터에 관련한 내용은 선행연구의 362~371면 참조.

\*\* 성균관대학교 유학동양한국철학과 박사수료, 제1저자

\*\*\* 성균관대학교 한국철학문화연구소 책임연구원, 공동저자

\*\*\*\* 성균관대학교 인간AI인터랙션융합전공 석사, 공동저자

\*\*\*\*\* 성균관대학교 심리학과 석사, 공동저자

\*\*\*\*\* 한국학중앙연구원 인문정보학 조교수, 공동저자

\*\*\*\*\* 성균관대학교 유학동양한국철학과 교수, 교신저자

문데이터에 나타나는 ‘저희’의 두 유형이 그리는 추이를 거시적으로 들여다본다. 본 논문은 기존에 산발적이고 한정적으로만 언급되었던 ‘저희’ 사용의 큰 흐름을 확인함으로써 ‘저희’의 용법이 정착되는 과정과 의미상의 변화를 탐색하였다. 또한 각 시대 별로 두 가지 ‘저희’의 연어 분석을 진행하여 각 용법과 연관되는 어휘들의 풍부한 사례들을 제공함으로써 향후 ‘저희’ 연구를 위한 토대를 마련하고 방향성을 제시하고자 하였다.

주제어 : ELECTRA 기반 의미 분류 모델, N-gram, 저희, 멀리서 읽기, 근대 소설, 신문데이터

## I. 서론: “저희” 멀리서 읽기의 필요성

오늘날 ‘저희’는 “우리”의 겸양어로 이해되며 일상생활에서 광범위하게 사용된다. 그런데 “우리”는 늦어도 10세기 중반부터 용례가 발견되며 현재에 이르기까지 일관되게 사용되어온 1인칭 대명사<sup>1)</sup>인 반면, ‘저희’는 사용 시기도 늦고 “우리”에 완전히 호응되는 겸양어 혹은 낮춤말이라고 하기에는 상당한 의미상의 편차를 보인다.<sup>2)</sup> 연구자들은 대체로 개화기에 구어체 문장이 시도되던 신문 광고 및 번역서 등에서 ‘저희’가 처음 사용되기 시작하였다고 보는데,<sup>3)</sup> 흥미롭게도 당시에는 3인칭 ‘저희’(=저들)와 1인칭 ‘저희’(“우리”의 겸양어)가 혼용되었음을 공통적으로 지적하고 있다.

현대 국어에서는 당연하게도 1인칭으로 여겨지는 ‘저희’가 처음 사용되기 시작하던 시기에는 3인칭으로 쓰였다는 사실<sup>4)</sup>은 학문적인 검토를 요한다. 또한 뒤늦게 “우리”의 겸양어로 쓰이게 된 ‘저희’와 “우리” 사이에 범위와 의미가 호응하지 않는다는 사실로부터 실제 언어생활에 있어서의 혼란도 야기되고 있는 실정이다.<sup>5)</sup> 본

- 1) 김미형, 「국어 대명사의 어휘사」, 『한국어 의미학』 9, 한국어의미학회, 2001의 <표 2·3·4·5> 참조.
- 2) 김정남의 연구(2003)는 “우리”를 주제로 한 것이기는 하지만 낮춤형인 ‘저희’와 비교함으로써 본 연구에 시사점을 준다. 요약하면 첫째 ‘저희’는 적극적인 [낮춤]의 의미를 가지지만 “우리”는 그 반대, 즉 적극적인 [높임]의 의미를 가지지 않는다는 점, 둘째 ‘저희’는 청자를 포함할 수 없지만, “우리”는 청자를 포함한다는 점 등을 들어 ‘저희’를 단순히 “우리”의 낮춤형이라고 보았을 때 맞닥뜨리는 문제점을 지적하고 있다 (김정남, 「한국어 대명사 우리의 의미와 용법」, 『한국어 의미학』 13, 한국어의미학회, 2003, 257~274면).
- 3) 정길남(2004); 채완(2003); 전무용(2006) 등 참조. 특히 전무용의 경우, 시기에 따라 같은 내용이 달리 표현되는 번역어를 추적할 수 있는 대상인 성경을 연구 대상으로 삼는다(정길남, 「개화기 교과서의 존재법 연구」, 『동악어문학』 42, 동악어문학회, 2004; 채완, 「개화기 광고문의 표현 기법」, 『한국어 의미학』 12, 한국어의미학회, 2003; 전무용, 「한국어 성경의 대명사 고찰」, 『성경 원문연구』 19, (재)대한성서공회 성경번역연구소, 2006).
- 4) 전무용(2006)에 따르면 <구역>과 <개역>(1938)과 <개역한글판> 성경에서는 3인칭을 가리키는 대명사로 “그/그들”도 쓰고 있고, “저/저희”도 쓰고 있다. 현대 국어로 올수록 이러한 3인칭의 자리에는 점차 “그”를 쓰고 있고, “저/저희”는 “자기를 낮추어 가리키는 1인칭 대명사”로 사용되고 있다(전무용, 앞의 2006 논문, 102면).
- 5) 가령 ‘우리나라’를 ‘저희 나라’라고 불러도 되는지에 대한 문제가 기사화되기도 하고 국립국어원에서 논란이 된 바 있다. 김정남(2003)의 경우 화자가 한국인이더라도 청자가 한국인이 아니라면 서로 ‘저희 나라’라고 표현해도 무방하다는 입장이다(김정남, 앞의 2003 논문, 268면).

연구는 이러한 문제들을 해결하기 위해 ‘저희’에 대한 포괄적이고 통시적인 연구가 필요하다는 인식에서 출발한다. 다만 아직까지 ‘저희’를 주제로 하여 본격적으로 다룬 연구가 없고, “우리”에 주목하거나 한국어 지시대명사를 고찰하는 과정에서 부차적이고 산발적으로 언급되어 왔을 뿐이므로 본 논문에서는 두 가지 ‘저희’가 어떻게 실제로 사용되고 변천해 왔는지에 대한 자료를 수집, 분석하여 제공함으로써 본격적인 연구의 기초가 되는 큰 그림을 제공하고자 한다.

본 논문에서는 ‘저희’의 용례들을 살펴서 얻어진 유형화의 결과로서 두 가지 ‘저희’를 각각 “겸양의 저희”와 “저편의 저희”로 부르기로 한다.<sup>6)</sup> 2장에서는 먼저 초기 근대 소설 텍스트 데이터에 나타나는 ‘저희’의 두 가지 유형을 살펴보고 각각의 특징을 파악하여 분류 기준을 마련하였다. 이로써 개화기의 특정 문헌에 주목한 기존 연구의 양적 한계를 극복함과 동시에 기존 연구에서 언급되지 않았던 ‘저희’에 대한 다양한 용례들을 확인함으로써 ‘저희’ 분류 모델을 만들 수 있는 질적 기준을 마련하여 대규모 데이터를 분석할 수 있는 토대를 갖추었다.

3장에서는 지난 70년간(1954년~2024)의 조선·동아일보 신문데이터로부터 두 가지 유형의 ‘저희’가 그리는 통시적 변화를 파악하고자 하였다. 선행연구에서는 ‘저희’의 특성을 분석하기 위해 한정된 문장에 대한 질적 분석을 위주로 하였으나, 오랜 시간 축적되어온 대량의 신문데이터를 관찰하면 보다 다양한 용례의 통시적 변천 과정을 살필 수 있을 것이라는 기대에서다. 이를 위하여 두 유형의 ‘저희’를 이진 분류하는 데 최적화된 ELECTRA 기반의 분류 모델을 개발하였다. 이 과정에서 근대 소설 데이터를 검토하며 얻은 기준으로 분류된 데이터를 활용하여 파인튜닝(Fine-tuning)을 시도하였다.

분류 모델을 활용하여 ‘저희’를 포함한 다량의 문장들이 각각 어느 유형에 속하는 지 자동 분류하고, 그 결과를 시각화하여 전체 추이를 한눈에 파악하고자 하였다.

6) 이제까지 ‘저희’가 언급된 연구에서는 두 가지 ‘저희’를 가리키는 용어가 연구자마다 서로 달라 독자들에게는 다소 혼란스러울 수 있다는 점을 고려하였다. 가령 김미형은 재귀칭 대명사 혹은 겸사말, 김정남은 낮춤말, 정길남은 겸양어, 겸양 대명사, 1인칭 대명사, 3인칭 대명사 등의 다양한 표현을 사용하며, 양영희는 재귀대명사로 보는 관점 대신 우선적으로 3인칭 대명사로 보아야 한다고 제안하는 등 학계에서 ‘저희’를 부르는 이름은 통일되어 있지 않다(김미형, 앞의 2001 논문; 김정남, 앞의 2003 논문; 정길남, 앞의 2004 논문; 양영희, 「중세 국어 ‘재귀대명사’ 설정에 대한 재고」, 『한글』 273, 한글학회, 2006).

또한 분류 모델에 이어, N-gram 방법론을 통하여 ‘전체’ 저희, ‘저편’과 ‘겸양’의 저희로 각각의 유형을 구분했을 때 등장하는 용례를 파악함으로써 통시적 흐름에서 ‘저희’와 그 연어의 쓰임이 어떤 양상으로 변화해오고 있는지 확인하고자 하였다.

이처럼 대량의 데이터를 통해 두 가지 ‘저희’를 살펴보면, 각각의 ‘저희’가 어떤 용례 상, 추이 상의 차이를 가지는지를 파악할 수 있을 것이다. 특히, N-gram 방법론을 활용하여 용례 변화를 연속적으로 추적하는 작업을 통해서 사람들이 같은 단어를 두 가지 다른 유형으로 표현할 때의 맥락과 그 변화상을 포착할 것이다.

이를 통해, 각 시기별 용례의 지속 혹은 단절 양상과 그 맥락을 파악한다면, 이는 이어지는 후속 연구에서 같은 모습의 두 가지 ‘저희’가 상호 간에 어떤 영향을 끼쳤는지 등을 추정하고 그 기저의 사유 방식을 이해하는 데에 핵심적인 토대가 될 것이다. 또한, ‘저희’에 대한 통시적인 연구를 위해 개발한 두 가지 유형 분류를 위한 ELECTRA 모델과 이어지는 언어 변화를 파악할 수 있는 N-gram 방법론, 그리고 수집한 데이터 등은 추후 본 연구와 유사한 ‘우리말 관념의 통시적 변화’를 추적하고자 하는 여러 후속 연구들에도 기여할 것으로 기대한다.

## II. 초기 근대 소설에 나타나는 ‘저희’의 두 가지 유형과 분류 기준

선행연구들을 통해 저희의 쓰임이 크게 두 갈래로 나뉘고 있음을 알 수 있었다. 한 갈래는 3인칭 지시어로서의 ‘저희’이고, 다른 한 갈래는 1인칭 겸양어로 사용된 ‘저희’이다. 본 연구에서는 이 두 가지 용법의 저희를 각각 ‘저편의 저희’와 ‘겸양의 저희’로 명명하여 다룰 것이며, ‘저희’의 두 양상을 함께 살펴볼 것이다.

이 장에서는 우리말로 ‘저희’라는 표현이 본격적으로 등장하던 시기의 근대 소설 텍스트 데이터를 살펴봄으로써 두 가지 ‘저희’ 용례의 출현 빈도와 특징 등을 파악

---

7) 익명의 심사자께서 지적해주셨듯, 1인칭 ‘저’와 복수형 ‘저희(들)’가 재귀대명사 용법으로 쓰이는 경우와 3인칭 원칭 지시사 ‘더’와 복수형 ‘더(희)들’이 구개음화를 거친 ‘저’와 복수형 ‘저(희)들’이 사람을 가리킬 경우 모두 동일한 형태, ‘저희(들)’로 등장한다. 이 두 경우 모두 ‘3인칭 대명사에 가까운 용법’으로 나타나는데, 본 연구에서는 이 두 경우를 직관적으로 구분하기 위하여 1인칭 ‘저희’를 ‘겸양의 저희’, 3인칭 ‘저희’를 ‘저편의 저희’로 명명하여 다루었다.

하고자 한다.

‘우리’의 겸양 표현으로 사용된 저희(이하, 겸양의 저희)를 다룬 연구에서는 대체로 어법상의 특징, 특정 시기에 연속적으로 간행된 간행물에서 보이는 특징<sup>8)</sup> 등을 근거로 ‘저희’가 사용되어온 기간이 비교적 짧을 것으로 추정하고 있다.<sup>9)</sup> 개화기에 등장한 우리말 소설들을 살펴보면 ‘저희’라는 말이 본격적으로 사용되기 시작하던 당시의 특징들을 살필 수 있을 것이다. 따라서, 1900년대 초기에서 중기에 이르는 근대 소설 텍스트 데이터에서 보이는 ‘저희’의 특성을 살피고, 그중에서도 ‘저희’의 용례가 발견되는 몇몇 소설들을 중심으로 양상을 파악하고자 한다.

〈표 1〉 근대 소설 데이터에서 ‘저희’의 출현 양상

시대 구분	전체	겸양의 저희	작품 수	저편의 저희	작품 수
E1 (1906~1919)	32	16	4	16	4
E2 (1920~1945)	216	79	25	137	57
E3 (1945~1957)	76	62	6	14	10
합 계	324	157	35	167	70

1900년대 초부터 중반 무렵까지의 근대 소설 데이터에서 두 가지 ‘저희’의 초기 사용 양상 및 시대 구분<sup>10)</sup>에 따른 비중 변화를 확인하였다. ‘저희’가 포함된 324개

8) 신문, 성경 등이 주 연구 대상으로 다뤄진다. 채완은 개화기 신문 광고로부터 겸양어 ‘저희’의 출현을 확인하였고 성경에서의 정길남은 1900년에 간행된 성서인 『신약전서』에서 1인칭 겸양어 ‘저’를, 전무용은 『성경전서』로부터 같은 ‘저’이더라도, 1인칭 겸양어와 3인칭 지시어로 사용될 때 그 표기가 각각 ‘저’와 ‘더’로 구분됨을 밝혔다(채완, 앞의 2003 논문; 정길남, 앞의 2006 논문).

9) 19세기 말에 들어서야 비로소 1인칭의 예사말과 겸사말이 구분되기 시작하는데, 그 중 대표적인 예가 1인칭 대명사 ‘저’이다. 중세 국어의 규칙에서 어미, 조사의 차이를 통해 존대를 나타낸다는 점에 비춰볼 때, 주체를 낮추는 ‘저희’는 무척 낮은 상황으로 고려된다. 조사 또는 동사에 표기하는 방식을 통해 존칭을 표기하던 고대와 중세 국어로 거슬러 올라가면 그 양상은 더욱 명확해진다(양영희, 앞의 2006 논문, 73~97면).

10) 시대 구분의 기준에 대해서는 다음 논문을 참조하라(서재현 외, 「멀리서 읽는 “우리”-Word2Vec, N-gram을 이용한 근대 소설 텍스트 분석」, 『대동문화연구』 115, 대동문화연구원, 2021, 371~2면 참조).

의 문장을 직접 수기 분류하여 검토한 결과, 이 중 157건이 ‘겸양의 저희’에 해당하며, 167건은 ‘저편의 저희’ 용법으로 사용된 것임을 알 수 있었다.

E1(1906~1919)시기에는 ‘저희’의 쓰임이 동일한 비율로 양분되고 있었으며, E2시기에는 저편의 저희가 우위를 점하다, E3시기에는 겸양의 저희가 더 많이 사용됨을 알 수 있었다. E1에서부터 E3시기까지 전체 시기를 놓고 봤을 때, 두 가지 ‘저희’는 대체로 유사한 비중으로 나타났다. 1900년대 중반 무렵까지는 ‘저편의 저희’ 역시 ‘겸양의 저희’만큼이나 빈번하게 사용되어왔다고 할 수 있다. 한편, ‘저편의 저희’는 다양한 작품에서 산발적으로 나타났지만, ‘겸양의 저희’는 저편의 저희가 나타난 작품 수의 절반 정도에서 전체 빈도가 유사하게 나타났다. 이는 ‘겸양의 저희’는 특정 작품에서 집중적으로 사용된 반면, ‘저편의 저희’는 비교적 다양한 맥락에서 사용되었을 가능성을 시사한다. 구체적으로 살펴보자면, E1(1906~1919) 시기에는 두 가지 쓰임이 동일한 비율로 나타나며, 특히 『무정』과 『원앙도』 두 소설에서 용례가 주로 발견되었다. 앞서 소개한 저희의 출현 배경을 다룬 선행연구들에 따르면 이 시기는 ‘겸양의 저희’가 등장하는 시기라고 할 수 있다. 그런 점에서 이 두 소설은 비록 소수의 사례이긴 하지만 겸양의 저희가 어떻게 쓰이기 시작했는지를 확인해볼 수 있는 자료라고 할 수 있다. 이에 대해서는 아래에서 자세하게 살펴 보도록 하겠다.

E2(1920~1945)시기에는 ‘저희’가 지시대명사로 사용된 경우가 압도적으로 많았는데, 소설의 종류에 있어서도 약 두 배 이상의 다양성을 띄어 복수의 저자들에게 의해 ‘저편의 저희’의 용법이 많이 사용되었음을 확인할 수 있었다. 한편, ‘겸양의 저희’가 사용되는 경우에는 청자가 ‘선생님, 스승, 하느님, 장모님, 어머님, 아버지’ 등으로 한정되었다.

마지막으로, E3(1946~57)시기에는 31:7의 비율로 상당한 양의 ‘겸양의 저희’가 사용되었다. 그러나 작품의 총수가 6건에 그쳤고 김내성에 의해 쓰인 『애인』과 『청춘극장』, 그리고 『백조의 곡』에서 각각 39건, 15건, 3건, 모두 57건이 출현하였다. 이는 약 91%가 동일 작가의 작품에서 집중적으로 나타난 것이므로, 이를 해당 시기에 ‘겸양의 저희’가 지배적으로 사용된 현상으로 해석하기는 어려움이 있다.

지금까지 시대별로 보이는 두 가지 저희의 출현 빈도와 양상을 위주로 설명했다면 아래에서는 각 작품에서 보이는 실제 사례들을 중심으로 그 특징과 용례를 살펴

보겠다.

### 1. '겸양의 저희'의 특징 및 용례

'저희'가 1인칭 겸양어로 사용된 용례를 근대 소설에서 살펴보면 주로 1인칭 주어의 자리에서 높임의 대상이 되는 청자에 대해 자신을 낮추는 기능을 하고 있음을 알 수 있다. 특히 각각의 소설이 설정하고 있는 시대적 배경을 감안하였을 때, '장군, 선생님, 신, 대왕' 등에 대하여 자신을 포함한 자신의 무리(병졸, 제자, 신도, 소인 등의 무리)를 낮춰 말할 때 사용됨을 알 수 있다.

(1) 화자와 청자를 구분하여 자신을 낮추는 경우

“예, **저희 무리**는 변산 도적이옵고, 소인은 팔도 도적의 도두목 체천대왕 조곰보라 하옵니다. …〈중략〉… 수 만 명 도적을 먹일 길이 바이 없사와서 오직 대왕님만 기다린 것이오니 대왕마마께옵서는 **저희 무리**를 불쌍히 여기시와 오늘부터 저희 무리의 왕이 되시옵고, 저희 무리에게 살길을 접지하시옵소서.”  
(『허생전』)

이렇듯 화자와 청자를 명확하게 구분하는 방식으로 '겸양의 저희'가 사용되는 양상은 근대 소설이 등장하던 초기에서부터 확인된다. 대표적으로 1911년에 발행된 이해조의 『원앙도』와 1917년에 발행된 이광수의 『무정』에 등장하는 '저희'의 용례를 통해 확인할 수 있다.

“에그 참말, 그것을 언다 두셨습니까? **저희**들 생각에는 사또께서 그것을 찾느라고 법석을 하실뿐더러 저희에게지도 의심이 돌아올 줄 알았더니, 지금껏 모르시는지, 알고도 모르는 체하시는지 알 수 없어요.” (『원앙도』)

이 나폴레옹 승배자는 형식에게 인사한 뒤에 엄연히 꿇어앉아, “**저희**가 선생님을 뵈오려 온 뜻은 ……” 하고 말을 시작한다. (『무정』)

『월양도』에 나타나는 겸양의 저희의 경우, 군수/기생처럼 사회적 계급에 있어서 윗사람과 아랫사람으로 신분 등이 확실하게 구분되는 경우, 신분이 낮은 아랫사람이 자신들을 낮추기 위해 저희를 사용하는 경우였다. 한편, 『무정』에 나타나는 겸양의 저희의 경우, 선생님/학생, 하느님/죄인 등 사회적 역할이 윗사람과 아랫사람으로 구분되는 관계에서, 그리고 종교적 헌신의 대상 앞에서 상대적으로 자신들을 낮추기 위해 ‘저희’를 사용하는 경우였다. 즉, 이 두 소설에 등장하는 ‘겸양의 저희’는 화자가 이 대상들에 대하여 자신을 포함한 무리를 명확히 구분 짓는 경우에 한하여서 ‘겸양의 저희’가 사용됨을 알 수 있었다.

(2) ‘저희’와 ‘우리’를 적재적소에 활용하여 각 인칭대명사에 포함된 사람들의 범위를 유연하게 조정하는 경우

선생님 **우리**는 정말 어떻게면 좋겠습니까? **저희들**의 갈 길을 진심으로, 꾸밈없이 단 한 마디만 깨우쳐 주십시오! (『청춘극장』)

선생님, **저희들**의 이 심각한 문제를 제 삼자로서는 한낱 강 건너 화재쯤으로 밖에는 더 생각키지 않을런지 모릅니다만 그러나 오늘날 **우리들**의 학도병 문제로 말하면 당자인 **우리들**만의 문제가 아니고 **우리** 삼천만 민족 전체의 문제입니다. (『청춘극장』)

**저희들**의 천사인 채영신 누이를 언제까지나 언제까지나 **우리** 청석골에서 떠나지 않도록 붙들어 주시옵소서! (『상록수』)

한편, ‘겸양의 저희’의 용례 중 눈여겨볼 만한 흥미로운 사례가 발견되었다. ‘겸양의 저희’가 1인칭 ‘우리’의 겸양어로 사용되에도 불구하고 ‘겸양의 저희’와 ‘우리’가 같은 문장에서 함께 사용된 경우가 발견된 것이다. 해당 문장들을 살펴보면, 자신보다 지위나 신분이 높은 청자를 배제한 상태에서 자신을 낮추어 지칭하는 ‘겸양의 저희’와 청자를 포함하는 ‘우리’를 명확히 구별하여 적재적소에 사용하고 있다는 점을 알 수 있다.

가령 『청춘극장』의 용례를 보면 선생님과 화자의 무리를 함께 포함할 때는 ‘우리’로 묶는 한편, (선생님으로부터 조언을 구해야 하는 대상인) 자신들의 무리를 특정

할 때는 ‘저희’를 사용하는 것을 확인할 수 있다. 위 문장이 출현하는 작중<sup>11)</sup> 선생님과 학생 간의 대화에서 ‘우리’와 ‘저희’는 화자의 의도에 따라 적절히 배치된다. 위 예시에서 운명을 공유하는 학도병, 선생을 포함하여 우리 민족을 지칭하는 경우에는 ‘우리’가, 선생님 앞에서 공손함을 드러내며 학생들만을 지칭할 때는 ‘저희’가 사용된다. 이어서, 소극적으로 에둘러 대답하는 선생님에 대하여 실망을 드러낼 때는 의도적으로 ‘공손함’을 표하지 않기 위해 또다시 ‘우리’를 사용한다. 이러한 용례를 통해, ‘저희’가 등장하기 시작한 시기 1인칭 대명사 ‘우리’와 그것의 겸양어 ‘저희’가 각 인칭대명사에 포함된 사람들의 범위를 유연하게 조정하기 위하여 적절하게 사용되어왔음을 확인할 수 있었다.

## 2. 우리편과 구분되는 ‘저편의 저희’

‘저희’가 3인칭 지시어로 사용되는 경우, 그 용법은 ‘저들’과 동일한 것으로 ‘우리’ 혹은 ‘나’에 대해 대비를 이루며 사용되었다. 이 용법은 대표적으로 『박명(薄命)』(한용운), 『환희』(나도향), 『사랑의 동명왕』, 『이순신』, 『재생』(이광수), 『인간문제』(강경애), 『흑치상지』(현진건), 『원앙도』(이해조) 등을 비롯하여 여러 소설 전반에 걸쳐 나타남을 알 수 있다. 그 중 ‘저편의 저희’의 쓰임이 명료하게 드러나는 사례는 다음과 같다.

바른 말이지 나는 당장과 당병은 그렇게 밍지 않소. **저희들도 저희 나라**를 위하여 피를 흘리는 것이니, **우리**에게 적은 될지언정 원수가 될 것 있소? 그런데 이 충상영이란 놈은 우리의 원수가 아니요? (『흑치상지』)

적병이 어제 이 포구에 들어 왔소. 어제 이 포구에 들어 와서 여염으로 돌아다니면서 재물과 우마를 약탈하여다가 지금 싸포께서 불살라 버리신 **저희들**의 배에다 갈라 신고 …〈중략〉… 놀기를 밤새도록 하였소. (『이순신』)

『흑치상지』에서는 저희(당나라 병사)와 우리가 대비를 이루고 있고, 『이순신』에

11) 『청춘극장』 2권 33장: 민족의 고달픔.

서는 저희를 통해 적병을 가리키고 있음을 알 수 있다. 이 두 가지 용례에서 볼 수 있듯이, ‘저희’는 명시적으로 ‘적’, ‘원수’처럼 ‘우리’와 대비를 이루는 상대편을 가리키는 상황에서 지시어로 기능하고 있음을 알 수 있다. 그런데, 이 용례를 지시대명사라고만 규정하는 데에는 부족함이 있다. 『흑치상지』에서 보이는 ‘저희 나라’의 예시에서와 같이 지시대명사의 역할만이 아니라 명사를 한정하는 기능을 가지고 있기 때문이다. 이는 우리말 지시어 ‘이’ - ‘그’ - ‘저(더)’ 중에서 원칭(遠稱)을 가리키는 ‘저(더)’에 해당하는 것으로 볼 수 있다. 바로 이러한 측면에서 본 연구에서는 ‘겸양의 저희’가 아닌 또 하나의 ‘저희’를 ‘3인칭 대명사’ 또는 ‘지시대명사’<sup>12)</sup>가 아닌 ‘저편의 저희’로 명명하였다.

1900년대 초에서부터 보이는 근대 소설 속 이러한 ‘저희’의 활용은 구어에 존재하던 ‘나’, ‘너’, ‘우리’ 혹은, ‘이편’ 또는 ‘저편’과 같은 구분이 문어로 옮겨지는 가운데 등장한 것으로 보인다. 즉, ‘나’보다 높은 청자인 ‘너’ 앞에서 ‘나’를 낮추기 위하여 사용된 ‘겸양의 저희’와 ‘우리들 편(이편)’과 ‘저들 편(저편)’을 나눌 때 사용된 ‘저편의 저희’를 통해 당시 문인들이 ‘저희’를 여러 형태로 활용하기 시작했음을 알 수 있다. 우리는 이러한 근대 소설 텍스트를 통해 이전부터 공유해왔던 우리말의 흔적이 당대 문인들에 의해 고안된 어휘를 통해 우리말로 옮겨지는 과정을 들여다 볼 수 있었다. 이와 같은 경향성은 문학작품이라는 제한된 공간 안에서 작가들의 문체가 강하게 반영된 것이다.

지금까지 근대 소설 텍스트를 통해 ‘저희’가 등장하기 시작하던 초기 용례부터 1957년까지의 흐름을 ‘가까이 읽기’를 통해 살펴보았다. 그 결과, 이 시기 동안 ‘저편의 저희’와 ‘겸양의 저희’라는 두 가지 용례가 공존하고 있음을 발견하였다. 이 데이터에서는 두 유형의 ‘저희’가 비슷한 빈도로 나타나, 각각의 특징을 분석하고 비교하는 데 적합한 자료가 되었다.

여기서 파악한 두 가지 ‘저희’의 특징은 다음과 같다. 먼저, ‘겸양의 저희’의 경우 주로 주어 자리에 쓰여 화자가 자신을 낮추는 용법으로, ‘높임의 대상이 되는 청자

12) 저희를 재귀적 용법으로만 분류하던 이전 연구(홍순성, 「한국어 재귀대명사의 특징」, 『동서인문학』 19, 인문과학연구소, 1987, 249~263면)와는 달리, 3인칭 지시어로서의 특성이 재귀어의 특성보다 더 포괄적인 상위 범주라는 점을 근거로, 서구 문법을 그대로 받아들인 재귀대명사 개념을 대신하여 ‘3인칭 지시어’로 칭할 것을 제안한다(양영희, 앞의 2006 논문, 73~97면).

가 노출'되는 경우가 많았다. 이때 계급을 명확하게 구분하는 어휘가 등장해 화자-청자 관계를 더욱 분명히 드러내기도 하였다. '저희'를 포함한 문장에서는 대부분 존대를 표하는 어미와 겸사를 사용하며, 때로는 극존칭까지 활용되는 것을 확인할 수 있었다.

한편, '저편의 저희'의 경우 대부분 비판적이거나 부정적 맥락에서 '적대시할 상대 집단을 지칭'하는 목적으로 사용되고 있었으며, '적', '원수', '왜놈' 등 우리편과 대립하는 상대편을 지칭하고 있었다. 따라서 '저편의 저희'를 대신하여 '저편'으로 치환하여도 의미가 잘 통하였다. '저편의 저희' 뒤에는 '-끼리·-들끼리·-네·-치·-놈' 등의 접미사가 자주 위치했으며, 동물을 지칭하는 경우도 눈에 띄었다.

'저희'와 '우리'가 함께 나타나는 문장의 경우 이 둘이 지칭하는 집단이 동일 집단인지 아닌지를 기준으로 삼으면 구분이 용이하였다. 앞서 살핀 바와 같이 같은 집단을 가리키고 있다면, '겸양의 저희'로 사용되어 청자에 대한 특정(불만 표출 등) 의도와 뉘앙스가 반영되고 있을 가능성이 컸다. 한편, 이 둘이 다른 집단을 지칭하고 있다면 위에서 상술한 특징들을 근거로 삼아 '겸양', '저편' 여부를 판별할 수 있었다. '겸양의 저희'의 경우에 '저희'는 '우리'에 비해 작은 집단을 가리키고 있는 것이며, '저편의 저희'의 경우에 '저희'는 '우리(편)'에 상대되는 상대편을 가리키는 것으로 볼 수 있다.

그렇다면, 근대 소설에서 비슷한 비중으로 보이던 '저편의 저희'와 '겸양의 저희'는 오늘날에 이르기까지 어떤 방식으로 자리 잡아왔을까? 이러한 변화는 사회적 언어 사용의 흐름에서 어떻게 설명될 수 있을까? 이 질문에 답하기 위해서는 근대 소설 데이터의 시대 범위 이후로 이어지는 데이터를 통시적으로 살펴볼 필요가 있었다. 또한, 작가들의 문체가 투영된 근대 소설의 작품 공간을 넘어, 일상 언어 사용 속에서 '저희'의 실제 용례를 파악하고자 하였다.

이 두 조건을 충족하면서도 충분한 양을 확보한 약 70여 년(1954~2024년) 간의 신문데이터를 통해 '저희' 용법의 통시적 흐름을 파악해보고자 한다. 이와 더불어 근대 소설에서 발견되는 특성에서 추가되거나 변화된 사용 양상이 있는지도 살펴볼 수 있을 것으로 기대한다. 이 과정에서는 앞서 파악한 두 가지 '저희' 용법의 특징을 기준으로 삼아, 이를 분류 모델에 적용할 학습 데이터를 코딩하는 작업에 반영할 것이다.

### Ⅲ. 1954년~2024년의 신문데이터에 나타나는 ‘저희’ 용례의 통시적 변화

지금까지 살펴본 1906년부터 1957년까지의 근대 소설에서의 ‘저희’ 용례로부터 파악한 여러 특성들을 바탕으로, 이 장에서는 보다 확장된 범위의 신문데이터에 대하여 두 가지 방법론을 적용한 통시적 분석을 시도할 것이다.

신문데이터에서 보이는 두 가지 ‘저희’의 추이 변화를 확인하기 위하여 직접 설계한 분류 모델을 통해 두 가지 저희의 사용 양상을 파악하고, N-gram 분석 방법론을 활용하여 시대별 용례들을 분석해봄으로써 추이가 나타난 원인에 대하여 해석을 시도하고자 한다.

지금부터 살펴볼 1954년부터 2024년까지 이어지는 조선 및 동아일보 신문데이터에서 발견되는 ‘저희’의 쓰임은 1900년대를 살아간, 그리고 2000년대를 살아온 이들의 삶 속에서 채록한 ‘저희’의 흔적이라고 할 수 있다. 이처럼 대량의 신문데이터를 통해 소설과 신문, 즉 각기 다른 용례가 담긴 다양한 텍스트를 살펴볼 수 있을 뿐만 아니라 무려 70년이 넘는 기간 동안에 사용되어 온 ‘저희’의 변화 추이와 특징을 통시적으로 살펴볼 수 있는 가능성이 열렸다고 할 수 있다.

#### 1. 신문데이터의 특징 및 수집방법 소개

현존하는 토큰나이저가 처리할 수 없는 옛한글을 제외하고, 1950년대부터 데이터의 범위로 설정하였다. 특히, 선행연구에서 ‘겸양의 저희’의 흔적을 개화기 신문들로부터 찾아왔다는 점<sup>13)</sup>과, 통시적 흐름을 파악하려는 연구들이 신문데이터를 활용한다는 점<sup>14)</sup>에 비추어 볼 때, 해당 기간의 신문데이터를 본 연구의 분석 자료로

---

13) 이 연구에서는 개화기 신문의 광고문으로부터 ‘저희’ 출현의 흔적과 동기를 찾고 있다(채완, 앞의 2003 논문).

14) 신문은 그 특성상 발행의 ‘정기성’, 사회의 주요 사건들을 다룬다는 측면에서 주제의 ‘시의성’을 모두 갖추고 있는 가치 있는 자료이며, 그 발행 역사 역시 100여년에 달하므로 통시적 분석 대상으로서 가치를 가지는 ‘롱 데이터’라고 할 수 있다(김일환, 「인문학을 위한 신문 빅 데이터와 텍스트 마이닝」, 『어문론집』 78, 중앙어문학회, 2019, 41~62면 참조).

활용하는 것이 적절하다고 판단하였다. 이를 통해 보다 장기적인 범위에 걸친 통시적 흐름 가운데서 ‘저희’의 두 가지 용례 추이를 확인할 수 있었다.

## 2. 두 가지 ‘저희’를 분류하기 위한 방법론: ELECTRA 분류 모델

### - 분류 모델의 도입배경 및 원리

1954년부터 2024까지의 신문데이터에서 나타나는 ‘저희’의 두 가지 용법을 분석하기 위해 ELECTRA 기반의 분류 모델<sup>15)</sup>을 사용하였다. <sup>16)</sup> ELECTRA 모델은 BERT<sup>17)</sup>와 같은 Masked Language Model(MLM)의 단점을 보완하여 고안된 모델이다.

‘저희’와 같이 복합적 용법을 가진 단어는 의미적 문맥을 고려하여 분류해야 한다. 따라서, 대규모 말뭉치로 사전 학습된 언어 모델에 복잡한 용례를 분류하여 미세조정된 모델을 활용하는 것이 적합하다. 다만, 실제 용례들을 토대로 데이터를 구축하고 학습에 반영하기 때문에, 상대적으로 데이터가 부족하거나 불균형할 것이 예상되었다. 모델 설계 단계에서 이 점을 주지하여 과적합(Overfitting)을 막고, 충분한 성능을 확보하고자 하였다. 이를 위해, ELECTRA 기반 모델을 활용하여 전이학습(Transfer Learning)을 통해 학습하는 것이 합리적이라고 판단하였다.

앞서 언급한 BERT와 같은 MLM 자연어 처리 모델의 경우 문맥을 이해하기 위해 15% 정도의 토큰을 mask로 가린 후에 mask에 어떤 단어가 들어가 있는지 추측하면서 문맥을 학습하는 구조이다. 이런 구조의 모델들은 높은 성능을 보여줬지만, 데이터 한 예시 당 15%로만 학습이 진행된다는 비효율성이 문제점으로 제시되었다. 반면, ELECTRA 모델은 이런 비효율성 문제를 해결하기 위해 Replaced Token Detection(RTD)이라는 방법론을 적용하였다. 아래 <그림 1>에 보이듯 Generator를 이용해 실제 입력된 토큰(Original) 중에 특정 부분을 가짜(Re-

---

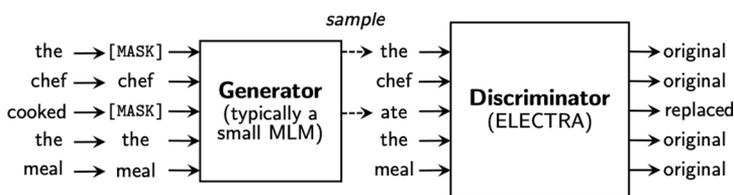
15) ELECTRA(Efficiently Learning an Encoder that Classifies Token Replacements Accurately).

16) Clark, K. et al, (2020). Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv: 2003.10555.

17) BERT(Bidirectional Encoder Representations from Transformers).

placed) 토큰으로 바꾸고, Discriminator는 각각 입력된 토큰이 실제 토큰인지 가짜 토큰인지 여부를 맞추는 작업을 통해 학습을 진행한다. MLM 모델과 달리 모든 토큰을 통해 학습을 진행하므로 더 빠르게 효율적으로 학습할 수 있고, 성능 측면에서도 더 뛰어난 것으로 확인되었다.

위와 같은 데이터와 그에 부합하는 모델의 특성을 고려하여, 확장된 신문데이터의 분류를 위해 겸양의 저희와 저편의 저희를 분류하는 심층 신경망 기반의 이진 분류 모델을 제작하였다. 사전 학습된 ELECTRA 모델에 ‘겸양의 저희’와 ‘저편의 저희’의 의미적 차이와 문맥적 차이를 정확히 구별한 데이터로 미세조정 (Fine-tuning)하여 두 가지 용법을 분류하는 모델을 학습하였다.



〈그림 1〉 ELECTRA 모델의 학습 구조

### - 분류 모델의 설계 및 학습

‘저희’ 분류 모델은 한국어 데이터로 기 학습된 KcELECTRA 모델<sup>18)</sup>을 저희 의미 분류 기능에 최적화하기 위해 데이터에 두 가지 ‘저희’를 분류 표기하는 사전 설정 작업인 파인 튜닝 (Fine-tuning) 과정을 통해 생성하였다.<sup>19)</sup> 먼저, 전체 데이터 셋을 구성하였다. 전체 데이터 셋은 수집한 신문데이터를 개별 문장 단위로 구분

18) KcELECTRA 모델은 (Junbum Lee. (2021). KcELECTRA: Korean Comments ELECTRA. 참조. 모델과 관련한 코드 자료들은 다음의 GitHub Repository를 참조. <https://github.com/Beomi/KcELECTRA>).

19) 파인 튜닝의 첫 번째 시도에서는 근대 소설 데이터와 제3의 데이터(연세 말뭉치 중 ‘저희’ 포함 문장)를 학습 데이터로 활용하여 모델을 훈련하였다. 그러나 문체의 차이, 불규칙적인 문장 부호의 사용 등으로 인해 코퍼스의 단락 구분에 어려움이 있었고, 그에 따른 토큰나이징 오류가 눈에 띄게 발생하였다. 따라서, 2차 시도에서는 신문 데이터에서 연도별 문장을 일정한 비율로 임의 추출하여 학습데이터를 구축하는 방식으로 개선하였다.

하고, ‘저희’가 포함된 문장만을 남겼다.<sup>20)</sup> 이때, 신문사, 발행 연도, 원문, ‘저희’ 포함 문장, 한 문장 안에 포함된 ‘저희’의 개수 등을 메타데이터로 표기하여 데이터를 구성하였다.

이후 전체 데이터셋에서 각 연도별로 30개의 데이터를 임의로 추출하여 총 2,100개의 학습용 데이터를 별도로 구성하였다. 해당 데이터를 세 개의 파트로 분할하여, 3인의 코더가 각각 2개의 파트를 교차하여 평정하였다(코더 1: A, B; 코더 2: B, C; 코더 3: A, C). 코더 간 일치도를 높이기 위해, 본격적인 코딩에 앞서, 세 명의 코더가 동일한 데이터를 대상으로 함께 코딩을 진행하며, 분류 기준을 공유하고 논의하였다. 이후 다음 <표 2>에서와같은 방식으로 3인이 각각 코딩을 진행하였으며, 분류가 일치하지 않는 경우에는 논의를 통해 합의된 기준으로 코딩을 완료하였다. 코더 간 일치도를 평가한 Krippendorff’s alpha 값은 0.741로, 이는 수용 가능한 수준의 일치도<sup>21)</sup>를 나타냈다. 데이터의 신뢰도를 더욱 높이기 위해, 1차 학습 후 불일치로 나타난 70개의 항목에 대해 코더들이 다시 코딩을 진행하며 추가적인 정교화 작업을 실시하였다.

<표 2> 코더가 분류한 학습 데이터 예시

신문사	연도	학습용 문장	분류
donga	1996	그 때 저희들끼리 수군거리던 비런치 일행이 고문간에게 다가왔다.	저편
donga	1996	저희부터 앞장서서 “인터넷 유스캠프”(IYC)운동을 벌여 나갈거예요	검양
chosun	1998	악독한 놈들, 저희도 사람인데 어떻게 이럴 수가 있나 ... “화전민으로 살림이 어려우나 단란한 가정”	저편
donga	2018	다만, 해외업체와 가격 경쟁을 하기엔 저희가 유리한 입장은 아닙니다.	검양
∴	∴	∴	∴

20) 본 논문에서 활용한 전체 코드 및 데이터는 다음 GitHub Repository ([https://github.com/J-Jaehyun-SEO/Project\\_Jeohui](https://github.com/J-Jaehyun-SEO/Project_Jeohui))를 참조하라.

21) Krippendorff’s alpha 값은 통상적으로 0.67 미만을 낮은 일치도, 0.67~0.80을 수용 가능한 일치도로, 0.80 이상을 높은 일치도로 본다.

이후, 세 명의 코더가 평정한 2100개의 데이터 중 약 90%에 해당하는 1,900개를 학습용 데이터, 약 10%인 210개를 테스트용 데이터로 분류하고, 학습용 데이터를 기준으로 모델을 학습하였다. 모델 구축 작업은 Python 언어와 딥러닝 프레임워크 PyTorch<sup>22)</sup> 그리고 Hugging Face<sup>23)</sup>를 사용하였다. 학습은 <표 3>의 파라미터를 기준으로 진행하였다.

### - 분류 결과 및 분석

본 연구의 테스트 데이터를 통한 검증 결과는, ‘정확도(accuracy)<sup>24)</sup> 0.9653, 정밀도(precision)<sup>25)</sup> 0.9207, 와 재현율(recall) 0.9721<sup>26)</sup>로 모두 유의한 수치를 나타내며 높은 수준의 분류 성능을 보였다.<sup>27)</sup> f1-score 또한 0.9457로 매우 우수한 결과를 보였다.<sup>28)</sup> 이는 모델이 정밀도와 재현율 간의 균형을 성공적으로

---

22) Pytorch는 Python 언어로 활용할 수 있는 심층 신경망 구동 프레임워크로 자연어 처리, 컴퓨터 비전 분야에서 활발하게 사용되고 있다(<https://pytorch.org/>).

23) Hugging Face는 트랜스포머 계열의 기 학습된 모델들을 손쉽게 불러오고 사용할 수 있도록 개발된 모듈로 모델을 처음부터 학습하는 과정에서 발생하는 시간과 비용을 절약할 수 있다(<https://huggingface.co/>).

24) 정확도(Accuracy)는 전체 예측(TP + TN + FP +FN) 중에서 정확히 예측한 (TP +TN) 비율을 의미한다. 이에 관해서는 Appendix에 수록된 <표 6>의 혼동 행렬 참조.

25) 정밀도(Precision)는 모델이 긍정 클래스로 예측한 것(TP + FP) 중에서 실제로 긍정인(TP) 비율을 의미하며, 이 값이 높다는 말은 모델이 긍정 클래스를 예측할 때, 그 예측이 높은 신뢰성을 가진다는 것을 의미한다.

26) 재현율(recall)은 실제 긍정 클래스(TP + TN) 중에서 모델이 올바르게 예측한(TP) 비율을 나타내며, 이는 모델이 실제 긍정인 데이터를 잘 찾아내는 능력이 매우 뛰어나다는 것을 의미한다.

27) ‘정확도(accuracy)’, ‘재현율(recall)’, ‘정밀도(precision)’, F-1 score는 혼동 행렬(Confusion Matrix)의 분포를 통해 확인하는 모델의 성능 지표이다. 테스트용 데이터의 실제 정답 클래스를 모델이 얼마나 정확하게 예측했는지 분석하여 행렬 형태로 표현한다. 실제 양성/음성, 예측 값의 양성/음성 여부(위 연구에서는 ‘검양의 저희’와 ‘저편의 저희’의 의미로 사용된 여부로 이해하면 된다.)를 비교하여 아래와 같이 True Positive(TP), True Negative(TN), False Positive(FP), False Negative(FN)의 네 가지 요소로 구성한다.

28) 분류 모델을 분석할 때 정밀도(Precision)과 재현율(Recall)은 모두 매우 중요한 지표이지만, 두 지표는 서로 상충되는 관계의 지표이다. 정밀도를 높이기 위해 긍정 클래스 예측 기준을 엄격하게 설정하면, 실제로 긍정인 데이터를 놓치는 경우가 많아 재현율이 낮아질 수 있다. 반대로, 재현율을 높이기 위해 긍정 클래스로 예측하는 기준을 완화하면, 잘못된 긍정 예측이 늘어나

유지하면서도 높은 분류 성능을 발휘하고 있음을 의미한다.

〈표 3〉 분류 모델 학습 파라미터

모델	학습 epoch	batch size	Optimizer
KcELECTRA	5	16	AdamW

이렇게 확정된 모델을 활용하여 1954년부터 2024년까지의 조선·동아일보 기사에서 확보한 ‘저희’를 포함하고 있는 문장 총 32,803개<sup>29)</sup>에 대하여 모델을 통한 분류 작업을 실시하였다. 그 결과 총 32,803개의 ‘저희’ 표현 중 30,221개가 ‘겸양의 저희’로 사용되었고, 2,582개가 ‘저편의 저희’로 사용되었음을 확인하였다. 다음의 〈표 4〉를 보면 연대별 사용 추세 역시 1950년대부터 오늘날까지 ‘겸양의 저희’로 사용되는 비율이 큰 폭으로 늘어나고 있음을 확인할 수 있다.

〈표 4〉 연대별 출현에 따른 ‘빈도(비율)’ 기술통계량

연대 구분	겸양의 저희	저편의 저희	전체 저희
1950년대 [1954 - 59]	548 (52.90%)	488 (47.10%)	1036
1960년대 [1960 - 69]	1179 (70.01%)	505 (29.99%)	1684
1970년대 [1970 - 79]	1426 (74.08%)	499 (25.92%)	1925
1980년대 [1980 - 89]	2067 (81.60%)	466 (18.40%)	2533

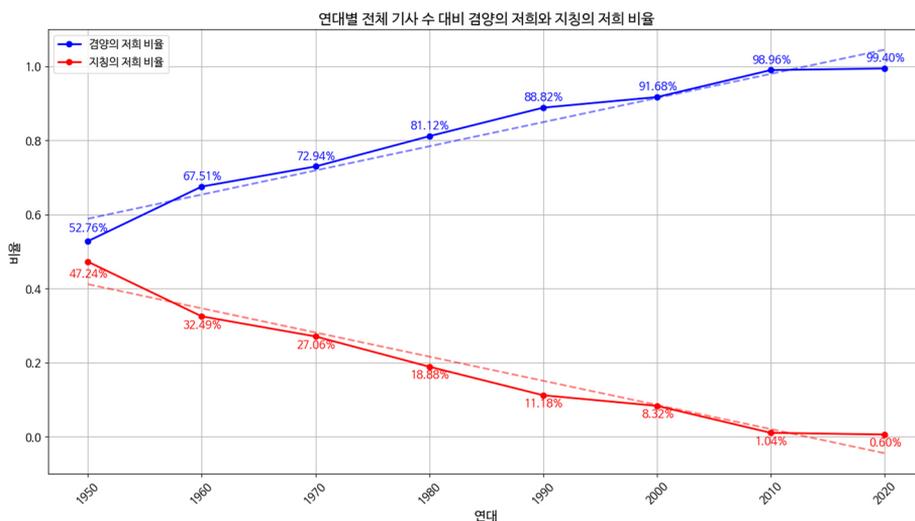
정밀도가 낮아질 수 있다. 이를 균형 있게 평가하기 위해 두 지표의 조화 평균인 F-1 score를 활용한다.

29) 1954년부터 1999년까지의 조선·동아일보 100년 아카이브로부터 ‘저희’의 단어가 포함된, 총 7,067건(동아일보 3,287건, 조선일보 3,781건)의 기사로부터 문장을 추출하였으며, 2000년부터 2024년 5월까지의 빅카인즈 DB의 조선·동아일보 기사에서 단어 ‘저희’가 포함된 16,195건(동아일보 9,010건, 조선일보 7,185건)의 기사를 합하여 ‘저희’가 포함된 문장 32,803개를 추출하였다. 문장 추출은 다음의 과정을 통해 이루어졌다(1단계: 문장 분리가 되지 않은 상태의 기사 텍스트를 kiwi의 문장 분리 기능을 활용해 문장 분리. 2단계: 이때 문장 분리는 마침표와 한국어의 종결어미 등을 종합적으로 판단해 기계학습 모형으로 판단. 3단계: 문장 분리 알고리즘을 통해 기사 당 문장 수 추출, kiwi 형태소 분석기 알고리즘과 관련하여 다음의 논문을 참조(이민철, 「Kiwi: 통계적 언어 모델과 Skip-Bigram을 이용한 한국어 형태소 분석기 구현」, 『디지털인문학』 1, 2024, 109~136면)).

멀리서 읽는 “저희”

1990년대 [1990 - 99]	2916 (89.72%)	334 (10.28%)	3250
2000년대 [2000 - 09]	1506 (91.49%)	140 (8.51%)	1646
2010년대 [2010 - 19]	8976 (99.09%)	82 (0.91%)	9058
2020년대 [2020 - 24]	11603 (99.42%)	68 (0.58%)	11671
총합	30221 (79%)	2582 (21%)	32803

\* 괄호 안의 수치는 전체 ‘저희’에 대해 각각의 ‘저희’가 차지하는 비율을 의미함.



〈그림 2〉 신문데이터에서 전체 기사 수 대비 두 가지 ‘저희’의 사용 추세 비율

위의 <표 4>와 <그림 2>에서 알 수 있듯이 ‘겸양의 저희’와 ‘저편의 저희’의 사용 빈도는 50년대까지만 해도 비슷하였으나, 10년씩 시간이 지날수록 그 차이가 큰 폭으로 벌어졌다.

특히 ‘겸양의 저희’ 사용량은 60년대부터 2020년대까지 꾸준히 증가하는 추세를 보인다. 이 추세는 점선으로 표시된 추세선을 통해서 확연히 드러난다. 60년대에 ‘저편의 저희’와 비교해 약 2.3배의 차이를 보였고, 90년대에는 이 격차가 8.7배로 확대되었다.

80년대에 들어서면서 ‘저편의 저희’는 출현 빈도가 급격히 감소해 18.4%를 기록 한 반면, ‘겸양의 저희’는 81.6%를 차지하며, 전년대인 70년대와 비교했을 때 약

10%의 증가율을 보였다. 이와 비슷한 추세는 2000년대 이후에도 이어지며, 2010년대에는 ‘겸양의 저희’가 99.09%, 2020년대에는 99.42%까지 증가했다. 반면, ‘저편의 저희’는 2010년대에 0.91%, 2020년대에는 0.58%로 그 비율이 매우 낮아졌다.

2000년대에는 ‘겸양의 저희’가 91.49%로 정점에 달하며, ‘저편의 저희’는 8.51%로 감소하였다. 특히, 2010년대와 2020년대에는 ‘겸양의 저희’의 사용이 압도적으로 많아져 ‘저편의 저희’는 거의 사용되지 않음을 보여준다. 이처럼, 60년대부터 2020년대까지 ‘겸양의 저희’는 꾸준한 상승세를 보인 반면, ‘저편의 저희’는 지속적으로 감소하여 20년대에 이르러서는 완전히 사라지지는 않은 채 드물게나마 존재하는 양상으로 변화하였다.

전체적으로 ‘저희’의 사용이 증가하는 가운데 이와 같이 달라지는 두 가지 저희의 출현 양상을 보면, 오늘날 ‘겸양의 저희’가 사람들의 인식 속에서 지배적인 위치를 점유하게 된 것은 자연스러운 일이었다고 할 수 있다.

### 3. ‘저희’에 연이어 오는 단어를 확인하기 위한 방법론: N-gram 분석<sup>30)</sup>

시간의 흐름에 따라 ‘겸양의 저희’와 ‘저편의 저희’의 사용량 추이는 뚜렷한 변화를 보였다. 시기별로 두 용법의 사용 빈도에 큰 변화가 나타나면서, 사용 양상에도 변동이 있었을 가능성이 제기된다. 이러한 변화를 체계적으로 살펴보기 위해, 본 연구에서는 엔그램(N-gram)분석<sup>31)</sup>을 활용하여 두 가지 용법의 ‘저희’에 연접하는 단어, 즉 연어(collocation)를 분석하였다. ‘저희’는 그 자체로 다의성을 지닌 표현이기 때문에 ‘저희’를 보는 것만으로는 정확한 용례를 확인하는 데 한계가 있다. 이에 ‘저희’에 곧바로 이어지는 명사를 파악함으로써, 어떠한 맥락에서 ‘저희’가 사

30) 본 연구에서는 ‘저희’가 한정하는 단어의 시대별 변화상을 파악하고자 하였다. ‘저희’와 그 뒤에 곧바로 연접하는 단어(연어, 連語)의 조합을 확인하기 위하여 N-gram 중 두 단어의 연어 관계를 살피는 Bi-gram 방법론을 활용하였다(서재현 외, 2021의 354, 372~376면 참조). 그러나 심사자들의 조언을 바탕으로 후속연구에서는 곧바로 이어지는 말 외에도 문장 단위에서 공출현하는 Co-occurrence를 살피는 것도 고려해보고자 한다.

31) 서재현 외, 앞의 2021 논문, 357면 참조.

용되었는지 구체적으로 확인할 수 있다. 이를 위해 bi-gram 분석을 활용하여 ‘저희’와 그 다음 단어 간의 연결 관계를 분석하였으며, 해당 결과를 연도별 메타데이터와 연동해 시각화하였다.

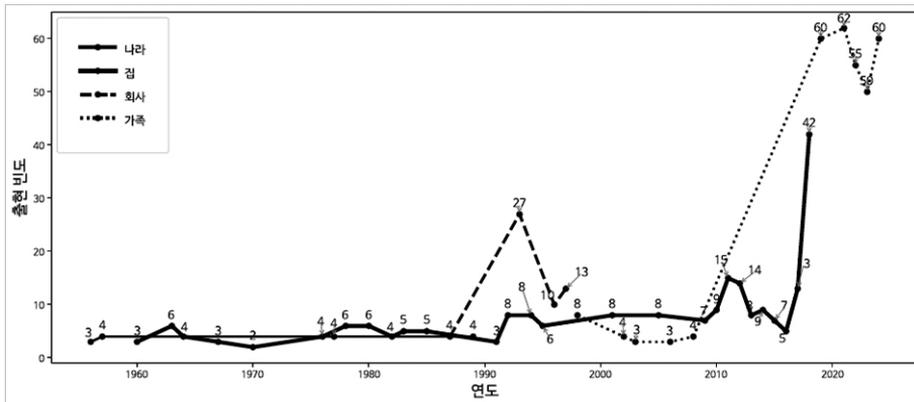
이 장에서 ‘저희’의 bi-gram 분석 결과는 다음과 같은 순서로 제시하고자 한다. 먼저, 1954년부터 2024년까지 연도별 분석 결과를 살펴본 뒤, 보다 거시적 관점에서 시대별 양상 변화를 파악하기 위해 10년 단위로 나눈 연대별 분석 결과를 제시할 것이다. 구체적으로는, 먼저 ‘점양의 저희’와 ‘저편의 저희’를 모두 포함한 ‘저희’ 용법의 연어 사용 추이를 보여준 후, 두 가지 용법 각각의 분석 결과를 제시할 것이다. 전체적인 ‘저희’ 사용의 추세를 파악한 뒤, 각 용법의 특징을 비교함으로써 더 명확하게 이해할 수 있을 것이라 기대한다.

#### - ‘저희’ 용법의 연도별 분석 결과

연도별 분석에서는 ‘점양의 저희’와 ‘저편의 저희’를 모두 포함한 전체 ‘저희’ 데이터와 각 용법의 데이터 모두 최다 빈도 중 1순위 연어를 추출하여 시각화하였다. 우선, 전체 ‘저희’ 용법의 연도별 분석 결과를 살펴보았다. 각 연도별로 ‘저희’의 연어로 가장 많이 출현한 단어를 분석한 결과, 가장 자주 등장한 단어는 ‘집’(27회), ‘가족’(10회), ‘나라’(4회), ‘회사’(4회)였다. 전반적인 양상을 살펴보았을 때, 1990년대 이전까지는 ‘저희’의 연어로 최다 출현 단어의 빈도가 1에서 6에 그쳐, 특정 단어가 두드러지게 많이 사용된 사례를 발견하기는 어려웠다. 이 시기의 등장한 단어는 ‘조합’, ‘선생님’, ‘본토’, ‘남매’, ‘당나귀’ 등으로 다양했다.

비록 빈도는 낮지만, 1990년대 이전에 눈에 띄는 단어로는 ‘나라’가 있었으며, 총 4개의 연도에서 최다 빈도 단어로 등장한 ‘나라’는 모두 1990년대 이전에 나타났다. 다음으로 주목할만한 단어는 ‘회사’인데, 1980년대 후반에 최다 빈도 단어로 등장하였고, 1993년에는 총 27차례 사용된 것으로 나타났다. 1990년 후반부터는 ‘가족’이 ‘저희’의 연어로 자주 등장하기 시작했으며, 특히 2019년부터는 그 사용 빈도가 급증하여 최근까지도 최다 빈도 단어로 유지되고 있다. 마지막으로, ‘집’은 1960년대부터 2018년까지 꾸준히 ‘저희’와 함께 사용된 단어로 분석되었다. 비록 출현 빈도는 높지 않았으나, ‘두’, ‘부부’, ‘학교’ 등의 단어가 두 개의 연도에서 최다 빈도 단어로 나타났다. 그래프<sup>32)</sup>에는 4개 이상의 연도에서 출현한 단어만을 표시하

여, 전체적인 추세를 파악할 수 있도록 하였다. 그래프에 표시된 숫자는 각 연도별 해당 단어의 출현 빈도를 나타낸다.



〈그림 3〉 연도별 '저희'의 추세

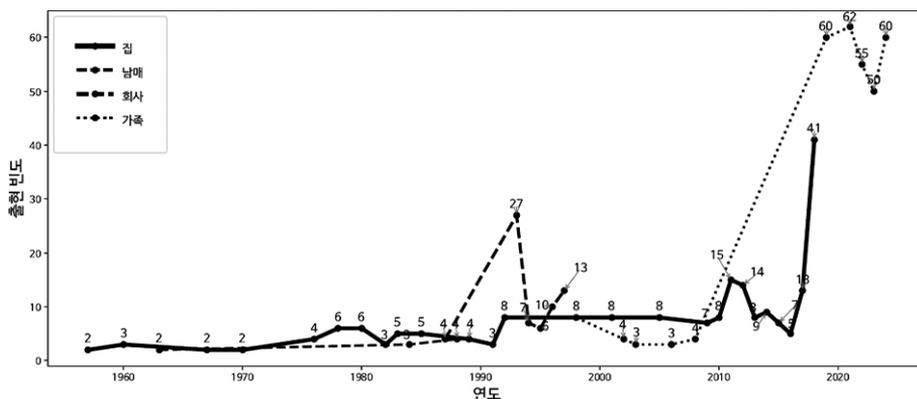
다음으로는 본 연구에서 초점을 맞춘 두 가지 '저희'의 용법 중 하나인 '겸양의 저희'의 연도별 분석 결과를 살펴보았다. 분석 결과, 전체 '저희' 데이터의 그래프와 유사한 추세를 보였다. '집(25회)', '가족(10회)', '회사(6회)', '남매(3회)' 등의 단어가 자주 등장하였으며, 이는 전체 데이터에서 빈번하게 나타난 단어들과 유사한 양상을 띠었다. 여러 차례 등장하지는 않았지만, 상대적으로 높은 빈도를 보인 단어로는 2020년에 53회 등장한 '당', 1990년에 10회 나타난 '내외', 1999년 10차례 기록된 '부부' 등이 있었다. 이 외의 단어들은 모두 5회 이하의 빈도수를 보였다.

자주 등장한 단어들을 자세히 살펴보면, 전체 데이터 분석에서는 등장하지 않았던 '남매'가 '겸양의 저희' 분석33)에서 등장하였으며, 1963년, 1984년, 1988년에

32) 연구에 사용할 데이터 수집이 완료된 시점이 2024년 5월이라는 한계로 인하여 2024년의 경우, 상반기인 1월부터 5월까지의 데이터만 수집할 수 있었다. 2024년 데이터를 전체연도와 비교 가능한 형태로 만들고, 그래프에서 추세에 대한 오해를 방지하기 위해 2024년 데이터에는 가중치 ( $12/5=2.4$ )를 적용하여 데이터를 보정한 후 제시하였다.

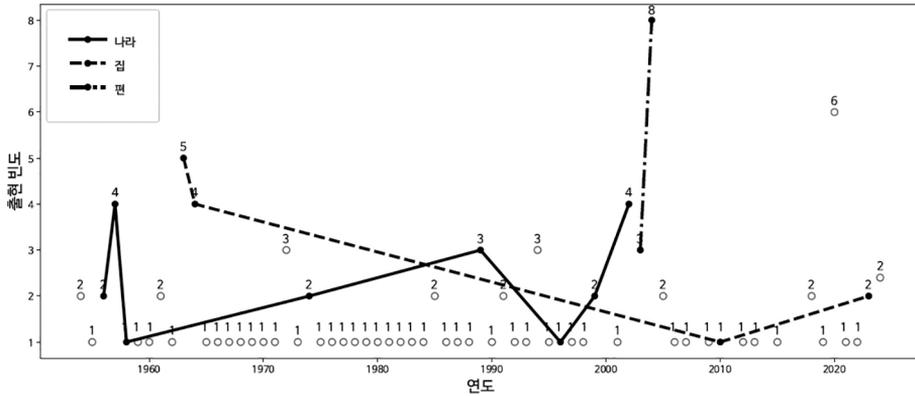
33) '겸양의 저희' 분석에서도 전체 데이터와 마찬가지로 통시적 추세 파악에 용이하도록 3개 연도 이상에서 나타난 단어들만을 그래프에 표시하였으며, 그래프의 각 숫자는 연도별 해당 단어의 출현 빈도를 나타낸다.

나타나 모두 1990년 전에 사용된 것으로 확인되었다. ‘집’, ‘가족’, ‘회사’는 전체 데이터 그래프와 유사한 패턴을 보였다. ‘회사’는 1980년대 후반에 최다 빈도 단어로 나타나기 시작하였고, ‘가족’은 1990년대 후반에 최다 빈도 단어로 등장하였다. ‘집’은 1957년에 등장해 2018년까지 지속적으로 ‘저희’와 함께 사용된 단어로 분석되었다. 또한, ‘두’, ‘부부’, ‘학교’는 각각 두 개의 연도에서 최다 빈도 단어로 나타났다.



〈그림 4〉 연도별 '겸양의 저희'의 추세

연도별로 ‘저편의 저희’ 용법을 분석한 결과, ‘저편의 저희’는 전체 ‘저희’ 데이터와는 상이한 양상을 보였다. ‘저편의 저희’는 대체로 낮은 빈도수를 보였으며, ‘겸양의 저희’처럼 자주 사용되는 연어가 있는 것이 아니라 다양한 단어들 산발적으로 사용되는 경향을 보였다. 여러 차례 나타난 단어로는 ‘나라(6회)’, ‘집(3회)’, ‘편(2회)’이 있었다. 여러 해에 걸쳐 반복적으로 등장하지는 않았으나, 일부 연도에서 상대적으로 높은 빈도를 보인 단어도 있었다. 2020년에 6회 등장한 ‘대신’, 1972년에 3회 나타난 ‘본국’, 1994년에 3회 출현한 ‘집안’이 그 예이다. 분석 결과는 〈그림 5〉에 나타냈으며, 그래프에서 ‘나라’, ‘집’, ‘편’과 같이 여러 차례 나타난 단어들은 선으로 연결되었고, 그 외 단어들은 빈 점(○)으로 표현되어, 데이터의 불균일하고 산발적인 분포를 쉽게 파악할 수 있다.



〈그림 5〉 연도별 ‘저희’의 추세

- ‘저희’ 용법의 연대별 분석 결과

연도별 분석에서는 다소 산발적으로 보였던 데이터들이 연대별로 취합되었을 때 의미 있는 사용 패턴이 발견되는지 확인하고자 하였다. 연대별 분석은 연도별 분석보다 전반적인 시대적 흐름을 파악하기 용이하여, 통시적 관찰을 가능하게 한다. 1950년대부터 2020년대까지 총 8개의 연대로 구분하여 bi-gram 분석을 진행하였다. 연대별 분석에서는 빈도가 가장 높은 연어 상위 열 개를 추출하여 살펴보았다. 여러 연도의 데이터를 통합했기 때문에, 연도별 분석처럼 하나의 단어만을 분석하는 것은 정보 손실이 크다고 판단되어, 열 개의 연어를 선택하여 분석하였다. 연대별 데이터를 통합하였음에도 불구하고 각 용법에서 빈도가 1인 단어는 의미가 크지 않다고 판단하여 분석 결과에 제시하지 않았다. 연대별 전체 분석 결과는 〈표 5〉에 제시하였다.

〈표 5〉 ‘저희’ 용법 연대별 N-gram 분석 결과

연대	전체	겹양의 저희	저편의 저희
1950년대	나라(9)	집(5)	나라(8)
	집(6)	두(4)	것(3)
	두/사람(4)	사람(3)	동무(3)
	것/동무(3)	원기소(2)	-
	군사/원기소(2)	-	-

멀리서 읽는 “저희”

1960 년대	집(19)	집(8)	집(11)
	남매/학교/형제(4)	학교(4)	것(3)
	것/나라/두/부부/조합(3)	남매/두/조합/형제(3)	나라/다랑이/동무/배(2)
	다랑이/동무/배/사람/선생님/ 아버지/애기/한국신문/후궁(2)	부부/사람/한국 신문 발행인 협회/후궁(2)	-
1970 년대	집(17)	집(16)	나라(4)
	나라(10)	신문(8)	것/본국(3)
	신문/학교(8)	나라/동네/아버님/학교(6)	무리/배/부하/꽤거리/학교(2)
	동네/들/아버님(6)	들/회사(5)	-
	당나귀/회사(5)	당나귀/본토/아버지/아빠/ 이리시/회원(4)	-
	두/본토/아버지/아빠/이리시/ 회원(4)	-	-
1980 년대	집(34)	집(33)	나라(6)
	집안(17)	회사(16)	재물/집안(3)
	회사(16)	집안(14)	것/동료/또래/속국/ 아버지/아빠/형제(2)
	학교(13)	학교(13)	-
	아버지(12)	아버지(10)	-
	나라(10)	어머니(8)	-
	어머니(8)	남매/선생님/학생(7)	-
	남매/선생님/학생/형제(7)	부부/아버님(6)	-
1990 년대	회사(83)	회사(82)	나라(6)
	집(62)	집(58)	집/집안(4)
	부부(29)	부부(29)	가슴/것(2)
	가족(23)	가족(23)	-
	학교(22)	학교(21)	-
	내외(17)	내외(17)	-
	세대(16)	세대(16)	-
	나라(13)	대학/들/아버님(11)	-
대학/들/아버님/집안(11)	-	-	

2000 년대	집(37)	집(36)	편(15)
	부부(22)	부부(22)	나라/장수/편이(4)
	회사(21)	회사(21)	군사/임금/진채(3)
	가족(20)	가족(20)	대군/대장/왕/주인(2)
	편(15)	학교(13)	-
	학교(14)	남편/시어머니/음악(7)	-
	남편/시어머니/음악(7)	동아일보/어머니/엄마/은행/자식(6)	-
	동아일보/아이/어머니/엄마/은행/자식(6)	-	-
2010 년대	집(155)	집(153)	집/부부(2)
	가족(107)	가족(107)	-
	부부(89)	부부(87)	-
	회사(62)	회사(62)	-
	아이(53)	아이(53)	-
	당(51)	당(51)	-
	어머니(41)	어머니(41)	-
	부모님(34)	부모님(34)	-
	엄마(29)	엄마(29)	-
딸/아버지/제품(26)	딸/아버지/제품(26)	-	
2020 년대	가족(239)	가족(239)	대신(6)
	집(175)	당/집(171)	집(4)
	당(171)	부부(129)	젓(3)
	부부(129)	아버지(104)	-
	아버지(104)	아이(102)	-
	아이(102)	회사(93)	-
	회사(93)	어머니(77)	-
	어머니(77)	입장(67)	-
	입장(67)	팀(66)	-

전체 데이터를 먼저 살펴보면, 각 연대별 순위는 다음과 같다. 1950년대에는 ‘저희’와 가장 많이 연관된 단어로 나라(9회), 집(6회), 두/사람(4회), 첫/동무(3회), 군사/원기소(2회)가 나타났으며, 이 시기에는 국가나 가족적인 맥락에서 ‘저희’가 주로 사용된 것으로 보인다. 1960년대에는 집(19회), 남매(4회), 학교(4회), 형제(4회), 첫/나라/두/부부/조합(3회)과 같은 단어들이 주로 등장해 가족 및 교육과 관련된 맥락에서 많이 사용되었다. 1970년대에는 집(17회), 나라(10회), 신문(8회), 학교(8회), 동네/들/아버님(6회)이 나타나, 여전히 가족, 교육, 사회적 맥락에서 ‘저희’가 사용되었으며, ‘신문’이라는 단어가 새롭게 등장하여 자주 사용되었다. 1980년대에는 집(34회), 집안(17회), 회사(16회), 학교(13회), 아버지(12회), 나라(10회), 어머니(8회)가 주로 등장하였고, 이 시기부터는 ‘저희 회사’와 같이 직장과 관련된 표현이 나타났다.

1990년대에는 회사(83회), 집(62회), 부부(29회), 가족(23회), 학교(22회), 내외(17회), 세대(16회), 나라(13회)가 주로 사용되었으며, 특히 ‘저희 회사’라는 표현이 급증하면서 직장 관련 맥락에서의 사용이 두드러졌다. 2000년대에는 집(37회), 부부(22회), 회사(21회), 가족(20회), 편(15회), 학교(14회)가 등장해, 가족과 직장에 대한 표현이 여전히 많았고 ‘편’이라는 새로운 연어가 등장했다. 2010년대에는 집(155회), 가족(107회), 부부(89회), 회사(62회), 아이(53회), 당(51회), 어머니(41회) 등이 자주 나타났으며, 특히 ‘저희 집’과 ‘저희 가족’이 급증하였고 ‘아이’와 관련된 표현도 자주 사용되었다. 2020년대에는 가족(239회), 집(175회), 당(171회), 부부(129회), 아버지(104회), 아이(102회)가 가장 많이 등장하여, ‘저희 가족’과 ‘저희 집’의 빈도가 지속적으로 높았고, 정치적 맥락에서 ‘당’과 관련된 표현도 크게 증가하였다.

다음으로 ‘겸양의 저희’ 용법에 대한 연대별 bi-gram 분석 결과를 살펴보았다. 연도별 분석 결과와 유사하게, 전체 데이터 분석에서 나타난 단어들이 비슷한 순위로 나타났다. 특히 ‘겸양의 저희’ 용법에서는 각 연대별로 반복적으로 사용된 단어들 두드러졌는데, 그 중에서도 집은 1950년대, 1960년대, 1970년대, 1980년대, 2000년대, 2010년대 모두에서 1순위를 기록하며 ‘겸양의 저희’와 가장 밀접한 연관성을 보였다.

1950년대와 1960년대는 다른 연대와 비교했을 때 상대적으로 꾸준히 사용된

단어가 뚜렷하게 드러나지 않았다. 1950년대에는 집(5회), 두(4회), 사람(3회)가 주로 사용되었고, 1960년대에는 집(8회), 학교(4회), 남매/두/조합/형제(3회) 등이 나타났다. 이 시기에는 개인적 공간이나 가족 중심의 단어들 이 주로 쓰였으며, 사회적 영역을 반영한 단어들은 드물었다. 1970년대에는 이전 연대와 비교해 다양한 단어들이 ‘겸양의 저희’와 함께 쓰이기 시작하였다. 집(16회), 신문(8회), 나라/동네/아버님/학교(6회), 들/회사(5회), 당나귀/본토/아버지/아빠/이리시/회원(4회)이 등장하며, 특히 ‘신문’과 ‘회사’와 같은 사회적 단어들이 눈에 띄기 시작했다.

1980년대부터는 ‘겸양의 저희’와 함께 사용된 연어들이 비교적 안정적인 패턴을 보였다. 이 시기에는 집(33회), 회사(16회), 집안(14회), 학교(13회), 아버지(10회), 어머니(8회) 등의 단어들이 높은 빈도를 기록했다. 이 시기에는 가족과 관련된 단어들이 중심을 이루면서도, 사회적 관계를 나타내는 단어들의 빈도가 꾸준히 증가하였다. 1990년대에도 비슷한 단어들이 나타났지만, 특히 회사(82회)와 부부(29회)의 급증이 두드러졌다. 집(58회), 가족(23회), 학교(21회) 등도 주요 단어로 계속해서 등장했다. 이는 1990년대의 경제적, 사회적 변화와 연관이 있을 가능성이 있을 것으로 추론된다. 2000년대에도 집(36회), 부부(22회), 회사(21회), 가족(20회), 학교(13회) 등이 꾸준히 사용되며 비슷한 패턴을 유지했다.

2010년대에는 집(153회), 가족(107회), 부부(87회), 회사(62회), 아이(53회), 당(51회)이 주요 단어로 나타났으며, 특히 가족과 부부의 빈도수가 이전보다 급격히 증가했다. ‘아이’와 ‘당’이 새롭게 등장한 것이 특징적이다. 2020년대에는 처음으로 가족(239회)이 가장 많이 쓰인 단어로 나타났으며, 그 뒤로는 당/집(171회), 부부(129회), 아버지(104회), 아이(102회), 회사(93회), 어머니(77회)가 많이 나타났다. 이와 더불어, 입장(67회), 팀(66회)와 같은 집단 간 구분을 나타내는 단어들이 새롭게 등장해 사회적, 집단적 관계를 강조하는 경향을 보였다.

연대별 bi-gram 분석 결과, ‘겸양의 저희’는 시대의 흐름에 따라 그 사용 패턴이 변화하면서도 일정한 연어들이 꾸준히 사용되었음을 알 수 있다. 특히 ‘집’, ‘회사’, ‘가족’, ‘부부’와 같은 단어들이 지속적으로 높은 빈도를 기록하며, 겸양의 표현이 가족과 사회적 관계를 중심으로 형성되어 왔음을 알 수 있다. 또한 2000년대 이후 ‘아이’, ‘당’, ‘입장’, ‘팀’과 같은 새로운 연어들이 등장하여, 사회적 맥락의 변화와 함께 ‘겸양의 저희’의 쓰임 또한 변화하고 있음을 보여준다.

‘저편의 저희’ 용법에서는 연대별 bi-gram 분석 결과 또한 다양한 단어들 이 산발적으로 나타났다. 그럼에도 불구하고, ‘나라’는 1950년대부터 2000년대까지 꾸준히 상위 순위에 자리하며 ‘저편의 저희’와 함께 자주 사용된 단어로 확인되었다.

1950년대에는 나라(8회), 것(3회), 동무(3회)가 주요 단어로 나타났고, 1960년대에는 집(11회), 것(3회), 나라/다랑이/동무/배(2회)가 상위 순위를 기록했다. 1970년대에도 유사하게 나라(4회), 것/본국(3회), 무리/배/부하/패거리/학교(2회) 등 다양한 단어들 이 나타났다. 1980년대에는 나라(6회), 재물/집안(3회), 것/동료/또래/속국/아버지/아빠/형제(2회) 등이 사용되었고, 1990년대에는 나라(6회), 집/집안(4회), 가슴/것(2회) 등이 나타났다.

2000년대에는 편(15회)이 가장 두드러지게 사용되었으며, 그 외 나라/장수/편이(4회), 군사/임금/진채(3회), 대군/대장/왕/주인(2회) 등이 나타났다. 2010년대에는 집/부부(2회)만이 상위에 나타나 ‘저편의 저희’ 용법이 매우 산발적으로 사용되었음을 알 수 있었다. 2020년대에는 대신(6회), 집(4회), 것(3회)이 주요 단어로 나타났다. ‘겸양의 저희’가 사회적 규범과 관련된 맥락에서 일관되게 사용된 것에 비해, ‘저편의 저희’는 상대적으로 특정 상황이나 대상을 강조하는 용법으로 사용된 것으로 보인다.

2장에서 살펴보았듯, ‘겸양의 저희’는 ‘저편의 저희’에 비해 시간이 흐를수록 사용 빈도가 크게 증가하였다. N-gram 분석 결과에 따르면, ‘저편의 저희’의 용례는 산발적으로 나타나는 반면, ‘겸양의 저희’는 시간이 지남에 따라 특정 용례에 집중되며 빈도가 높아졌다. 이러한 추세는 각 용법의 쓰임에도 영향을 미쳤을 것으로 추정되는데, ‘겸양의 저희’는 ‘집’과 같은 단어가 1950년대부터 2020년대까지 꾸준히 사용되었으며, 1980년대 이후로는 ‘집’, ‘회사’, ‘학교’, ‘어머니’, ‘부부’, ‘가족’ 등의 단어가 지속적으로 쓰였다.

반면, ‘저편의 저희’의 경우 ‘나라’를 제외하고는 꾸준히 사용되는 용례를 찾아보기가 어려웠다. 특히 2010년대에 이르러서는 ‘집’과 ‘부부’가 각각 2회 등장한 것이 최다 출현 빈도로 기록되었을 뿐, 그 외에 1회 이상 반복 사용된 단어는 발견되지 않았다. 이는 ‘저편의 저희’가 전반적으로 정해진 쓰임이 없이 산발적으로 사용되다가 점차 그 빈도가 감소하고 있는 것으로 추론 가능하다.

이는 ‘겸양의 저희’가 ‘저편의 저희’에 비해 비교적 안정적인 용례를 가지고 있으며, 그 결과로 그 쓰임이 계속 증가하고 있는 것으로 해석된다.

#### IV. 결론

지금까지 근대 소설과 신문데이터를 통해 ‘저희’의 두 가지 용법인 ‘겸양의 저희’와 ‘저편의 저희’에 대한 통시적 추이와 그 특징을 살펴보았다. ‘겸양의 저희’는 ‘우리’의 겸양어로 쓰이는 ‘저희’를 나타내고, ‘저편의 저희’는 ‘우리 편’과 대비되는 ‘저희’를 말한다. 먼저, 약 50여 년에 걸친 근대 소설 데이터로부터 ‘저희’가 두 가지 방식으로 사용된 것을 확인할 수 있었다. 데이터를 시기별로 나누어 분석한 결과, 1990년대 중반까지는 ‘저편의 저희’가 작지 않은 비중으로 출현하고 있음을 알 수 있었다. 이 표현은 화자의 집단(우리 편)과 대비되는 집단을 지칭하는 데 사용되었는데, 이후 점차 시간이 흐르며 청자에 대하여 자신의 집단을 겸손하게 낮추는 ‘겸양의 저희’에게 자리를 내어주고 있음을 알 수 있었다.

근대 소설을 창작한 작가들은 작중에서 화자를 낮추고 청자를 배제하는 동시에, 여러 집단 가운데 화자가 속한 영역을 특정하기 위해 ‘저희’를 사용하고 있었다. 한편, 더 큰 집단인 ‘민족’과 같은 범주를 상정하여 모두를 포섭하고자 할 때는 ‘우리’를 사용하며 그 범위를 상황에 따라 유연하게 조정하고 있었다. 또한, 청자에게 존경하지 않는다는 태도를 의도적으로 드러내기 위해 ‘저희’ 대신 ‘우리’를 사용하는 경우도 관찰되었다. 이러한 ‘저희/우리’의 사용 방식은 특히 몇몇 소설가들의 작품에서 두드러졌으며, 여기에는 작가들의 문체가 상당히 반영되어 있었다. 근대 소설에서 나타나는 두 가지 ‘저희’의 특징을 도출하여 3장에서 사용할 학습 및 파인 튜닝 과정의 토대자료로 삼는 동시에, 데이터의 종류와 범위를 확장함으로써 보다 일반적이고 통시적인 해석의 가능성을 열어보고자 하였다.

이를 위해 근대 소설 데이터에 이어, 분석 대상을 오늘날까지의 신문데이터로 확장하여 연구를 진행하였다. 데이터를 분류하고 세부 용례를 분석하기 위해 ELECTRA를 기반으로 한 ‘의미 분류 모델’을 도입하였다. 이 분류 모델을 활용함으로써 방대한 데이터의 용례를 효과적으로 분류하고 통계적으로 분석하는 효율을

크게 향상시킬 수 있었다. 1954년에서부터 2024년에 이르는 신문데이터에 대하여 이 분류 모델을 적용한 결과, 겸양의 저희는 저편의 저희를 큰 폭으로 압도하는 한편, 저편의 저희는 소멸하지 않은 채 낮은 빈도로 사용을 이어오고 있음이 확인되었다. 이는 오늘날 ‘저희’가 겸양어로 당연하게 받아들여지는 현상을 반영하는 것이다.

이어서 ‘저희’와 함께 자주 사용되는 용례를 확인하기 위해 N-gram 분석을 시행하였다. 신문데이터에 대하여 두 가지 ‘저희’ 용법을 나눠 bi-gram 분석을 진행한 결과, 먼저 연도별로 살펴본 ‘겸양의 저희’는 전체 ‘저희’ 데이터와 유사한 추세를 보였다. 즉, 전체 데이터에서 자주 나타난 단어들이 ‘겸양의 저희’의 연어(連語)로도 다수 등장하였다. 구체적으로는 ‘집’, ‘가족’, ‘회사’, ‘남매’ 등의 단어가 빈번하게 사용되었다. 반면, ‘저편의 저희’는 전체 ‘저희’ 데이터와는 상이한 양상을 보였다. ‘저편의 저희’는 전반적으로 낮은 빈도수를 기록하였으며, 산발적으로 사용된 연어들이 발견되었다. 주요 연어로는 ‘나라’, ‘집’, ‘편’ 등이 있었다.

연대별로 분석한 결과도 연도별 분석과 유사하게 나타났다. ‘겸양의 저희’의 연어와 전체 데이터에서 자주 나타난 단어들이 비슷한 순위로 유지되었으며, 특히 ‘집’은 여덟 개의 연대 중 여섯 개의 연대에서 가장 많이 사용된 단어로 기록되며 ‘겸양의 저희’와 가장 긴밀하게 연결되었다. ‘저편의 저희’의 연어는 연대별 분석에서도 산발적인 경향을 보였으나, 그 중 ‘나라’는 1950년대부터 2000년대까지 꾸준히 ‘저편의 저희’와 함께 사용된 것으로 나타났다. 결과적으로, ‘겸양의 저희’는 시간이 흐를수록 유사한 단어들이 꾸준히 사용되며 안정적인 쓰임을 보인 반면, ‘저편의 저희’는 특정 상황이나 대상을 강조하기 위한 용법으로 산발적으로 사용된 것으로 보인다. 이러한 안정성과 산발성의 차이는 시대를 거치며 그 사용 빈도와 비중이 크게 변동했지만, 근대 소설 분석을 통해 추측한바—‘겸양의 저희’는 특정 작품에서 집중적으로 사용된 반면, ‘저편의 저희’는 비교적 다양한 맥락에서 사용되었을 가능성—를 구체적인 데이터로 확인하는 작업이었다고 할 수 있다.

두 가지 방법론을 통해 근대 소설 텍스트 전반에서 유사한 비율로 등장했던 두 가지 ‘저희’가 지난 70여 년간 다양한 용례로 분화되고, 변용되는 과정을 거쳐왔음을 알 수 있었다. 이 과정에서 ‘겸양의 저희’는 가족과 사회 집단을 나타내는 용례를 중심으로 안정화되고, ‘저편의 저희’는 ‘나라’를 제외하고는 여러 갈래로 산재해 오

며 두 갈래로 선명히 갈리는 추이를 그린 것이다. 이는 앞서 선행연구에서 추정하였던 근대화과 같은 시대변화에 따라 증가한 ‘겸양의 필요’에 의해 ‘우리’를 낮추기 위해 고안된 표현인 ‘저희’가 사회, 관계적 맥락을 반영한 특정 용례들과 더 많이, 더 제한(집약)적으로 결합하여 나타난 결과라고 볼 수 있다.

특히, 두 가지 자아 지칭, ‘우리’와 ‘저희’가 함께 사용된 문장을 살펴보면, ‘저희’의 특성을 보다 선명하게 읽어낼 수 있는 문장들을 발견할 수 있다. ‘겸양의 저희’는 ‘우리’의 낮춤말이라는 점에서, ‘저편의 저희’는 우리(내) 편에 대비된다는 점에서 그러한데, 화자들은 겸양의 의도, 배제의 의도, 지시의 의도를 나타내기 위한 경우에는 자신들을 ‘저희’로 표현하는 한편, 청자를 아울러 자신의 영역에 포함시키는 경우에는 ‘우리’를 사용하고 있었다. 이처럼 맥락에 따른 적절한 자아 지칭을 선택하기 위해서는 자신이 처한 상황, 화자와 청자의 관계, 발화의 뉘앙스 등을 고려하면서 ‘저희’가 가진 기능과 한정 범위를 고려해야 한다.

만일 이를 무시한 채, 단순히 사전적 의미의 ‘우리의 낮춤말’로 ‘저희’에 평면적으로 접근하게 되면 낮춰야 할 대상이 아님에도 높여야 할 사람까지도 모두 낮춰버리는 과도한 낮춤의 상황 혹은, ‘나라’를 낮춰 불필요하게 품격을 해치는 문제 상황 등이 발생하기도 한다. 실제로 우리가 들여다본 데이터 곳곳에는 이 문제 상황을 피하기 위하여 ‘저희’를 이러저러한 방식으로 사용하는 것이 적절하다는 논조가 상당수 발견되기도 하였다.

하지만 우리가 이 연구에서 ‘저희’의 적절한 사용에 대한 확정된 해결책을 제안할 수는 없다. ‘겸양의 저희’ 및 ‘저편의 저희’와 함께 짝을 이뤄 등장하는 수많은 크고 작은 관계들 사이에 통용되는 ‘하나의 단일한 법칙’을 만드는 것은 불가능에 가까워 보이기 때문이다. 그 대신, 우리는 데이터 가운데서, 화자들이 매 순간 고민 끝에 선택하여 사용해 온 ‘저희’의 궤적들, 그리고 크고 작은 범위에서 ‘우리’와 ‘저희’가 조응해 온 흔적들을 있는 그대로 보이고자 하였다. 이 통시적 흐름은 근대 소설 데이터를 가까이 들여다보고 얻은 추정을 실제 우리의 언어 사용을 비교적 잘 보존하고 있는 신문데이터로부터 멀리서 보고, 또다시 조금 더 다가가 보아 확인한 끝에 비로소 얻을 수 있는 결론이었다.

우리의 연구에는 다음과 같은 한계도 존재하였다. 지시어 ‘저희’와 가까운 낱말로 ‘저’와 ‘더’를 비롯하여 ‘저희’의 여러 이형이 존재함을 알 수 있었다. 연구 설계

## 멀리서 읽는 “저희”

차원에서 적정 규모를 고려해야 할 필요가 있었고, 지면의 한계로 인하여 ‘저편의 저희’의 여러 다른 형태로 간주되는 ‘즈이·저이·저·즈·지’ 등 ‘저희’의 여러 이형을 검증하여 데이터 추이를 파악하는 과정에 반영하지 못했다는 한계가 있다.

또한, 토큰나이징 기술의 한계로 인하여, 수집한 신문데이터를 모두 활용하지 못했다는 점 역시 아쉬움이 남는다. 형태소 분석 기술의 제약으로 인하여 당초 수집하였던 1906년부터의 데이터를 사용하지 못하였다는 점 등은 본 연구가 더 개선하고 나아가야 할 한계점이다. 마지막으로, 본 연구에서는 ‘저희’를 중심으로 수집된 데이터를 들여다보는 한 가지 매개로써 ‘우리’를 제한적으로 활용하였다. 후속 연구에서 전격적으로 ‘우리’와 ‘저희’가 함께 사용되어 온 전체 통시적 흐름 등을 파악해 본다면 두 가지 ‘저희’가 그리는 추이를 넘어서 보다 풍부한 해석을 제시할 수 있을 것으로 기대한다.

투고일: 2024.10.20

심사일: 2024.11.20

게재확정일: 2024.12.20

참고문헌

- 국립국어원, '저희', 『국립국어원 표준국어대사전』, 국립국어원, n.d
- 민영진, 『성경전서 개역개정판-이렇게 달라졌다』, 대한성서공회 성경원문연구소, 2003
- 이기창, 『한국어 임베딩』, 에이콘, 2019
- Franco Moretti, *Distant Reading*, London: Verso, 2013
- HuggingFace, 『트랜스포머 모델 사용 모듈』, (<https://huggingface.co/>)
- Junbum Lee, 『Kclectra: Korean Comments ELECTRA』, 2021, (<https://github.com/Beomi/KcELECTRA>)
- PyTorch, 『사용 가능한 자연어 처리, 컴퓨터 비전용 딥러닝 프레임워크』, (<https://pytorch.org/>)
- 김미형, 「국어 대명사의 어휘사」, 『한국어 의미학』 9, 한국어의미학회, 2001
- 김일환, 「인문학을 위한 신문 빅 데이터와 텍스트 마이닝」, 『어문론집』 78, 중앙어문학회, 2019
- 김정남, 「한국어 대명사 우리의 의미와 용법」, 『한국어 의미학』 13, 한국어의미학회, 2003
- 서재현 외, 「멀리서 읽는 "우리"-Word2Vec, N-gram을 이용한 근대 소설 텍스트 분석」, 『대동문화연구』 115, 대동문화연구원, 2021
- 양영희, 「중세국어 3 인칭 대명사의 존재와 기능 검증」, 『우리말글』 33, 우리말글학회, 2005
- \_\_\_\_\_, 「중세 국어 '재귀대명사' 설정에 대한 재고」, 『한글』 273, 한글학회, 2006
- 이민철, 「Kiwi: 통계적 언어 모델과 Skip-Bigram을 이용한 한국어 형태소 분석기 구현」, 『디지털인문학』 1, 2024
- 이봉일, 「개화기 문예에 나타난 근대적 내면성의 성립 과정 연구」, 『국제어문』 42, 국제어문학회, 2008
- 전무용, 「한국어 성경의 대명사 고찰」, 『성경원문연구』 19, (재)대한성서공회 성경번역연구소, 2006
- 정길남, 「개화기 교과서의 존대법 연구」, 『동악어문학』 42, 동악어문학회, 2004
- 채완, 「개화기 광고문의 표현 기법」, 『한국어 의미학』 12, 한국어의미학회, 2003
- 홍순성, 「한국어 재귀대명사의 특질」, 『동서인문학』 19, 인문과학연구소, 1987
- Cavnar, W. B., Trenkle, J. M., N-gram-based text categorization, In *Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval* Vol. 161175, 1994
- Clark, K., et al, *Electra: Pre-training text encoders as discriminators rather than generators*, 2020 (arXiv preprint arXiv:2003.10555)

## Distant Reading on ‘Jeohui’ — A Study on the Diachronic Changes of Two Types of ‘Jeohui’ Applying an ELECTRA-Based Semantic Classification Model and N-gram Analysis

Seo, Jaehyun·Ha, Na·Kim, Min-woo·Han, Jimin·  
Kim, Byungjun·Park, So-jeong

This paper adopts quantitative research methodologies collectively referred to as “distant reading” to analyze the characteristics of two types of the term “*Jeohui*” found in modern Korean novel texts. By applying these findings to an expanded dataset up to the present day, the study examines the diachronic trends in the usage of “*Jeohui*” to understand its overall patterns of change. Utilizing digital methodologies such as an ELECTRA-based semantic classification model and N-grams, the research provides a macro-level view of the trajectories of the two types of “*Jeohui*” in the data from newspapers (Chosun and Donga), spanning approximately 1954 to 2024.

The study examines the broader trends in the usage of the term “*Jeohui*”, which have previously been discussed only sporadically and in a limited scope, to explore how its usage became established and how its meaning has evolved over time. Additionally, it conducts collocation analyses of the two types of “*Jeohui*” across different periods, offering extensive examples of vocabulary associated with each usage. Through this, the paper aims to lay the groundwork and provide direction for future research on “*Jeohui*”.

Key Words : ELECTRA, N-gram, Jeohui, Distant Reading, Modern Korean Novels,  
Newspaper Data

[부록]

〈표 1〉 혼동 행렬(Confusion Matrix)

	실제 음성('저편의 저희')	실제 양성('곁양의 저희')
예측 음성('저편의 저희')	True Negative(TN)	False Negative(FN)
예측 양성('곁양의 저희')	False Positive(FP)	True Positive(TP)