

[특집] 한글의 과학성을 문화로 바꾼 분들

최현배의 《우리말 말수 사용의 찾기 조사》*

서상규 연세대학교 국어국문학과 교수
언어정보연구원 에이치케이(HK)사업단 단장

1. 국어 정보화의 첫걸음

오늘날 국어의 정보화라든가 국어 정보라는 말은 이제 더 이상 전문적이거나 학술 분야에서만 쓰이는 낱말이 아니게 되었으며, 누구나 한 번쯤 들어 보았을 테고 그 뜻을 얼추는 알고 있는, 거의 일상적인 낱말이라 해도 지나치지 않다. 그러면 우리말의 정보화는 언제부터 시작된 것일까? 이 물음에 대한 답은, 무엇을 기준으로 하느냐에 따라서 여러 갈래가 될 것이다.

국어 자료를 다루거나 연구를 하는 데 컴퓨터나 전산화 기술을 활용하기 시작한 것을 기준으로 한다면 대체로 1980년대가 그 시작이 될 것이다. 서상규(2002)에서도 지적하듯이, 1980년대 말부터 ‘국어정보학회’, ‘한국정보학회’, ‘한글 및 한국어 정보 처리 대회’ 등 한글 관련 전산 기술에 대한 논의가 활성화되기 시작했기 때문이다.¹⁾

* 이 글은 서상규(2008)를 바탕으로 하여, 이후의 연구를 통해서 새롭게 알게 된 것들을 덧붙이며 일반 독자를 위하여 고쳐 쓴 것이다.

1) 이러한 연구의 흐름에 대해서 더 소상히 알고 싶을 때는 서상규(2002)가 좋은 참고가 된다.

실증적 자료인 ‘말뭉치’(대량으로 전산화된 문어나 구어 자료를 뜻하는 코퍼스(corpus)를 번역한 우리말)를 국어 연구에 실제로 활용하기 시작한 것을 기준으로 본다면, 1980년대 중반부터 1990년대에 걸쳐 말뭉치의 개념과 말뭉치언어학 방법론을 우리나라에 소개하고 정착시키는 데에 큰 몫을 해낸 학술지《사전편찬학 연구》(연세대학교 한국어사전편찬실, 지금은 언어정보연구원)와 거기에 실린 논문들이 크게 영향을 끼쳤다.²⁾

국어학과 언어학, 전산학, 정보과학이 융합되어 ‘국어정보학’이라는 새로운 학문 분야가 싹튼 것을 시초로 삼는다면 1997년 연세대학교의 대학원에 국어정보학 협동 과정의 석박사 과정이 설치된 것이 그 첫걸음이다. 대학에서 국어정보학과 같은 이름의 강의가 개설되고, 또 국어정보학 분야의 교원을 뽑기 시작한 것은 대체로 2000년대 초에 이르러서이다. 국가적인 큰 규모의 정보화 사업을 기준으로 삼는다면 1990년대 중반 카이스트를 중심으로 한 STEP2000이나 1998년부터 10년에 걸쳐 이루어진 21세기 세종 계획 국어 정보화 기반 구축 사업이 그 시초라고 할 수 있다.

그런데 이 글에서 소개하려는 것은, 우리말 정보화의 처음은 우리가 생각하고 있는 것보다도 훨씬 이전에 외솔 최현배가 1930년대에 이미 시작하였으며, 그로부터 최현배가 이룩한 여러 결실은 오늘날까지도 여전히 우리에게 매우 새로운 정보를 제공해 주며 ‘살아 있는’ 것이라는

2) 그중 “뭉치 언어학적으로 본 사전 편찬의 실제 문제”(1988), “뭉치언어학: 사전 편찬의 필수적 개념”(1990), “말뭉치: 그 개념과 구현”(1995), “뭉치 언어학의 기본 전제”(1995) 등을 통해 ‘말뭉치’라는 말이 처음 쓰였다.

점이다. 이 글에서는 특히 우리는 우리말의 빈도 조사라는, 국어 정보화의 가장 뚜렷한 영역에 초점을 맞추어 살펴보고자 한다.

2. 우리말 빈도 조사의 현황

이제까지 한국어를 대상으로 한 전면적인 빈도 조사를 모두 모아 보면 다음과 같다.

최현배(1930), 한글의 낱낱의 글자의 쓰히는 빈수, 《조선어문연구》(연희전문학교 문과 연구집) 제1집, 연희전문학교출판부.

최현배·이승화(1955),³⁾ 《우리말에 쓰인 글자의 찾기 조사-문자 빈도 조사》, 문교부.

최현배·이승화(1956), 《우리말 말수 사용의 찾기 조사-어휘 사용 빈도 조사》, 문교부.

문영호 외(1993), 《조선어 빈도수 사전》, 과학백과사전종합출판사.

김홍규·강범모(1997), 《한글 사용빈도의 분석》, 고려대 민족문화연구원.

서상규(1998), 《현대 한국어의 어휘 빈도》, 연세대 언어정보개발연구원.

김홍규·강범모(2000), 《한국어 형태소 및 어휘 사용 빈도의 분석(1)》, 고려대 민족문화연구원.

조남호(2002), 《현대 국어 사용 빈도 조사》, 국립국어연구원.

강범모·김홍규(2004), 《한국어 형태소 및 어휘 사용 빈도의 분석(2)》, 고려대 민족문화연구원.

3) 이제까지는 보통 이 두 책의 저자 이름으로 ‘문교부’라고 써 왔지만, 실제로는 조사 책임자 격인 최현배와 실무를 전반적으로 담당해서 완수한 이승화 두 사람의 공동 저작으로 표시하는 것이 옳다.

김한샘(2005), 《현대 국어 사용 빈도 조사 2》, 국립국어원.
서상규(2014), 《한국어 기본어휘 의미 빈도 사전》, 한국문화사.

1990년대 이후 컴퓨터와 대량의 말뭉치를 활용함으로써 국어의 계량적 분석은 조사 자료의 양적인 확충은 물론이요 이제 낱말이나 형태소의 빈도에 머물지 않고, 서상규(2014)의 의미별 빈도의 조사에까지 이르고 있다.⁴⁾

3. 외솔의 우리말 찾기 조사

우리말 찾기 조사에 관한 외솔의 글은 “한글 낱말의 쓰히는 변수(사용(使用)되는 도수(度數))⁵⁾로써의 차례잡기(차서배정(次序排定))” (1930)가 유일하지만, 《외솔 최현배 박사 고회 기념 논문집》(1968:14)에 실린 목록을 보면, 《우리말 말수 사용의 찾기 조사》(1956)와 《우리말에 쓰힌 글자(한글, 한자)의 찾기 조사》(1955)라는 두 권이 외솔의 “지도 또는 공동 편찬의 업적”으로 언급되어 있어, 모두 세 편으로 볼 수 있다.⁶⁾

다음은 《우리말에 쓰힌 글자(한글, 한자)의 찾기 조사》(1955)에 나타나는 ‘최현배’에 관한 기록을 모두 모은 것이다.⁷⁾

4) 우리말을 대상으로 한 계량 연구의 흐름을 전반적으로 이해하는 데는 김한샘(2012)이 좋은 참고가 된다.

5) 최현배는 “변수(使用되는 度數), 찾기, 빈도”를 같은 뜻으로 쓰고 있으며, 이 글에서도 마찬가지이다.

6) 《외솔 최현배 박사 고회 기념 논문집》(1968:14)에서는 연도가 뒤인 1956년 책이 먼저 언급되어 있다.

속표지 사람은 아름다운 꽃을 좋아한다. 그러나, 그 아름다운 꽃이 피기까지에는 많은 노고가 숨어 있음을 모르는 이가 많다.

못술(외술)

7쪽 이 비율의 결정은, 출발 당시에, 편수국장(최 현배)과 비서관(박 창해)과 편수과장(조 병욱), 국어 편수관(국민학교 최 태호)(중고등학교 홍 응선), 말수 조사 주간(이 승화)의 합석한 자리에서 의논하여 결정지은 것이다.

14쪽 들은말 적는 법. (중략) 참여한 회원은 다음과 같다. (중략) 최현배 (중략)

22쪽 본 조사에서 취한 기준이, 최현배 교수 지은 “김고 고친 우리 말본”이란 책의 (중략)

51쪽 다음과 같은 것들을 갖추어서 편수국장께 보고하였고 (중략)

56쪽 제2차 검토를 외부 학자(이론) 교육자(실천)에게 부탁하였었는데, 맡아 보아 주신 분은 다음과 같다. (중략) 최현배 (★) (중략)

57쪽 제3차 검토에서 종결지우려다, 한자말 관계를 중심으로 말수(어휘)의 단위성 문제에 여러 가지로 통일을 지을 수가 없어서, 서울로 환도하여, 다시, 다분한 편수국장(최현배)의 지도 아래서 (중략)

57쪽 이때는 예산 관계로 사무실을 한글학회 옛 사전편찬실로 옮겼는데, 뒤이어 국장이 사임하고, (중략)

57쪽 한자말의 색출과 한자 사용의 잣기(빈도) 조사를 새로 하게 되어서, (중략) 다시 외부 학자들의 수고를 빌었다. (중략)

7) 인용할 때는 띄어쓰거나 글자의 틀림도 원문 그대로 따오기로 한다.

최현배(★) (중략)

227쪽 편의상 "최현배" 교수 지은 "우리말본" 체계에 좇아서 (중략)

232쪽 본 사무의 조직은, (중략) 편수국장의 직할(直轄) 아래서,
(중략) 편수국장 1. 최현배 2. 신 태현 (중략)

233쪽 (중략) 과정을 넘어설 적마다의 검사와 수정에 골머리를 앓고, 편수국장 님의 결재서의 너무도 엄격하신 검열에, 혼이 나고 도망치듯 그만 두는 사람도 많았었다.

235쪽 (마) 외부 교육자 및 학자의 협조 (중략) ㄱ. 말수조사 소리 차례표의 제3 차 검토를 해 주신 분네(환도하기 전 부산에서) (중략) 최현배 ☆ (중략) ※이 중에서 가장 많은 분량의 수고를 하신 이는 최현배 선생님과 김 계원 선생이다.

235쪽 ㄴ. 말수조사 소리차례표의 제6차 검토를 해 주신 분 *최현배 ☆ (중략) ※이 가운데서 *는 수고를 특히 많이 하신 분이다.

237쪽 잡비 소모를 적게 하려고, 최 선생님의 주선으로 다시 이사한 곳은; 한글학회의 단 하나 남은 부동산인 "옛 사전 편찬실"이었다.

239쪽 학적 양심을 지키기 위해서, 시간과 예산을 몇 번이곤 어기었다. 그럼에도 불구하고, 장관님과 차관님을 비롯하여, 국장님(최현배 선생님, 신 태현 선생님)과 과장님께서 이 일을 할 수 있게끔 주선해 주셨음에 대하여 감사한다.

이와 같이 외솔은 우리말 찾기 조사의 각 기준과 방법론, 실제 자료의 검토에 이르기까지의 모든 과정은 물론, 추진 조직과 재정적 지원까지 큰 수고로움을 마다하지 않았음을 알 수 있다. 이 기록들을 통하여 우리는 최현배·이승화(1955, 1956)의 기획에서 완성에 이르기까지 외솔이

어떠한 역할(지도 또는 공동 편찬)을 감당하였는지를 이해할 수 있다.

언어 조사를 통해서 믿을 수 있는 결과를 얻기 위해서는, 그 목적과 대상, 조사의 방법론을 어떻게 설정하느냐가 매우 중요하다. 조사의 목적이 정해지면 그에 맞추어 조사 대상이 될 언어 단위가 설정되며, 그 결과의 내용과 형태가 정해지기 때문이다. 조사 방법론에는 조사 자료의 구성과 선정, 나아가서는 실제 언어의 분석 기준과 각종 원칙의 수립과 적용에 걸친 전반의 문제가 포함된다. 이제 이러한 요건 등을 중심으로 하여 최현배(1930)와 최현배·이승화(1955), 최현배·이승화(1956)를 분석하여 살펴보기로 한다.

3.1. “한글의 낱말의 글자의 쓰히는 번수” 조사(1930)

최현배(1930)는 10쪽의 짧은 논문인데, 그 첫 문장에서 이 연구가 이루어진 시기에 대해 말하고 있다.⁸⁾

- (a) 일은 발서 해 놓앗건마는, 이제 세상(世上)에 공포(公布)하기를 뜻하고 그이름을 붙이랴니 별로 간단명료(簡單明瞭)한 것이 되지 아니하고, 이렇게 좀 지저분하게 되었다.(최현배 1930:100)

이 논문집은 1930년 12월에 발간되었지만 논문의 꼬트머리에 ‘1930,3,3(一九三〇,三,三)’이라는 날짜가 기록되어 있고, 이 조사 대상에 1929년 11월 8일 자 신문이 사용되었으므로 그 4개월 사이에 조사된 것으로 짐작된다.

8) 읽는 이들의 불편함을 덜고자, 원문에 한자로 표기된 부분에는 한글을 앞에 내세우고 원래의 한자는 괄호 속에 넣었다.

3.1.1. 조사의 목적

최현배(1930)의 조사 목적은 아래와 같다.⁹⁾

- (b) 우리의 한글(훈민정음(訓民正音), 언문(諺文))의 이십팔자(二十八字) 가운데 이미 폐용(廢用)된 ㆍ △ ○ ㅎ 사자(四字)를 제(除)한 이십사자(二十四字)가 일상어문(日常語文)에 쓰힐 적에 어느것이 가장자조 쓰히며, 어느것이 가장드물게 쓰히며, 또 그 중간(中間)에 오는 글자의 순서(順序)는 어떠한가? 이 낱낱의 글자의 쓰히는 변수를 조사하여 그것으로써 각자(各字)의 차례잡기(차서배정(次序排定))을 해보라는 것이 이 글월의 내용(內容)의 목적(目的)이다.(최현배 1930:100)
- (c) 위선 낱낱의 글자의 일상어문(日常語文)에 쓰히는 잦은번수(頻度)를 조사하여 그 차례잡기(차서배정(次序排定))을 하여 본 것이 나의 이 글월의 내용(內容)이다.(최현배 1930:101)

또 다른 목적으로, 이 변수 조사의 통계를 사람들이 각자 자기의 목적에 따라서 활용할 수 있을 것이라고 하면서, 하나의 구체적인 사례로 '한글 가로쓰기'를 위한 새 글자 모양 변경의 순서 잡기에 바로 쓸모가 있음을 말하고 있다.

- (d) 이러한 통계(統計)는 어대에 쓰히는 것인가? 이것은 내가 여기에 구체적(具體的)으로 대답(對答)하지 아니해도 좋을 것이다. 그것은 사람을 딸아 그 목적(目的)을 딸아 쓰는대로 쓰힐것이기

9) 인용된 글 속의 밑줄은 필자가 친.

때문이다. 다만 내가 여기에 한가지말해 두고자 하는 것은 우리 조선어연구자(朝鮮語研究者)들의 당면(當面)의 한 문제(問題)가 되어있는 한글의가로쓰기(횡서(橫書))를 실현(實現)하고자 그 자체(字體;글자모양)를 다소(多少) 변경(變更)하지 아니하지 못할 형편(形便)인데, (중략) 그 우열(優劣)의 정도(程度)는 따로 과학적(科學的) 실험방법(實驗方法)에 의(依)하여 결정(決定)될 것이다. 그리하여 그 된품의 우열(優劣)의 정도(程度)로써 잡은 각자(各字)의 차례가 여기서 쓰힘의 변수의 다소(多少)로 써 잡은 차례하고 서로 꼭 일치(一致)할것같으면 이상적(理想的)이라 할만할 것이다.(최현배 1930:109)

3.1.2. 조사의 단위

‘한글 각 글자의 변수에 따라 차례를 정하’려는 목적을 이루기 위한 조사 단위는 당연히 (c)에서 본 바와 같이, ‘날날의 글자’이다. 그런데 우리말 맞춤법이 통일되기 전이었던 당시의 현실을 고려해 볼 때, 조사 단위인 글자의 목록을 정함에 있어서, 실제로 자료에 나타난 표기 실태를 있는 그대로 반영할 것인지, 아니면 일일이 바른 표기로 고치면서 조사할 것인지에 대한 태도를 정하지 않으면 안 되었을 것이다.

(e) 그러나 현금(現金)의 실제(實際)의 글자는 꼭 학리적(學理的)으로 보아 근본적(根本的)인 기초문자(基礎文字)만을 쓰지 아니하고, 그 밖에 실제적(實際的) 편의(便宜)를 따아, 이것 저것을 서로 합(合)하여서 쓴다. 그러므로 이제 글자의 쓰히는 변수를 조사(調査)함에는 이 실제(實際)의 용법(用法)을 따르는 것이 편리(便利)하다.(최현배 1930:101)

이와 같이 최현배(1930)에서는 우리말의 ‘실제의 용법’을 잘 드러내면서도 어느 정도 일관성 있는 표기를 하고 있는 신문을 조사 대상으로 삼은 것이다.

3.1.3. 조사 대상 자료

오늘날에는 빈도 조사에 컴퓨터와 대규모의 언어 자료(말뭉치)를 활용하는 것이 일반적이는데, 이때 어떠한 자료를 어떻게 선택해서 조사 대상을 구성할 것인가가 매우 중요하다. 최현배(1930)의 조사에서 어떤 자료를 대상으로 했는지는 앞에 인용한 (c)에 잘 드러난다. 여기서는 ‘일상어문(日常語文)’이라는 말을 쓰고 있는데, 그렇다면 ‘일상어문’을 어떻게 찾아서 정할 것인가 하는 문제가 뒤따른다.

(f) 다압에 일어나는 문제(問題)는 이렇게 세운 마흔자의 쓰히는 번수를 조사(調査)함에는 어떠한 말과 글을 쓸 것인가이다. (중략) 그래서 나는 현금(現今) 우리말로써 내는 신문(新聞) — 동아일보(東亞日報), 조선일보(朝鮮日報), 중외일보(中外日報)에서 각각(各各) 가장 적당(適當)하다고 생각되는 일면(一面)을 택취(挾取)하였다. (최현배 1930:102)

여기에는 조사 자료 선택의 두 가지 조건을 밝히고 있다. “우리말로써 내는 신문”, “가장 적당하다고 생각되는 면”이 그것이다.

3.1.4. 자료의 표본 추출과 최초의 신문 말뭉치의 구성

우리말 번수 조사에서는 “우리말로써 내는” 3종 신문으로 ‘일상어문’을 대표하는 것으로 여기고 있는데, 신문을 활용한 최현배의 구체적인

조사 방법에서 우리는 오늘날의 언어 조사에 널리 쓰이는 균형 말뭉치 구성 방법에 매우 가까운 모습을 발견하게 된다. 그 근거는 신문의 선택 방법에서 살펴볼 수 있다.

외솔은 3종의 신문에서 자료 표본을 선택하는 과정에서 의도적으로 달(2월, 3월, 11월)과 날짜를 각각 서로 다르게 정하였을 뿐 아니라, 신문의 면을 선택할 때에도 기사의 내용은 물론 표기 방식까지 꼼꼼히 고려하였음이 틀림없다. 기사의 내용에서도 역시 대체로 사회 일반의 사건을 다룬 내용이 중심을 이루고 있어, 일상적인 언어 사용의 실태가 드러나도록 한 것이 분명하다(서상규 2008:13~19).

변수 조사 대상으로 신문이 선택된 또 하나의 이유로는 신문이 당시의 자료 가운데서 표기의 원칙을 일정하게 적용한 자료라는 점, 기사마다 그 주제와 내용이 서로 독립적이며 따라서 내용이 다양하다는 점 등이었으며, 조사 대상을 선정하는 단계에서 이 점이 당연히 고려되었을 것이다. 그러므로 ‘변수 조사’에서 사용한 신문 자료는 오늘날의 기준으로 보아도 한글 중심 표기의 신문 3개 면, 79개의 독립된 내용의 표본(기사)으로 구성된 훌륭한 ‘신문 말뭉치’였던 것이다.

신문 표본의 선택과 변수 계산 방법에서도 매우 높은 수준의 정보학적 고려가 이루어져 있다. 최현배(1930)에서는 한글이 쓰인 변수를 계산할 때에는 3개 신문의 각 면을 10개의 구역으로 나누어서 조사하여 따로 센 수와 합친 수를 표에 함께 보여 줌으로써, 면 전체뿐 아니라 기사 구역에 따른 한글 사용의 실태를 비교해 볼 수 있게 했다. 3개 신문(면)을 각각 10개의 구역으로 나누어 글자 수를 각각 세었으므로, 결국 실제 변수의 조사는 신문당 10번, 모두 합해 30번 이루어진 셈이고, 이 30개의 변수 조사 결과표를 비교해 본다면, 충분히 신뢰할 수 있는 통계를 얻을 수 있을 뿐 아니라 당시 글자 사용의 실태를 분명하게 파악할

수 있게 된다. 이와 같이, 외솔의 ‘변수 조사’는 그 조사 대상의 선택 과정으로부터 변수 세기의 절차와 과정에 이르기까지 매우 합리적으로 짜였던 것이다.

결국 최현배(1930)의 조사 자료가 당시의 ‘일상적인 어문’을 충분히 대표하는 것이라는 결론에 이른다.

3.1.5. 조사 결과와 해석

‘변수 조사’의 결과는 구획별로 나누어진 통계표와 그것을 하나로 뭉뚱그린 종합통계표의 넷으로 되어 있다.

(g) 이것은 상기(上記) 삼통계표(三統計表)를 하나로 뭉뚱거린 것이다. 곧 이 조사통계(調查統計)의 결론(結論)이다.

(표 생략) (중략)

이것은 다시 홑소리와 닿소리로 가를것같으면

홑소리(一一字)

ㅏ ㅣ ㅓ ㅡ ㅗ ㅜ ㅠ ㅋ ㅊ ㅌ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅈ ㅊ ㅋ ㅌ ㄴ ㄷ ㄹ

닿소리(一九字)

ㄴ ㄱ ㄹ ㅅ ㄷ ㄹ ㅇ ㅎ ㅅ ㅂ ㅅ ㅈ ㅊ ㅊ ㅌ ㅌ ㅌ ㅌ ㅌ ㅌ ㅌ

로 된다. 이것을 다시 다른 말로 하면, 홑소리나 닿소리나를 물론(勿論)하고 홑소리(단음(單音)) 곧 기초문자(基礎文字)가 제이차적(第二次的)인 거듭소리(중음(重音))보다 자주 쓰히며, 그중에도 거듭홑소리는 거의 안 쓰인다 할만하며, 뒤시웃한 닿소리 곧 짝소리(쌍자음(雙子音))는 매우 적어서 모다 숨하여도 그 변수의 분수(比喩)가 천(千)에 칠(七)밖에 되지 아니한다.(최현배 1930:108~109)

여기에 보인 한글 글자의 사용 빈수에 따른 차례와 상대적인 분포의 차이는 바로 1930년대 우리 글자 생활의 단면을 있는 그대로 보여 주는 ‘사진’인 셈이다. 이러한 조사 결과가 남아 있기에 비로소 우리는 그 이후의 빈도 조사 결과들과 비교함으로써 당시 국어의 특성은 물론이요 시대를 달리하면서도 이어지고 있는 국어의 보편적 특성을 잘 파악할 수 있는 것이다.

3.1.6. 자료와 계산에서 결함과 정확성의 문제

최현배(1930:102~103)에서는 자료 자체가 가진 문제점으로, ‘마침법’ 즉 철자법의 혼란에 따른 문제점과 계산의 정확성에 관한 문제점을 스스로 밝혔다.

(h) 그러나 이 재료(材料)는 그 스스로가 다소(多少) 불완전(不完全)한 흠함(欠陷)이 없지 아니하다. 그는 곧 마침법(철자법(綴字法))이 바르게 정리(整理)되지 못한 때문이다. (중략)

다함에 재료자체(材料自体)의 흠함(欠陷)밖에 또 조사(調査)하는 계산(計算)이 절대정확(絶对正確)을 보증(保証)하기 어렵은 흠함(欠陷)이 있다. 이것은 누구든지 이러한 시험(試驗)을 해본 경험(經驗)이 있는 분은 다 아시는바와 같이, 아모리 조심해서 센다 하더라도 그 절대정확(絶对正確)은 썩 기대(期待)하기가 어렵다. 그러나 한가지 애답은 미봉적(弥縫的) 위안처(慰安处)가 없지 아니하다. 그는 모든것이 골고로 다 다소간(多少間)의 불확실성(不確實性)이 있다 하면 이는 그 평균비율(平均比率)을 상(傷)히는 일이 없어질 것이다. 다만 이 통계표(統計表)가 조심으로써 한 것은 물론(勿論)이요, 한가지를 두세번씩 센 것이 여간 많잖음을 붙여말하여 둔다.(최현배 1930:102~103)

빈도 조사 과정에서는 분석의 단위나 기준, 사용 자료의 성격에 따라서 갖가지 문제에 부닥치게 된다(서상규 1998b:234~238을 참조 바람). 그러므로 빈도 조사의 결과물에는 단지 통계 수치만을 보일 것이 아니라 조사의 목적부터 단위 설정, 기준, 작업 과정, 적용 과정의 문제점 해결 등 상세한 기록을 덧붙이는 것이 바람직하다.

이와 같이, 최현배(1930)의 변수 조사는 그 방법론적인 면만 보아도, 이제까지 최현배·이승화(1955)를 우리말 찾기 조사의 시초로 보아 온 우리말 연구 역사를 고쳐 써야 할 만큼의 수준 높은 내용을 보여 준다. 국어의 계량적 연구의 기점을 1955년이 아니라 1930년으로, 그보다 25년이나 앞당길 수 있는 것이다. 이것은 맞춤법 통일 이전인 1920년대 말의 한글 표기와 자모 실태를 기록한 실증적 자료로서 가치를 가질 뿐 아니라, 신문 기사 선택과 통계 제시 방식, 조사 단위의 설정과 방법론에 관한 요건을 다 갖추어, 오늘날의 관점에서 보아도 바람직한 빈도 조사의 요건을 충분히 갖춘 것이었다.

3.2. 《우리말 글자 사용의 찾기 조사》(1955)/《우리말 말수 사용의 찾기 조사》(1956)

앞서 밝힌 바와 같이, 외솔의 “지도 또는 공동 편찬의 업적”으로 기록된 《우리말에 쓰힌 글자(한글, 한자)의 찾기조사》(1955)와 《우리말 말수 사용의 찾기 조사》(1956)로 실현된 조사의 필요성은, 최현배(1930)에 이미 다음과 같이 언급되어 있다.

- (i) 우리 조선말가운데 어떠한 낱말(단어(單語))이 가장 자조 쓰히는 지? 가장 자조 쓰히는 변수로 보아서, 오백단어(五百單語), 일천단어(一千單語), 이천단어(二千單語), 삼천내지사천단어(三

千乃至四千單語)의 각(各) 범위(範圍)에 소속(所屬)할 낱말이 과연(果然) 어떠한것인지, 이러한 통계적(統計的) 조사(調查)에는 많은 노력(努力)과 오랜 세월(歲月)을 요(要)한다. 이것은 아직 두고, 위선 낱말의 글자의 일상어문(日常語文)에 쓰히는 잦은빈수(빈도(頻度))를 조사하여 그 차례잡기(차서배정(次序排定))을 하여 본 것이 나의 이 글월의 내용(內容)이다.(최현배 1930:101)

최현배가 낱말의 빈도수를 조사하려는 뜻은, 일정한 기준에 따른 낱말의 목록(이를테면 교육용 기본 어휘)을 만들고자 하는 것이었다. 1930년의 빈수 조사에서는 나중으로 미룰 수밖에 없었던 이 일이 20년 이상 지난 뒤에야 비로소 실현된 것이 바로 《우리말에 쓰인 글자의 잦기 조사(文字 頻度 調査)》였다.¹⁰⁾ 이것은 크게 세 가지 조사로 이루어져 있다.

- A. 우리말 말수(어휘) 사용의 잦기(빈도) 조사 (3가지)
- B. 우리말에 쓰인 글자의 잦기 조사
 - I. 한글 사용의 잦기 조사 (104가지)
 - II. 한자 사용의 잦기 조사 (3가지)
- C. 우리말 풀이씨(용언)의 도움줄기(보조어간)와 씨끝(어미)의 잦기 (빈도) 조사 (18가지)

10) 최현배·이승화(1955)의 책머리 ‘이 책을 내면서’에서 “이 조사는 4264(1951)년 3월에 계획하고, 동년 4월 13일부터 조사하기 비롯하여, 4288(1955)년 6월 초에 마쳤는데”라고 하여, 조사에 약 5년 이상의 시간이 소요된 것으로 밝히고 있다.

최현배·이승화(1955)에는 B와 C, 최현배·이승화(1956)에 A가 수록되어 있어, 책의 구성 순서와 간행된 순서가 거꾸로 되었는데, 그 까닭을 최현배·이승화(1955)의 속표지에서 “다만, 예산 관계로 우선 B만을 박아 내기로 한 것이 이것이다.”라고 밝히고 있다.

3.2.1. 조사의 목적

[표 1] 최현배·이승화(1955, 1956)의 조사 목적

	A. 말수(어휘) 사용	B. 글자 빈도		C. 도움줄기와 씨끝 빈도
		I. 한글 사용	II. 한자 사용	
목적	우리말 말수(어휘)가 사용되는 짓기(빈도)의 실태를 조사하여, 과학적인 국어의 기본 형태를 파악하고, 우리말의 합리적인 사용을 꾀하며, 국어의 정상적인 발달 및 정화 운동을 목표하는 교과서 편집이나 계몽 활동에 활용하고, 나아가서는 국어학 연구의 참고 자료로 제공(7쪽)	우리말에 쓰이는 글자의 실태를 과학적으로 알아내어, 글자와 관계있는 각종 문화 활동에, 참고로 삼거나 활용(63쪽)	우리 말에서 쓰이고 있는 한자 말에서 한자말을 뽑아내어 가지고, 한자에 대한 온갖 문제의 검토 및 해결에 제공(167쪽)	도움줄기와 씨끝의 실태를 과학적으로 밝히어 국어의 창조적 아름다움과 발전적 사용을 위하여, 문법연구, 수사법연구, 등의 연구 검토의 참고 자료로 제공(227쪽)

3.2.2. 조사의 단위와 조사 방법

[표 2] 최현배·이승화(1955, 1956)의 조사 단위와 방법

	A. 말수(어휘) 사용	B. 글자 빈도		C. 도움줄기와 씨끝 빈도
		I. 한글 사용	II. 한자 사용	
단위 설정	말수(어휘) 문교부(1955:20~33)에서 상술함	낱내(음절), 낱자(단자)	한자말 또는 한자말 이 끼어들어 서 된 우리말 말수(어휘)	줄기 이외의 것 도움줄기(보조어 간), 씨끝(어미)
조사 방법	모든 글월 ¹¹⁾ 을 낱말의 단위에 좇아 분석하여 조사(1955:7)			
조사 기준	가) 목차와 삽화 속의 글월 포함 나) 보통 사용하는 말만 조사, 홀로이름씨(고유명사) 제외 다) 현대말을 주종으로 하여, 대중말(표준말)로 조사 라) 준말(약어)과 안준말은 따로(것/거, 무엇/무어/뭐……) 마) 씨가지는 예외로만 따로 세움(들, 꿈, 금, 네) 등	a) 낱내(음절) 따른 조사 b) 낱자(단자) 따른 조사	A) 한글학회 맞춤법에 따름 B) 글자는 찾기 별/부수별로 배열 C) 뜻이나 음이 같은 경우	최현배의 《우리말본》 체계에 따름. A) 줄기(어간)에 잇달은 그대로의 조사 B) 도움줄기(보조어 간)만의 조사 C) 씨끝(어미)만의 조사

11) 이에 대해서는 “산 말인 입말(口調)을 녹음한 것, 곧, 실사회의 대중이 많이 모인 시장, 극장, 오락처, 술자리, 각종 대합실 등에서 몰래 녹음한 것을 가지고, 조사함이 이상적인 국어의 실태 조사이라고 하겠으나, 시설과 비용 등 온갖 사정이 허락지 않아서”(1955:7) ‘글월을 조사 대상으로 삼았다고 밝히고 있다. 사실, 최현배·이승화(1956) 이후 2010년 대에 이르기까지 꽤 많은 말뭉치가 구축되고 국어 자료 처리 기술도 좋아졌음에도 불구하고, 여전히 살아 있는 입말의 제대로 된 빈도 조사의 성과는 아직도 이루어지지 않고 있다. 필자는 21세기 세종 계획 구어 주석 말뭉치를 바탕으로 동음이의어 구분까지 끝난 말뭉치를 몇 년 간에 걸쳐 완성했으며, 머지않아 첫 구어 빈도 조사의 성과를 이룩할 것으로 기대하고 있다.

3.2.3. 조사 대상 자료

[표 3] 최현배·이승화(1955, 1956)의 조사 대상 자료

	A. 말수(어휘) 사용	B. 글자 빈도		C. 도움줄기와 씨갈 빈도
		I. 한글 사용	II. 한자 사용	
표본 추출 방식	(1) 조사목적물의 부문 별 비율(%) ¹²⁾ 초중등 교과서(50%) <국어, 가사, 사회생 활 30%, 과학, 실업류 20%> 일반 간행물(50%) <문학, 예술류 30%, 신문, 잡지, 방송원고, 국회 의사록 20%>	A(말수 사용의 찾 기 조사)의 5%분 량에 해당, 그 각 분 야에 좇아서의 그 비율은, 말수 조사 에서의 비율과 같 이 함	A(말수 사용의 찾 기 조사)의 전체	A(말수 사용의 찾기 조사)의 5% 분량에 해당, 그 각 분야에 좇아서의 그 비율 은, 말수 조사에서 의 비율과 같이 함
규모	56,485날말 /총찾기 2,218,727	A조사의 5.18% (11종) /총찾기 114,225	글자 총수 3,312자 /총찾기 1,016,731	

12) 부문별 비율 결정에 대해 “이 비율의 결정은, 조사 출발 당시에, 편수국장(최현배)과 비서관(박창해)과 편수과장(조병욱), 국어 편수관(국민학교 최태호)(중고등학교 홍웅선), 말수 조사 주간(이승화)의 합석한 자리에서 의논하여 결정지은 것”(1955:7)으로 기록하고 있다.

3.2.4. 조사 결과와 해석

[표 4] 최현배-이승화(1955, 1956)의 조사 결과

	A. 말수(어휘) 사용	B. 글자 빈도		C. 도움줄기와 씨끝 빈도
		I. 한글 사용	II. 한자 사용	
결과	찾기(빈도) 차례 결과표, 소리(頭音) 차례 결과표	낱내(음절)따른 조사 결과표, 낱자(단자)따른 조사 결과표	찾기(빈도) 차례 결과표, 부수 차례 결과표	소리 차례, 찾기 차례
효용	<ol style="list-style-type: none"> 1) 국어 사용의 과학적 근거/국어의 합리적인 사용/교과서 편찬에 사용 2) 용어의 제정/계몽적인 문화 활동 3) 우리말의 실태 짐작 4) 우리말 연구 및 정화 운동 5) 순수 우리말과 한자말, 들은말의 비율 파악 6) 사전의 말수 보충 및 취사선택 7) 국어 생활 실태 조사의 참고 	<ol style="list-style-type: none"> 1) 기초 학년 교과서 편찬시의 제시 낱내 음절의 합계나 순서 결정 2) 우리말의 언어미학으로의 검토 3) 언어심리학 면에서의 참고 자료 4) 한글 적는 법의 연구 자료 5) 한글 자모, 자형의 재고찰 및 개량의 자료 6) 인쇄 공장의 활자 수량 준비와 나열 순서 결정, 개량 7) 타자기의 글자판 배치 8) 맞춤법의 연구 자료 9) 한글 가로 풀어쓰기 연구 및 운동 10) 언어 및 글자 조사 방법 연구 	<ol style="list-style-type: none"> 1) 일상생활에서의 한자 폐지 또는 제한 사용에 관한 문제 검토 2) 한자의 자형, 자체 연구 3) 한자 타자기, 인쇄 공장 설비 전반 검토에 참고 4) 국어 정화 운동, 언어 심리학적 연구, 언어 미학적 연구, 운필상의 물리적 연구, 서도상의 미학적 연구 	<ol style="list-style-type: none"> 1) 우리말 사용의 연구 자료 및 참고 자료 2) 우리말의 구성적 연구 곧 문법 연구의 참고 자료 3) 수사법 연구 4) 국어의 제반 문제 해결

3.2.5. 자료와 계산에서 결함과 정확성의 문제

최현배·이승화(1955:164~165)에서도, 앞서 소개한 최현배(1930) 때와 마찬가지로 어려움을 겪었음을, 이미 출판된 책에 종이를 덧대서 표를 고친 부분이 여러 곳인 것이라든가 “부산 고등학교 상급 학생 연 196명의 협조로 카아드의 작성 및 대조를 마치고서, 부산 제일 상업 고등학교 3학년 전원의 협조로, 잣기(빈도)의 합계와 검산을 한 것을 다시 정리원들의 재검산으로 마치었다”(최현배·이승화 1955:53), “그런데, 두 곳에서 나온 총계의 숫자가 서로 맞지 않아서, 무척 오랫동안 많은 이가 고생을 다시 하였었다”(최현배·이승화 1955:58), “말수 조사의 최종 통계인 잣기별 및 첫소리별의 계급적 집계표의 숫자의 이가 맞지 않아, 고생을 하다가 최후로 서울 상고의 주판 교사의 손으로 단락을 지웠다”(최현배·이승화 1955:232)는 등의 기록을 통해서도 확인할 수 있다.

이와 같이, 최현배·이승화(1955, 1956)는 언어 조사에서 이루어질 수 있는 여러 단위의 설정 방법과 구체적인 적용, 조사 대상 자료의 구성과 선택, 다양한 목적에 따른 언어 정보의 추출 방식, 결과 표시의 모형 제시 등에 이르는, 조사 방법론상의 갖가지 요소에 대해 매우 상세히 기술함으로써,¹³⁾ 이후에 이루어질 계량적 연구의 기대 수준을 크게 높였다. 또 조사의 과정에 필요한 현실적인 여러 절차와 비용, 시간, 일손의 구체적인 내역, 갖가지 문제점 등을 소상히 기록으로 남겨 후속 연구의 길잡이가 되도록 하였다.

13) 이에 대해 이승화(1955:238)는 “조사 방법에 있어서도, 불 만한 점이 한 둘이 아니다. 그러나 이 본 보고서에 있는 내용의 조사 방법이 장차에 판이들이 조사하는 경우에 있어서, 좋은 참고 자료가 될 것이다.”라고 말하고 있다.

4. 맺음말

이 글에서는 빈도 조사의 요건으로서 목적과 대상의 설정, 조사 대상의 자료 선정과 활용의 방법과 절차, 표본의 선택과 계량, 조사 결과와 해석 등 여러 측면에서 최현배의 업적을 살펴보았다.

이제까지 국어 계량 분석, 나아가 국어정보학적 연구의 시초로 삼아 온 최현배·이승화(1955, 1956)는 오늘날의 시점과 기준으로 본다 해도, 매우 훌륭한 국어 정보화의 성과임은 두말할 나위가 없다. 그뿐만 아니라, 외솔 최현배의 이러한 국어 정보화에 대한 앞선 깨달음과 노력은 이미 최현배(1930)에서 싹트었으며, 그 성과 역시 국어에 대한 최초의 실증적, 계량적 연구임이 이미 서상규(2008)에 의해서 밝혀진 바 있다. 다시 말해 최현배·이승화(1995, 1956)는 최현배(1930)에서 세워진 계량적 연구의 방법론을 대규모의 연구로써 실현한 성과였다. 방법론을 수립하고 실현하는 데 실로 20년 넘게 걸렸던 것이다. 이는 우리말에 대한 정보화, 국어 정보학적 연구의 시초가 1930년이라는 것을 뜻하는 것이며, 외솔 최현배의 탄생 120주년을 맞으면서, 이제까지 우리가 잘 못 알아온 국어 정보화의 시초에 대한 우리의 인식과 국어학사의 기술을 고쳐 써야 할 때에 이르렀다고 할 수 있다.

최현배·이승화(1955)의 속표지에는 외솔의 짧은 글이 적혀 있다.

“사람은 아름다운 꽃을 좋아한다. 그러나, 그 아름다운 꽃이 피기까지는 많은 노고가 숨어 있음을 모르는 이가 많다. 못술(외솔)”

참고문헌

- 강범모·김홍규(2004), 《한국어 형태소 및 어휘 사용 빈도의 분석(2)》, 고려대 민족문화연구원.
- 고영근(1995), 《최현배의 학문과 사상》, 집문당.
- 김석득(1983), 《우리말 연구사》, 정음문화사.
- 김석득(2000), 《외솔 최현배 학문과 사상》, 연세대출판부.
- 김한샘(2005), 《현대 국어 사용 빈도 조사 2》, 국립국어원.
- 김한샘(2012), 한국어 어휘 계량 연구의 성과, 《한민족문화연구》 제41집, 한민족문화학회, 39~74.
- 김홍규·강범모(1997), 《한글 사용 빈도의 분석》, 고려대 민족문화연구소.
- 김홍규·강범모(2000), 《한국어 형태소 및 어휘 사용 빈도의 분석(1)》, 고려대 민족문화연구원.
- 문영호 외(1993), 《조선어 빈도수 사전》, 과학백과사전출판사.
- 서상규(1998a), 《현대 한국어의 어휘 빈도》, 연세대 언어정보개발연구원.
- 서상규(1998b), 말뭉치 분석에 기반을 둔 낱말 빈도의 조사와 그 응용, 《한글》 242, 한글 학회, 225~270.
- 서상규(2001), 말뭉치를 이용한 국어 문법 연구의 현황과 방향, 《21세기 국어 정보화와 국어연구》, 고려대 민족문화연구원 국어연구소 편, 월인.
- 서상규(2002), 국어정보학 연구의 현황과 방향, 《국어학 연구 50년》, 이화여자대학교 한국문화연구원 편, 해안, 431~463.
- 서상규(2006), 의미 빈도 사전과 어휘 연구, 《조선학보》 198호, 조선학회.
- 서상규(2008), 한글의 변수 조사와 외솔 최현배, 《한글》, 한글학회, 35~71.
- 서상규(2014), 《한국어 기본어휘 의미 빈도 사전》, 한국문화사.
- 서상규·한영균(1999), 《국어정보학 입문》, 태학사.
- 외솔 최현배 박사 고회 기념 논문집 간행회(1968), 《외솔 최현배 박사 고회 기념 논문집》, 정음사.
- 이승화(1955), <우리말 말수 조사의> 본 조사의 경과, 《우리말에 쓰인 글자의 찾기 조사》, 문교부.

- 조남호(2002), 《현대 국어 사용 빈도 조사》, 국립국어연구원.
- 최현배 선생 환갑 기념 논문집 간행회(1954), 《최현배 선생 환갑 기념 논문집》, 사상계사.
- 최현배(1930), 한글의 낱말의 글자의 쓰이는 변수, 《조선어문연구》(연희전문 학교 문과 연구집) 1집, 연희전문학교출판부, 100~109.
- 최현배·이승화(1955), 《우리말에 쓰인 글자의 잦기 조사-文字 頻度 調査-》, 문교부.
- 최현배·이승화(1956), 《우리말 말수 사용의 잦기 조사: 어휘 사용 빈도 조사》, 문교부.