

조선 후기 왕실 한글 어필의 자형 분포와 대표 글자 도출 연구

서체 디자인에서 빈도·구조·자모 커버리지의 삼중 기준을 중심으로

Glyph Distribution and the Derivation of Representative Letterforms in Late Joseon Royal Hangeul Calligraphy

Focusing on the Three Criteria of Frequency, Structure, and Jamo Coverage in Typeface Design

주 저 자 : 홍지경 (Hong, Ji Kyung) 부산대학교 디자인학과 박사과정

교 신 저 자 : 김성계 (Kim, Sung Kye) 부산대학교 디자인학과 교수
sungkye@pusan.ac.kr

<https://doi.org/10.46248/kidrs.2026.3.912>

접수일 2026. 08. 16. / 심사완료일 2026. 08. 28. / 게재확정일 2026. 09. 07. / 게재일 2026. 09. 30.

Abstract

Typeface design from historical materials begins with a sampling problem: which letterforms to draw first. This study builds a corpus from two late-Joseon royal Hangeul manuscripts — *Jeongjo Eopil Hangeul Pyeonjicheop* and *Gonjeon Eopil*, Queen Hyoui's Hangeul transcription — and derives representative letterforms under three criteria: frequency, structure, and *jamo* coverage. Decomposing 13,327 tokens and 674 distinct glyphs into *jamo* shows that the archaic vowel *araea*(·) ranks among the most frequent vowels, that vertical clusters occur 8-9 percentage points more often than in modern Korean, and that 80% of cumulative usage is reached at 156 glyphs. The 156 most frequent glyphs deviate from the corpus cluster composition by 6.9 percentage points, whereas a stratified set of 156 reduces this to 0.4. Yet both sets score almost identically on the Jamo Coverage Index (0.691 and 0.689), showing that structural and *jamo* representativeness are independent criteria. Adding nine glyphs that carry the uncovered *jamo* yields a 165-glyph set retaining 79.5% coverage, covering every *jamo* observed, and raising the JCI to 0.828 — above the 0.792 of the 400 most frequent glyphs. The study proposes a triple-criterion sampling procedure for allocating design effort at the outset of historical Hangeul typeface development. JCI is used as a diagnostic measure of input composition, not as a predictor of output quality.

Keyword

Hangeul Typography(한글 타이포그래피), Royal Hangeul Calligraphy Corpus(왕실 어필 역사 코퍼스), Representative Letterforms(대표 글자)

요약

역사 자료를 바탕으로 한 서체 디자인은 어떤 글자를 먼저 설계할 것인가라는 표집 문제에서 출발한다. 본 연구는 『정조어필한글편지첩』과 『곤전어필』을 서체 연구용 코퍼스로 구축하고, 빈도·구조·자모 커버리지의 삼중 기준으로 대표 글자를 도출하였다. 13,327 토큰, 674 자형을 낱자로 분해한 결과 아래아(·)가 상위 활자로 나타났고, 세로모임 비중은 현대 한글보다 8-9%p 높았으며, 누적 사용량 80% 지점은 156자였다. 빈도순 156자의 모임꼴 구성 편차는 최대 6.9%p였으나 총화 156자는 0.4%p로 줄었다. 두 세트의 자모 커버리지 지수(JCI)는 0.691과 0.689로 유사하여 구조 대표성과 자모 대표성이 독립적 기준임이 확인되었다. 미포괄 자모를 담은 9자를 보강한 165자는 커버리지 79.5%를 유지하며 자모를 전수 포괄하고 JCI를 0.828로 높였다. 본 연구는 착수 단계의 조형 검증 자원을 배분하는 삼중 기준 표집 절차를 제안한다.

목차

1. 서론

- 1-1. 연구 배경 및 목적
- 1-2. 연구 범위 및 방법

2. 이론적 배경

- 2-1. 한글 서체 대표 글자 선정 기준의 검토와 한계

- 2-2. 자모 커버리지 지수: 참조 세트의 설계 진단 지표

- 2-3. 연구 대상 자료: 왕실 어필 두 종의 서체적 성격

- 2-4. 역사 자료 서체 복원에서 현대 한글 대표 글자의 부적합성

3. 연구 방법

- 3-1. 서체 연구용 코퍼스의 구축

- 3-2. 날자 분해와 모임꼴 분류: 조형 단위의 설정
- 3-3. 분석 지표: 설계 판단을 위한 다섯 척도

4. 분석 결과

- 4-1. 코퍼스 규모와 자형 분포: 복원 난이도의 계량
- 4-2. 날자 사용 빈도: 기준 날자의 재설정
- 4-3. 모임꼴 구조 분포: 조형 검증 자원의 재배분
- 4-4. 대표 글자의 임계 규모
- 4-5. 구조 대표성: 총화 156자의 도출

- 4-6. 자모 대표성: JCI에 의한 세트 평가
- 4-7. 자모 보강: 최종 대표 글자 165자
- 4-8. 대표 글자 165자의 조형 변수 배분

5. 결론

참고문헌

1. 서론

1-1. 연구 배경 및 목적

한글 서체 하나를 제작하는 일은 KS 코드 기준 2,350자, 유니코드 기준 11,172자를 설계하는 작업이며, 통상 3개월에서 6개월 이상의 시간이 소요된다. 이 때문에 서체 디자이너는 모든 글자를 순차적으로 그리지 않고 일부 글자를 먼저 설계한 뒤 나머지를 파생시키는 방식으로 작업한다. 이때 먼저 설계하는 글자를 '대표 글자(representative letterforms)'라 부르며, 어떤 글자들을 대표 글자로 삼느냐에 따라 전체 글자의 조형적 통일성과 작업 효율이 결정된다.¹⁾

대표 글자 선정에는 조형적 기준과 계량적 기준의 두 계열이 있으며, 양자는 상호 보완적이다. 조형적 기준은 어떤 구조의 글자를 함께 검토해야 균형과 비례가 흐트러지지 않는지를 알려 준다. 계량적 기준은 어떤 글자에 설계 자원을 우선 투입해야 실사용 지면의 인상이 개선되는지를 알려 준다.

문제는 문화유산 서체의 복원 연구가 확산되면서 발생한다. 역사 자료의 서체를 복원하고 확장하는 작업에서도 어떤 글자부터 설계할 것인가라는 질문은 동일하게 제기된다. 그러나 이 경우 현대 기준의 대표 글자는 두 가지 이유로 적용될 수 없다. 첫째, 역사 자료에는 아래아(·), 아래아(·), 합용병서 계열 첫달자 등 현대 한글에 존재하지 않는 날자가 사용되며, 이들은 현대 코퍼스의 빈도 조사에 포함되지 않는다. 둘째, 역사 자료는 현존하는 자형의 수가 제한되어 있어 설계자가

임의로 표본을 확대할 수 없다. 즉 역사 자료의 서체 복원에서는 해당 자료 자체의 자형 분포에서 대표 글자를 도출해야 한다.

대표 글자 문제는 자료 해석이 아닌 디자인 의사결정의 문제이다. 착수 시점에 무엇을 먼저 그릴지에 대한 선택은 이후 수천 자의 골격과 무게, 공간 배분을 규정한다. 대표 글자는 이 선택을 명시적 규칙으로 바꾸어 놓은 장치이므로, 대표 글자 목록은 자료의 통계적 요약이 아니라 설계 지침의 성격을 갖는다. 문화유산 서체를 디지털 자산으로 전환하는 작업에서도 마찬가지이다. 이 지침이 부재하면 서체의 조형적 일관성은 서체 디자이너의 숙련도에 전적으로 의존할 수밖에 없다. 본 연구가 대표 글자 도출을 절차로 정식화하려는 이유가 여기에 있다.

그럼에도 현재까지 특정 역사 자료를 대상으로 자형 분포를 계량하고 그로부터 대표 글자를 도출한 연구는 확인되지 않는다. 조선 왕실 한글 자료에 관한 기존 연구는 판독과 주석, 서지적 해제, 서예사적 감식에 집중되어 왔으며, 자형을 조형 단위로 계량한 자료는 축적되지 않았다.

본 연구의 목적은 조선 후기 왕실 한글 필사 자료의 자형 분포를 조형 단위로 계량하고, 빈도-구조-자모 커버리지의 삼중 기준을 순차적으로 적용하는 대표 글자 도출 절차를 정식화하는 데 있다. 이는 착수 시점의 조형 검증 자원을 어디에 먼저 배분할 것인가라는 설계 판단을 재현 가능한 규칙으로 바꾸기 위한 것이다. 이 목적은 다음 다섯 개의 연구 문제로 전개된다.

- 연구 문제 1. 왕실 어필 자료의 날자 사용 빈도는 현대 한글과 어떻게 다른가(4-2절).

1) 이용재, '효율적인 한글 활자디자인을 위한 대표 글자 연구: 글자의 균형과 비례를 중심으로', 글짜씨, Vol.6, 한국타이포그래피학회, 2012. 06. p.1186

• 연구 문제 2. 모임꼴 구조 분포는 현대 한글과 어떻게 다르며, 두 자료 간에는 어떤 차이가 있는가(4-3절).

• 연구 문제 3. 누적 사용량을 기준으로 할 때 대표 글자의 임계 규모는 어느 수준인가(4-4절).

• 연구 문제 4. 빈도를 기준으로 선정한 대표 글자는 구조 대표성을 확보하는가(4-5절).

• 연구 문제 5. 구조적 대표성을 확보한 대표 글자는 자모 수준의 대표성도 확보하는가(4-6, 4-7절).

1-2. 연구 범위 및 방법

연구 대상은 『정조어필한글편지첩』과 『곤전어필』의 한글 본문이다. 두 자료 모두 국립한글박물관 소장 자료총서의 판독문을 저본으로 삼았으며, 본 연구는 판독을 새로 수행하지 않았다. 한문으로 작성된 부분과 필자가 다른 발문은 계수 대상에서 제외하였다. 제외 기준은 3-1절에서 상술한다.

연구 방법은 코퍼스 구축, 낱자 분해 및 모임꼴 분류, 분포 분석, 대표 글자 도출 및 지표 평가의 네 단계로 구성된다. 분석에는 Python 3.10을 사용하였다. 현대 한글과의 비교 준거로는 박재홍(2018)의 21세기 세종계획 빈도 조사 결과를 사용하였으며, 이는 본 연구와 동일하게 음절글자와 낱자를 모두 계량하고 여섯 모임꼴 분류를 채택하고 있어 직접 대조가 가능하다. 본 연구는 서체 생성 모델의 성능이나 복원 결과물의 품질 평가는 다루지 않는다. 연구의 범위는 서체 복원에 선행하는 자료 설계 단계에 한정된다.

2. 이론적 배경

2-1. 한글 서체 대표 글자 선정 기준의 검토와 한계

대표 글자에 대한 논의는 크게 두 계열로 전개되어 왔다. 하나는 조형적 근거에 기반한 계열이다. 김두식(2008)은 활자 제작용 대표 글자 138자와 활자 보기용 대표 글자 116자를 제시하였다.²⁾ 이용제(2012)는 글자의 균형과 비례를 확인할 수 있는 대표 글자 109자를 제안하면서 디자인 단계를 세분화하고 각 단계에 맞는 대표 글자를 배정할 필요성을 지적하였다.³⁾ 다른

하나는 계량적 근거에 기반한 계열이다. 박재홍(2018)은 21세기 세종계획 코퍼스 약 4,500만 어절에서 3,025 종의 음절글자를 채집하여, 누적 사용량의 80%를 점하는 상위 181자를 대표 글자로 제안하였다.⁴⁾

조형적 계열의 논지는 다음과 같다. 이용제(2012)는 서체 디자이너들이 전체 글자의 틀을 잡기 위해 관행적으로 선택하는 대표 글자가 있다고 지적한다. ‘ㅁ’ 당자를 기본으로 삼아 합자한 <마, 맘, 모, 몸, 와, 땀>, <미, 밎, 므, 뎡, 뎡, 뎡> 등이 그것이다. ‘ㅁ’을 기준으로 삼는 이유는 시각 삭제가 없고 글자의 공간 구획을 계획하기 쉽기 때문이다.⁵⁾ 그러나 한글에는 ‘ㅅ’이나 ‘ㅇ’처럼 성격이 다른 당자가 있고, 이들을 상하좌우로 모아쓰면 ‘ㅁ’을 기준으로 설정한 균형과 비례를 유지하기 어렵다.⁶⁾

이 지적은 대표 글자 선정의 두 가지 요건을 시사한다. 하나는 구조 포괄성이다. 대표 글자 집합은 한글의 다양한 모임꼴 구조를 모두 포함해야 한다. 다른 하나는 대비 가능성이다. 좌우 대칭 글자와 비대칭 글자, 밀도가 고른 글자와 고르지 않은 글자를 함께 포함하여 서로 비교하며 설계할 수 있어야 한다. 이용제(2012)는 이 요건에 따라 대표 글자 109자를 제시하였고, 김두식(2008)이 제안한 138자에 대해서는 “138 대표 글자 중에서 어떤 글자부터 디자인해야 하는지 제시하지 않았고, 전체 글자의 균형과 비례 등을 일정하게 유지하기 위해 먼저 고려해야 하는 대표 글자들도 빠져 있다”고 지적한다.⁷⁾ 즉 대표 글자 논의의 관건은 목록의 존재 자체가 아니라 도출하는 기준의 명시성에 있다.

계량적 근거에 기반한 접근은 “빈도가 높은 글자에 설계 시간을 더 투입하는 것이 효율적”이라는 명확한 논리는 갖는다.⁸⁾ 한편 이은하와 남기춘(2020)은 세종말뭉치를 대상으로 한국어 음절의 빈도와 분포를 계량하면서 음절 빈도의 분포와 그 음절을 구성하는 낱자의 분포가 서로 다른 양상을 보인다는 점을 보고하였다.⁹⁾ 이는 음절 단위의 빈도만으로 표집을 설계할 경

4) 박재홍, ‘한글 글자디자인을 위한 21세기 세종계획 한글 빈도수 조사분석’, 조형미디어학, Vol.21, No.3, 한국일러스트아트학회, 2018. 08. p.299

5) 이용제, Op. cit., p.1189

6) 이용제, Ibid., p.1189

7) 이용제, Ibid., p.1191

8) 이용제, Ibid., p.1191

9) 이은하, 남기춘, ‘세종 말뭉치에 나타난 한국어 음절의

2) 김두식, 『한글 글꼴의 역사: 15세기~19세기 활자본 및 목판본을 중심으로』, 시간의 물레, 2008; 이용제, Op. cit., p.1186에서 재인용.

3) 이용제, Op. cit., p.1191-1192

우 낱자 수준의 노출 상태가 통제되지 않는다는 점을 시사한다.

두 계열의 관계를 정리하면 다음과 같다. 조형적 접근은 대표 글자 집합의 내부 구성, 즉 어떤 구조의 글자가 함께 들어 있어야 하는지를 규정한다. 계량적 접근은 대표 글자 집합의 규모, 즉 몇 자까지 설계하면 실사용 효과가 확보되는지를 규정해 글자의 양적 기준을 확보한다. 두 접근 모두 대표 글자를 선정하는 데 필요한 기준을 제시하고 있지만, 빈도순으로 상위 N자를 취하면 특정 모임꼴이 과대 또는 과소 대표될 수 있다는 점에서 두 기준은 충돌할 수 있다.

2-2. 자모 커버리지 지수 참조 세트의 설계 진단 지표

자모 커버리지 지수(Jamo Coverage Index: JCI)는 참조 세트 단위로 낱자 수준의 대표성을 계량하는 기준으로 앞서 살펴본 대표 글자 선정 기준의 한계를 극복하는 방법이다.¹⁰⁾ JCI는 첫당자, 홀자, 받침당자의 세 위치에서 아래 두 항의 값을 곱한 뒤 이를 평균하여 하나의 값으로 종합한다. 하나는 다양성 항으로, 대상 위치의 자모 가운데 몇 종이 세트에 나타나는지를 비율로 나타낸다. 다른 하나는 균등성 항으로, 나타난 자모들이 세트 안에서 고르게 분포하는지를 정규화 엔트로피로 나타낸다.

JCI를 두 항의 곱으로 구성한 이유는 다음과 같다. 다양성 항만 사용하면 특정 자모가 단 한 번만 나타나도 포괄된 것으로 간주되어, 세트가 사실상 편향되어 있어도 값이 높게 나온다. 반대로 균등성 항만 사용하면 소수의 자모만 고르게 분포하는 세트가 과대평가된다. 두 항을 곱하는 구성은 많은 종류의 자모를 고르게 담은 세트에만 높은 값을 부여한다.

선행 연구에서는 세 종류의 참조 세트를 대상으로 개념 수준에서 예시적으로 JCI를 산출되었으며, 관찰된 사례 수가 적어 지표의 타당성을 단언하기 어렵다는 한계가 지적되었다. 본 연구는 이 지표를 역사 자료 코퍼스라는 새로운 대상에 재적용하여, 대표 글자 세트의 평가 도구로서 실효성을 검토한다.

JCI는 참조 세트가 대상 서체의 자모 체계를 얼마나

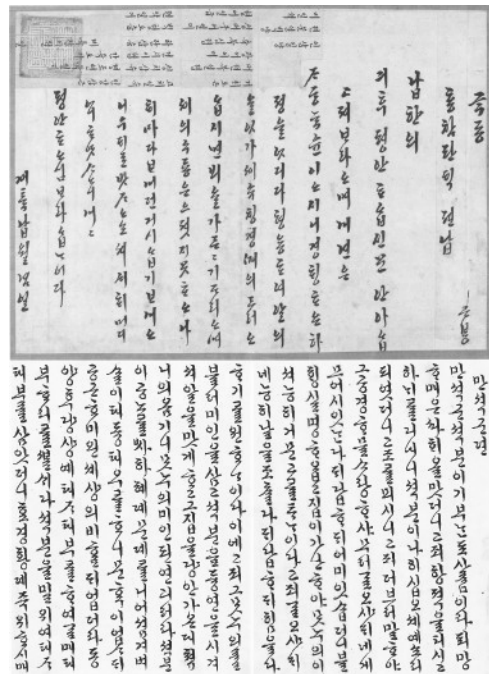
빈도와 분포', 언어과학연구, 제92호, 언어과학회, 2020. pp.92-93

- 10) Hong, J. K., Kim, S. K., 'Few-Shot Learning for Irregular Hangeul Typeface Expansion: A Comparative Study of GAN, VQGAN, and Diffusion Models', Electronics, Vol.15, 2633, 2026. 06. p.11

고르게 노출하는지를 기술하는 입력 구성 지표이다. 따라서 JCI 값은 생성 결과의 조형적 품질을 예측하지 않으며, 두 지표의 관계는 별도의 실험으로 검증해야 한다.

2-3. 연구 대상 자료: 왕실 예필 두 종의 사체적 성격

『정조어필한글편지첩』은 정조가 외국도 여행민씨 등에게 보낸 한글 편지를 모은 첩이다. 예필 2점, 예찰 6점, 어찰 8점 등 총 16점으로 구성되며, 이 가운데 한글 편지는 14통이고 나머지 2점은 한문 자료이다. 2022년 국가지정문화유산 보물 제2167호로 지정되었다. 작성 시기는 원손 시절인 1755년경부터 재위 22년인 1798년까지 걸쳐 있다. 어린 시절의 글씨와 장년기의 글씨가 함께 남아 있어 한 필자의 필적 변화를 관찰할 수 있는 자료로 평가된다.¹¹⁾



[그림 1] 정조어필한글편지첩(위), 곤전어필(아래)¹²⁾

『곤전어필』은 정조의 비 효의왕후 김씨(1753~1821)가 1794년(정조 18)에 필사한 자료이다. 조카 김종선에

- 11) 박정숙, '조선시대 (18세기) 왕실 한글 필사료 자료의 서예적 고찰', 국립한글박물관 소장자료총서, 1, 2014. pp.126-129

- 12) 국립한글박물관, '소장자료총서 1', 국립한글박물관, 2014, pp.38, 58

게 『사기』의 만석군전과 『신당서』의 곽자의전을 한글로 번역하게 한 뒤, 그 번역본을 효의왕후가 직접 필사하였다. 자료에는 효의왕후의 발문과 김기후·김기상의 발문이 함께 실려 있어 필자가 다르다. 「효의왕후 어필 및 함 — 만석군전·곽자의전」이라는 명칭으로 보물 제2087호로 지정되어 있다.¹³⁾

두 자료는 서체 연구의 관점에서 상보적 성격을 갖는다. 『정조어필한글편지첩』은 사적 서간 유형의 자료로 분량은 적으나 필자의 필이가 직접적으로 드러난다. 반면, 『곤전어필』은 전기(傳記) 필사 유형의 자료로 분량이 많고 자형이 정제되어 있다. 두 자료를 함께 분석함으로써 왕실 한글 서체에서 나타나는 개인적 서체와 정제된 서체 양식을 구분하여 관찰할 수 있다.¹⁴⁾

2-4. 역사 자료 서체 복원에서 현대 한글 대표 글자의 부적합성

역사 자료 서체 복원에 현대 기준의 대표 글자를 적용할 수 없는 이유는 한글 표기 체계의 차이에 있다.¹⁵⁾ 현대 한글의 출자는 21자이지만, 조선 후기 한글에는 아래아(·)와 그 결합형인 아래애(·) 등이 사용되었다. 아래아는 단순한 추가 날자가 아니다. 아래아는 수평축 출자로 기능하여 첫달자와의 상하 배열, 즉 세로모임 구조를 형성한다. 따라서 아래아의 사용 빈도가 높은 자료는 모임꼴 구조 분포 자체가 현대와 다르다.¹⁶⁾

이는 서체 디자인 실무에 직접적인 함의를 갖는다. 이용제(2012)는 날자가 대부분 대칭 형태이거나 대칭에 가까워 날자 단위에서는 균형과 비례를 유지하기 쉬우나, 이를 모아쓰면 “대칭이었던 날자의 특징은 세로모임꼴 온자에서만 일부 유지되고, 가로모임꼴과 섞임모임꼴 온자에서는 대칭이 흐트러진다”고 설명한다.¹⁷⁾ 즉 모임꼴별 조형 난이도가 다르므로 대표 글자 집합은 모임꼴을 균형 있게 포함해야 한다. 만약 특정 자료에서 세로모임꼴 비중이 현대보다 현저히 높다면,

현대 기준 대표 글자를 그대로 사용할 경우 세로모임꼴 자형의 조형 검증이 구조적으로 부족해진다.

또한 현대 한글 대표 글자에 대한 계량적 접근의 전제는 역사 자료에는 바로 적용하기 어려워 재검토가 필요하다. 박재홍(2018)은 현대 한글 텍스트에서 실제 출현하는 3,025종의 자형을 모집단 대상으로 하지 만,¹⁸⁾ 역사 자료의 자형 종류는 현전 자료의 분량에 의해 상한이 결정된다. 이 조건에서 상위 80% 커버리지 지점이 몇 자에 해당하는지는 선험적으로 알 수 없으며, 실측을 요구한다.

나아가 역사 자료에서는 자모 체계 자체가 실측 대상이 된다. 현대 한글의 자모 목록은 확정되어 있으나, 특정 역사 자료의 자모 종류와 그 수는 자료마다 다르다. 따라서 자모 수준의 대표성을 평가하려면 해당 코퍼스에서 실제로 관측된 자모 목록을 모집단으로 설정해야 한다. 본 연구는 이 원칙에 따라 JCI의 모집단을 코퍼스 내 관측 자모로 정의한다.

3. 연구 방법

3-1. 서체 연구용 코퍼스의 구축

코퍼스 구축은 판독문 확보, 정규화, 자형 분절, 계수의 순서로 수행하며, 판독문은 국립한글박물관 소장 자료총서를 저본으로 사용하였다. 계수 대상에서는 다음 세 가지 항목을 제외하였다. 첫째, 한문으로 작성된 부분이다. 『정조어필한글편지첩』 16점 중 한문 자료 2점 전체와, 한글 편지 내에 삽입된 한자 표기가 이에 해당한다. 둘째, 필자가 다른 발문이다. 『곤전어필』의 김기후·김기상 발문은 효의왕후의 필적이 아니므로 제외하고, 효의왕후 자신의 발문만 포함하였다. 셋째, 물목 단자의 한자 표기이다.

자형 분절에서 가장 주의를 요하는 지점은 옛한글 처리이다. 유니코드에서 현대한글 음절은 완성형 영역(U+AC00~U+D7A3)에 단일 부호로 존재하지만, 옛한글 음절은 첫가끝 날자(U+1100~U+11FF)의 연쇄로 표현된다. 예컨대 ‘ㅎ’은 U+1112(ㅎ)와 U+111E(·)의 두 부호로 구성된다. 따라서 이를 부호 단위로 계수하면 옛한글 음절이 날자로 분해되어 자형 빈도집계에서 누락되거나 중복 계수된다. 본 연구는 첫가끝 연쇄를 하나의 음절 단위로 병합하여 계수하는 절차를 적용하였다. 이 처리를 통해 ‘ㅎ’, ‘ㄹ’, ‘ㄷ’, ‘읍’ 등이

13) 박정숙, Op. cit., pp.135-137

14) 박정숙, Ibid., pp.135-137

15) 양효정, 임순범, 안병학, ‘고문헌 집자를 통한 옛한글 글꼴 디자인 토대 연구’, 기초조형학연구, 23권 5호, 한국기초조형학회, 2023. 02. pp.153-154

16) 우형식, ‘문자의 보수성과 아래아 표기의 변천’, 한국어문화교육, 15권 2호, 한국어문화교육학회, 2022. 01. pp.50-51

17) 이용제, Op. cit., p.1189

18) 박재홍, Op. cit., p.299

독립 자형으로 집계되었다.

본 연구에서 계수 단위는 판독문상의 음절글자 종이며, 동일 음절글자의 필사상 이체는 구분하지 않았다. 따라서 본 연구의 분포는 조형 변이량이 아니라 설계 대상 목록의 규모를 기술한다.

3-2. 낱자 분해와 모임꼴 분류: 조형 단위의 설정

먼저 수집한 모든 자형을 첫당자, 홀자, 받침당자로 분해하였다. 완성형 음절은 유니코드 산술식으로 분해하고, 첫가끝 연쇄는 부호 영역에 따라 낱자를 판별하였다. 이 절차를 통해 완성형 자형과 옛한글 자형을 동일한 분석 틀에서 다룰 수 있다.

모임꼴 분류는 박재홍(2018)과 이용제(2012)가 공통으로 채택한 여섯 분류 체계를 따랐다. 즉 홀자의 축 방향에 따라 가로모임, 세로모임, 섞임모임으로 나누고, 각각을 받침당자의 유무에 따라 민글자와 받침글자로 세분하여 총 여섯 유형을 구성하였다. 홀자의 축 방향 판별 기준은 [표 1]과 같다.

굵게 표시한 아래아(·)와 아래애(·) 등은 현대 한글에 없는 홀자이다. 아래아를 세로모임 홀자로 분류한 근거는 다음과 같다. 아래아는 자형상 점 또는 짧은 획으로 실현되며 첫당자의 아래에 놓여 상하 배열을 형성한다. 이는 ‘ㅡ’, ‘ㄴ’, ‘ㄱ’가 형성하는 배열과 동일한 구조적 성격을 가진다. 아래애(·)는 아래아(·)에 ‘ㅣ’가 결합한 형태로 상하와 좌우가 혼합되므로 섞임모임으로 분류하였다. 한편 본 코퍼스에서는 [표 1]에 열거한 23종 외에 ‘ㅞ’, ‘ㅟ’ 등 2 종의 옛한글 복합 홀자가 관측되었으며 이들 역시 동일한 축 방향 판별 원칙에 따라 각각 섞임모임과 세로모임으로 귀속하였다.

[표 1] 모임꼴 분류 기준

모임꼴	홀자	구조	예시
가로모임	ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅝ ㅡ	첫당자와 홀자의 좌우 배열	마, 맘
세로모임	ㅓ ㅕ ㅗ ㅛ ㅜ ㅝ ㅞ ㅟ	첫당자와 홀자의 상하 배열	모, 몸
섞임모임	ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅝ ㅞ ㅟ	좌우와 상하의 혼합 배열	와, 맘

이 분류 절차는 방법론적 이점을 갖는다. 자형을 표기 형태 그대로 유형화하는 방식에서는 ‘ㅟ’, ‘ㅞ’, ‘ㅟ’ 등 옛한글 복합 홀자 자형이 세 유형 중 어디에도 명확히 속하지 않아 잔여 범주가 발생한다. 하지만 낱자

단위로 분해하여 홀자의 축 방향을 판별하는 방식에서는 이들이 섞임모임으로 일관되게 귀속되며, 잔여 범주가 소멸한다.

3-3. 분석 지표: 설계 판단을 위한 다섯 척도

본 연구는 자형 분포를 서체 설계의 판단 근거로 전환하기 위해 다섯 개의 지표를 사용하였다. 앞의 세 지표는 코퍼스의 성격과 복원 난이도를 기술하고, 뒤의 두 지표는 도출된 대표 글자 집합이 원자료를 얼마나 충실히 대리하는지를 평가한다. 각 지표의 정의와 서체 디자인상의 해석은 다음과 같다.

첫째, 유형-토큰 비율(TTR)은 고유 음절글자 종수를 총 토큰 수로 나눈 값이다. 값이 클수록 자형이 반복되지 않고 다양하게 사용되었음을 뜻한다. 둘째, 1회 출현 자형 비율은 전체 자료에서 한 번만 나타난 자형이 고유 자형 전체에서 차지하는 비율이다. 서체 복원의 관점에서 이 비율은 참조 가능한 자형 사례가 단 하나뿐인 경우의 비중을 뜻하므로, 복원 난이도의 지표로 해석할 수 있다. 셋째, 누적 커버리지는 빈도 상위 N개 자형이 총 토큰에서 차지하는 누적 비율이다. 대표 글자의 규모를 결정하는 직접적 근거가 된다. 넷째, 모임꼴 구성 편차는 대표 글자 집합의 여섯 모임꼴 구성 비율과 전체 코퍼스의 여섯 모임꼴 토큰 비중 사이의 차이를 백분율 포인트로 나타낸 값이다. 이 값이 작을수록 해당 집합이 전체 코퍼스의 구조를 잘 대표한다. 다섯째, JCI는 대표 글자 집합이 코퍼스의 자모 체계를 얼마나 고르게 담고 있는지를 나타내는 지표이다. 위치 p 를 첫당자, 홀자, 받침당자로 두고 다음과 같이 산출한다.

$$D_p = \frac{K_p}{N_p}, \quad E_p = \frac{H_p}{\log K_p},$$

$$H_p = - \sum_{i=1}^{K_p} f_i \log f_i,$$

$$JCI = \frac{1}{3} \sum_p D_p E_p$$

여기서 N_p 는 코퍼스 전체에서 관측된 위치 p 의 자모 종류 수이고, K_p 는 평가 대상 집합에 나타난 위치 p 의 자모 종류 수이다. f_i 는 집합 내에서 자모 i 가 차지하는 상대 빈도이며 $\sum_i f_i = 1$ 로 정규화한다. 따

라서 D_p 는 자모 종류의 포괄률을, E_p 는 포괄된 자모들의 분포 균등도를 나타낸다. $K_p = 1$ 인 경우 $E_p = 0$ 으로 정의한다.

본 연구는 이 지표를 산출할 때 다음 세 가지 조건을 명시한다. 첫째, N_p 는 현대 한글의 자모 목록이 아니라 본 코퍼스에서 관측된 자모 목록으로 설정한다. 역사 자료의 자모 체계는 자료마다 다르므로, 해당 자료 내부를 모집단으로 삼는 것이 타당하다. 둘째, 받침당자의 N_p 는 받침글자에 있는 글자에 한하여 산출하며 받침 없음은 범주에서 제외한다. 받침 없음을 하나의 범주로 포함할 경우 그 압도적 비중이 균등성 향을 왜곡할 수 있다. 셋째, f_i 는 자형 단위로 계수한다. 즉 집합에 포함된 각 자형이 자신의 자모를 한 번씩 기여하며, 코퍼스 내 출현 빈도로 가중하지 않는다. 대표 글자 집합은 설계 대상의 목록이므로 목록 내에서 각 자형의 지위가 동등하다고 보는 것이 지표의 목적에 부합한다.

4. 분석 결과

4-1. 코퍼스 규모와 자형 분포: 복원 난이도의 계량

두 어필 자료의 성격 차이는 지표에 뚜렷하게 나타난다. 『정조어필한글편지첩』(이하 정조어필)은 유형-토큰 비율이 0.149, 1회 출현 자형 비율 31.6%로, 자형이 적게 반복되고 다양하게 분산된 소규모 자료이다. 사적 서간이라는 유형의 특성상 인명, 물목명, 안부 표현 등 어휘가 폭넓게 등장하기 때문이다. 반면 『곤전어필』은 유형-토큰 비율이 0.057, 1회 출현 자형 비율 20.0%로 반복률이 높다. 전기 필사라는 유형에서 문법 형태소와 상용 한자어가 반복적으로 출현하기 때문이다.

이 차이는 서체 복원의 관점에서 실천적 함의를 갖는다. 정조어필에서는 고유 자형의 3분의 1가량이 단 한 번만 출현하므로, 해당 자형의 조형 판단에 활용할 수 있는 사례가 하나뿐이다. 반면 곤전어필은 대부분의 자형에 복수의 사례가 존재하여 자형의 변이 폭을 관찰할 수 있다. 즉 전자는 필자의 개인 양식을 특정하는데, 후자는 자형의 일반형을 확보하는데 적합하다. 구축된 코퍼스의 개요는 [표 2]와 같다.

옛한글 자형은 통합 기준 107자로 전체 고유 자형의 15.9%를 차지하며, 토큰 기준으로는 18.3%에 이른다. 이는 현대 완성형 부호계로는 코퍼스의 약 6분

의 1을 표현할 수 없음을 뜻하며, 서체 복원 시 옛한글 자형 설계가 선택 사항이 아니라 필수 요건임을 보여 준다.

[표 2] 왕실 어필 코퍼스 개요

구분	정조어필	곤전어필	통합
총 토큰 수	1,843	11,484	13,327
고유 음절글자 종수	275	649	674
유형-토큰 비율	0.149	0.057	0.051
1회 출현 자형	87자 (31.6%)	130자 (20.0%)	125자 (18.5%)
옛한글 자형 수	29자 (10.5%)	102자 (15.7%)	107자 (15.9%)
옛한글 토큰 비중	15.8%	18.7%	18.3%

코퍼스 내 관측 자모는 첫당자 19종, 홀자 25종, 받침당자 9종으로 집계되었다. 종수만 보면 현대 한글의 19종, 21종, 27종과 비슷하거나 적으나, 구성이 다르다. 홀자에는 아래아와 아래애 등 옛한글 홀자 4종이 추가되며, 첫당자에는 합용병서 계열이 포함되는 대신 현대의 각자병서 일부가 나타나지 않는다. 받침당자는 9종에 불과하여 현대 27종에 비해 단순하다. 이 관측 자모 목록은 이후 JCI 산출의 모집단이 된다.

4-2. 낱자 사용 빈도: 기준 낱자의 재설정

본 연구에 사용하는 낱자 용어는 한글 조형 연구의 관례에 따른다. 낱자는 음절글자를 이루는 최소 단위로 놓이는 자리에 따라 첫당자, 홀자, 받침당자로 구분된다. 이 가운데 홀자는 음절글자의 축을 결정하는 낱자로 세로 방향의 기동과 가로 방향의 보, 그리고 이들에 붙는 결줄기의 조합으로 실현된다. 홀자가 어떤 축을 취하는지에 따라 첫 당자와 받침당자가 놓일 공간의 위치와 크기가 함께 규정되므로, 홀자는 모임꼴을 결정하는 낱자이면서 글자의 비례를 좌우하는 낱자이기도 하다.¹⁹⁾ 따라서 낱자 사용 빈도를 검토할 때 홀자를 먼저 다루는 것은 어떤 홀자를 기준으로 골격을 세울 것인지를 판단하기 위한 절차적 선택이다.

낱자 사용 빈도를 박재홍(2018)의 현대 한글 조사 결과와 대조하였다. 홀자 사용 빈도 비교 결과는 [표 3]과 [그림 2]와 같다.

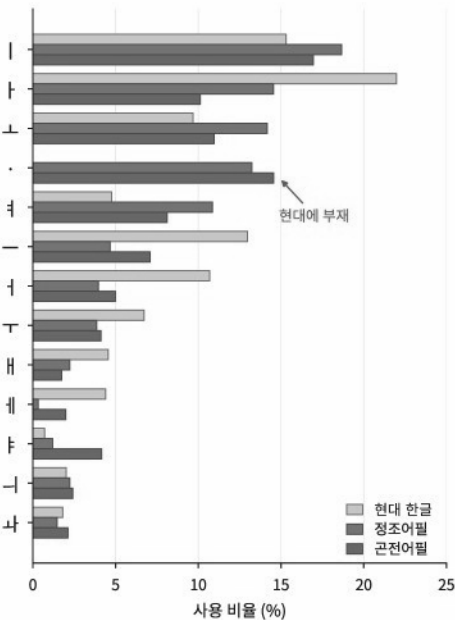
19) 안상수, 한재준, 이용제, 『한글 디자인 교과서』, 안그래픽스, 2009, pp.280~281

[표 3]은 세 자료의 흘자 사용 빈도를 상위 10위까지 병렬한 것으로 각 열은 해당 자료에서 그 흘자가 전체 흘자 토큰 가운데 차지하는 비율이다. 순위 구성과 비율의 절대값을 함께 대조하면 왕실 어필의 흘자 사용 방식은 단순한 빈도 차이를 넘어 현대 한글과 세 가지 지점에서 뚜렷하게 다르다는 것을 알 수 있다. 이는 서체 설계 시 어떤 흘자를 기준으로 골격을 잡아야 하는지에 영향을 미친다.

[표 3] 흘자 사용 빈도 비교

(단위 %)

순위	현대 한글	정조어필	곤전어필
1	ㅏ 21.98	ㅣ 18.67	ㅣ 16.95
2	ㅣ 15.31	ㅏ 14.54	ㅂ 14.55
3	ㅡ 12.97	ㅓ 14.16	ㅓ 10.96
4	ㅑ 10.69	ㅕ 13.24	ㅏ 10.12
5	ㅓ 9.68	ㅑ 10.85	ㅑ 8.12
6	ㅓ 6.72	ㅡ 4.67	ㅡ 7.09
7	ㅑ 4.75	ㅑ 3.96	ㅑ 4.99
8	ㅑ 4.55	ㅓ 3.85	ㅏ 4.15
9	ㅑ 4.40	ㅑ 2.50	ㅓ 4.12
10	ㅑ 2.02	ㅑ 2.22	ㅑ 2.92



[그림 2] 흘자 사용 빈도 비교

첫째, 아래아가 정조어필에서는 13.24%로 4위, 곤

전어필에서는 14.55%로 2위를 차지한다. 두 값 모두 현대 한글의 3위 흘자인 ‘ㅏ’의 12.97%를 상회한다. 즉 현대 코퍼스의 빈도 조사에 원리적으로 포함될 수 없는 날자가 왕실 어필에서는 최상위 사용 빈도를 갖는다. 여기에 아래애의 2.50~2.92%를 더하면 현대에 부재하는 흘자의 비중은 15.7~17.5%에 달한다. 이는 현대 한글 기준의 대표 글자 목록이 역사 자료 복원에 부적합함을 보여 준다.

둘째, ‘ㅏ’와 ‘ㅣ’의 순위가 역전된다. 현대 한글에서 ‘ㅏ’가 21.98%로 압도적으로 가장 높은 빈도를 보이지만, 어필 자료에서는 ‘ㅣ’가 가장 높은 빈도를 보이며, ‘ㅏ’는 14.54%, 10.12%로 하락한다. 현대의 서체 디자인 실무에서는 ‘ㅏ’가 기동과 결줄기의 관계를 정하는 기준 흘자로 다루어진다. 하지만 어필 서체 복원에서는 ‘ㅣ’, 즉 결줄기가 없는 기동의 조형을 먼저 확정하는 것이 자료 실태에 부합한다.

셋째, 현대 한글에서 4.75%의 비중을 차지하는 ‘ㅑ’가 어필에서는 8.12~10.85%로 두 배 이상 상승하는 반면, 현대 한글에서 12.97%의 빈도를 차지하는 ‘ㅡ’는 정조어필에서는 4.67%로 급감한다. ‘ㅑ’의 경우 <여>, <며>, <데> 등 겹흘자 표기가 빈발함에 따라 높은 빈도를 차지하는 것으로 보인다. 이에 겹줄기 흘자의 간격 처리가 어필 서체에서 상대적으로 중요한 과제를 알 수 있다.

첫당자와 받침당자의 비교 결과는 흘자만큼의 큰 차이는 없이 상위 날자의 구성을 대체로 공유한다. 하지만 중위 빈도에서 문헌 유형에 따른 편차가 뚜렷하게 드러나는 양상을 보인다. 상세 내용은 [표 4]와 같다.

[표 4] 첫당자·받침당자 사용 빈도 비교

(상위 6위, 단위: %)

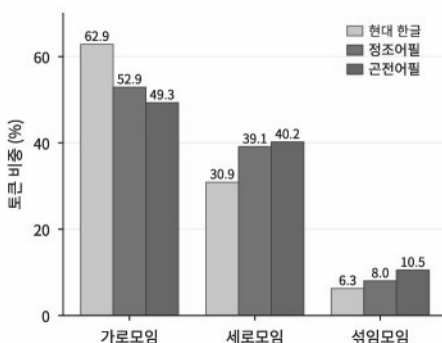
순위	첫당자			받침당자		
	현대 한글	정조어필	곤전어필	현대 한글	정조어필	곤전어필
1	ㅇ 23.91	ㅇ 29.30	ㅇ 21.69	없음 54.17	없음 58.11	없음 62.38
2	ㄱ 13.14	ㅅ 13.13	ㄱ 11.56	ㄴ 15.09	ㄴ 11.83	ㄴ 10.85
3	ㄷ 9.19	ㄱ 10.31	ㅎ 11.49	ㄹ 8.86	ㅇ 7.92	ㅇ 9.22
4	ㅅ 8.45	ㅎ 8.52	ㅅ 10.51	ㅇ 6.89	ㅂ 6.95	ㄹ 8.25
5	ㅈ 8.34	ㄴ 8.46	ㄷ 9.41	ㄱ 4.61	ㄹ 6.19	ㄱ 4.00
6	ㅎ 7.07	ㄷ 8.03	ㄹ 7.35	ㅁ 2.83	ㄱ 3.91	ㅁ 2.26

세 자료 모두 첫당자에서는 ‘ㅇ’이 가장 높은 빈도를 차지한다. 구체적인 수치에서는 차이가 있는데, 정조어필의 29.30%는 현대 한글의 23.91%보다 5.4%p 높다. 편지에서 어미와 조사, 그리고 <읍> 계열의 경양 표현이 반복된 결과이다. 현대 한글과 빈도수의 차이가 있는 첫당자들도 있다. ‘ㅅ’은 현대 한글에서는 4위를 차지하나 정조어필에서는 2위로 상승한다. 현대 한글에서 6위를 차지하는 ‘ㅎ’은 곤전어필에서는 3위로 상승한다. 반면 ‘ㄹ’은 정조어필에서 2.50%로 순위 밖에 머무는 데 비해 곤전어필에서는 7.35%로 6위를 차지하여, 두 자료 사이의 편차가 가장 큰 첫당자로 나타난다.

받침당자에서 주목해야 할 결과는 정조어필의 ‘ㄴ’이다. 현대 한글에서 ‘ㄴ’ 받침은 상위 6위 이내에 들지 않지만, 정조어필에서는 6.95%로 4위를 차지한다. 이는 <읍>, <습> 등 경양법 어미 표기가 편지 유형에서 집중적으로 사용된 결과이다. 즉 문헌 유형의 특성이 받침당자 분포에 직접 반영된 것으로 볼 수 있다. 또한 현대 한글에서 2.52%로 7위를 차지하는 ‘ㅃ’ 받침은 두 어필 자료에 전무하며, 겹받침도 곤전어필에서 ‘ㄹ’ 0.05%, ‘ㄷ’ 0.01%로 사실상 나타나지 않는다. 받침당자의 종류가 현대보다 단순하다는 점은 서체 복원에서 설계 부담을 줄이는 요인으로 작용하지만, 뒤에서 밝혀듯이 소수의 회소 받침이 자모 커버리지에서는 중요한 변수가 된다. 따라서 어필 서체의 기준 낱자는 첫당자 ‘ㅇ’, 출자 ‘ㅣ’와 아래아(·)로 재설정된다.

4-3. 모임꼴 구조 분포: 조형 검증 자료의 재배분

모임꼴 구조 분포는 현대 한글과 두 어필에서 가장 뚜렷한 차이가 확인된 영역이다. 세 자료의 모임꼴 토큰 비중은 [그림 3]과 [표 5]와 같다.



[그림 3] 모임꼴 구조 분포 비교

[표 5] 모임꼴 구조 분포 비교

(토큰 비중, 단위: %)

모임꼴	현대 한글	정조어필	곤전어필
가로모임	62.9	52.9	49.3
세로모임	30.9	39.1	40.2
석임모임	6.3	8.0	10.5

어필 자료의 세로모임 분포는 39.1~40.2%로, 현대 한글의 30.9%보다 8.2~9.3%p 높고, 석임모임 역시 현대 한글보다 1.7~4.2%p 높다. 반대로 가로모임 분포는 현대 한글보다 10.0~13.6%p 낮다.

이 차이의 원인은 아래아이다. 아래아는 세로모임 홀자로 가능하며 그 비중은 13.24~14.55%에 이른다. 아래아를 포함하면 현대에 없는 출자가 세로모임과 석임모임에 기여하는 몫은 15.7~17.5%에 이른다. 다만 세로모임과 석임모임의 실제 증가폭 합계는 9.9~13.5%p로 이보다 작는데, 이는 현대 한글에서 세로모임 홀자로 높은 비중을 차지하는 ‘ㅡ’가 정조어필에서는 8.3%p, 곤전어필에서는 5.9%p 감소하여 아래아의 기여를 부분적으로 상쇄한 결과이다. 즉 왕실 어필의 모임꼴 분포가 현대 한글과 다른 것은 어휘나 문체의 차이가 아니라 표기 체계의 차이에서 비롯된 구조적 결과이다.

이 결과는 서체 디자인 실무에 구체적인 함의를 준다. 이용제(2012)가 지적한 대로 낱자의 대칭성은 세로모임꼴 온자에서만 부분적으로 유지되고 가로모임꼴과 석임모임꼴에서는 흐트러진다. 어필 자료는 세로모임 비중이 현대보다 9%p 가까이 높고 석임모임 비중도 높으므로, 조형적 대칭 처리와 비대칭 보정이 요구되는 자형의 실사용 비중 구성이 현대와 상당히 다르다. 따라서 현대 한글 빈도를 근거로 도출된 대표 글자 목록을 그대로 사용하면 세로모임꼴과 석임모임꼴 자형의 검증 기회가 실제 필요보다 적게 배분된다. 여섯 모임꼴 세부 분포는 [표 6]과 같다.

민글자와 받침글자의 비율은 정조어필에서 58.11% 대 41.89%, 곤전어필에서는 62.38% 대 37.62%로, 현대 한글의 54.17% 대 45.83%와 비교하면 받침글자에 비해 민글자 비중이 높다. 자형 수와 토큰 비중의 관계도 주목할 만하다. 가로모임 받침글자는 자형 수가 224자로 가장 많으나 토큰 비중은 18.56~24.74%에 그친다. 반대로 세로모임 민글자는 자형 수가 74자에 불과하나 토큰 비중은 약 24%에 이른다. 즉 세로모임 민글자는 적은 수의 자형이 높은 빈도로 반복되는 유형으로, 서체 인상에 미치는 영향 대비 설계 부담이 낮다. 이는 설계 순서를 정할 때 우선순위를 부여할 근거가 된다.

[표 6] 여섯 모임꼴 분포

(단위: %, 개)

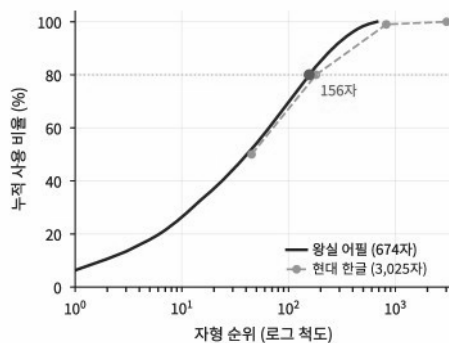
모임꼴	정조 토큰	곤전 토큰	통합 자형 수
가로모임 민글자	28.16	30.76	97
가로모임 받침글자	24.74	18.56	224
세로모임 민글자	24.31	23.86	74
세로모임 받침글자	14.76	16.33	187
섞임모임 민글자	5.64	7.76	55
섞임모임 받침글자	2.39	2.73	37
계	100.00	100.00	674

4-4. 대표 글자의 임계 규모

통합 코퍼스의 누적 커버리지를 산출한 결과는 [표 7]과 [그림 4]와 같다. 두 자료의 모집단 규모는 4.5배 차이가 나고 토큰 규모는 3,000배 이상 차이가 난다. 그럼에도 50% 커버리지 지점은 41자와 45자, 80% 지점은 156자와 181자로 근사하다. 반면 99% 지점은 543자와 822자로 격차가 벌어진다.

[표 7] 누적 커버리지 비교

누적 커버리지	왕실 어필 (모집단 674자)	현대 한글 (모집단 3,025자)
50%	41자	45자
80%	156자	181자
90%	260자	-
95%	363자	-
99%	543자	822자



[그림 4] 누적 커버리지 곡선 비교

이 결과는 두 가지로 해석된다. 첫째, 분포의 상단, 즉 고빈도 영역의 구조는 코퍼스의 규모나 시대와 무

관하게 유사하다. 두 분포 모두 소수의 자형이 사용량을 지배하는 지프(Zipfian) 분포의 형태를 따르며, 그 기울기가 유사하다. 둘째, 분포의 하단, 즉 저빈도 꼬리는 코퍼스 규모에 의해 제약된다. 어필 코퍼스의 꼬리가 짧은 것은 왕실 한글의 자형 종류가 적었기 때문이 아니라 현전 자료의 분량이 제한적이기 때문이다.

또한, 이 결과는 대표 글자 논의에 다음과 같은 함의를 갖는다. 기존에 제안된 대표 글자의 규모는 김두식(2008) 138자, 박재홍(2018) 181자, 이용제(2012) 109자로, 모두 100자에서 200자 사이에 분포한다. 본 연구에서 산출한 80% 커버리지 지점 156자 역시 이 범위에 속한다. 조형적 근거에서 출발한 논의와 계량적 근거에서 출발한 논의, 그리고 현대 한글과 18세기 왕실 한글이라는 상이한 대상이 모두 유사한 규모로 수렴한다는 점은, 대표 글자의 임계 규모가 한글 모아쓰기 구조 자체에서 비롯되는 상수일 가능성을 시사한다. 다만 본 연구의 자료만으로 이를 단정할 수는 없으며, 추가 자료의 검증을 요한다.

4-5. 구조 대표성: 총화 156자의 도출

빈도만을 기준으로 상위 156자를 취할 경우 구조 대표성이 확보되는지 검토한 결과는 [표 8]과 [그림 5]과 같다.

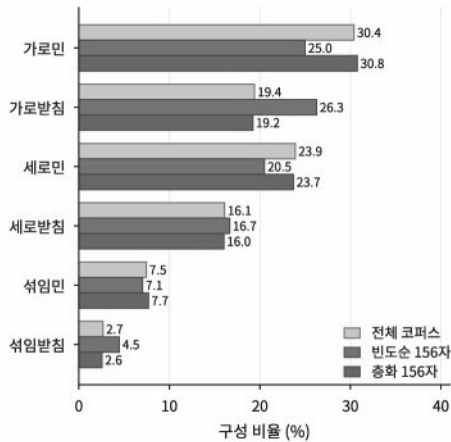
[표 8] 여섯 모임꼴 구성 비교

(단위: %)

모임꼴	전체 코퍼스 (토큰 비중)	빈도순 156자	편차	총화 156자	편차
가로모임 민글자	30.4	25.0	-5.4	30.8	+0.4
가로모임 받침글자	19.4	26.3	+6.9	19.2	-0.2
세로모임 민글자	23.9	20.5	-3.4	23.7	-0.2
세로모임 받침글자	16.1	16.7	+0.6	16.0	-0.1
섞임모임 민글자	7.5	7.1	-0.4	7.7	+0.2
섞임모임 받침글자	2.7	4.5	+1.8	2.6	-0.1
최대 절대 편차	-	-	6.9	-	0.4
토큰 커버리지	-	80.0	-	79.3	-

빈도순 156자는 가로모임 받침글자를 6.9%p 과대 대표하고 가로모임 민글자를 5.4%p 과소 대표한다. 최대 절대 편차는 6.9%p이다. 이 왜곡이 발생하는 이

유는 받침글자의 자형 종류가 민글자보다 많아, 상위 구간에 받침글자가 상대적으로 많이 포함되기 때문이다. 결과적으로 순수 빈도 기준의 대표 글자 집합은 이용제(2012)가 요구한 조건, 즉 전체 글자의 균형과 비례를 확인할 수 있어야 한다는 조건을 충분히 충족하지 못한다.²⁰⁾



[그림 5] 여섯 모임꼴 구성 비교: 빈도순 대표와 총화 대표

이에 본 연구는 빈도와 구조 대표성을 결합한 표집 절차를 적용하였다. 절차는 다음과 같다. 첫째, 전체 코퍼스의 여섯 모임꼴 토큰 비중을 산출한다. 둘째, 목표 집합 크기에 이 비중을 곱하여 모임꼴별 할당 수를 결정한다. 셋째, 각 모임꼴 내부에서 빈도 상위순으로 할당 수만큼 자형을 선정한다.

이 절차로 도출한 총화 156자는 최대 절대 편차를 0.4%p로 축소하였다. 토큰 커버리지는 79.3%로, 빈도순 집합의 80.0%보다 0.7%p 낮다. 즉 0.7%p의 커버리지를 양보하여 구조 편차를 6.9%p에서 0.4%p로, 약 17분의 1 수준으로 줄인 것이다. 빈도순 집합과 총화 집합은 141자를 공유하며, 총화 과정에서 15자가 교체되었다. 새로 편입된 자형은 <겨, 니, ㄱ, 데, 비, 파, 뉴, 쇼, 사, 게, 두, ㅁ, 허, 바, 서>로, 대부분 가로모임 민글자와 세로모임 민글자에 속한다. 총화 156자의 모임꼴별 할당과 옛한글 포함 현황은 [표 9]와 같다.

[표 9] 총화 대표 글자 156자의 구성

모임꼴	할당 자형 수	코퍼스 내 가용 자형 수	옛한글 자형 수
가로모임 민글자	48	97	1
가로모임 받침글자	30	224	0
세로모임 민글자	37	74	10
세로모임 받침글자	25	187	7
섞임모임 민글자	12	55	5
섞임모임 받침글자	4	37	2
계	156	674	25

4-6. 자모 대표성: JCI에 의한 세트 평가

앞에서 도출한 두 156자 세트가 자모 수준의 대표성까지 확보하는지 JCI로 평가하였다. 결과는 [표 10]과 같다. 여기서 각 위치의 값(DE)은 D_p 와 E_p 의 곱이다.

가장 주목할 결과는 빈도순 156자와 총화 156자의 JCI가 0.691과 0.689로 유사하다는 점이다. 두 세트의 자모 포괄 중수도 완전히 같다. 첫당자 15종/19종, 홀자 21종/25종, 받침당자 7종/9종이다. 즉 모임꼴 총화는 구조 대표성을 크게 개선하면서도 자모 대표성에는 유의한 개선을 가져오지 않았다.

[표 10] 대표 글자 세트의 자모 커버리지 지수

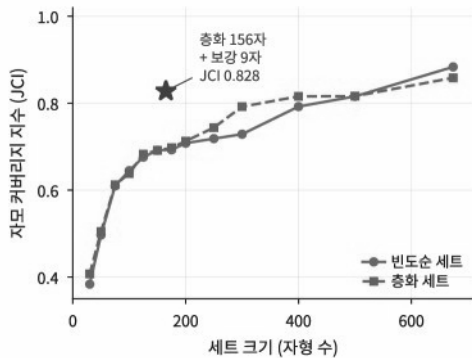
세트	자형 수	토큰 커버리지 (%)	첫당자 (DE)	홀자 (DE)	받침당자 (DE)	JCI
전체 코퍼스	674	100.0%	0.907	0.869	0.874	0.884
빈도순 156자	156	80.0%	0.707	0.740	0.626	0.691
총화 156자	156	79.3%	0.709	0.747	0.610	0.689
빈도순 50자	50	54.1%	0.412	0.609	0.470	0.497
빈도순 100자	100	69.7%	0.585	0.714	0.636	0.645
빈도순 200자	200	85.2%	0.707	0.744	0.673	0.708
빈도순 400자	400	96.2%	0.803	0.777	0.797	0.792

이 결과는 각각 실무적, 이론적으로 해석할 수 있다. 실무적으로는 총화 절차가 자모 대표성을 희생하지 않으므로 안전하게 채택할 수 있다는 것을 의미한다. 또한 이론적으로는 더 중요한 함의를 갖는다. 모임꼴 구성 편차와 JCI는 서로 독립적인 기준이라는 사실이 확인된 것이다. 모임꼴은 홀자의 축 방향과 받침의 유무라는 두 개의 이항적 속성으로 정의되므로, 여섯 유형의 비례를 맞추는 일은 개별 자모의 포괄과 논리적으

20) 이용제, Op. cit., pp.1190-1191

로 무관하다. 예컨대 가로모임 민글자를 48자 확보했다 하더라도 그 48자가 모두 소수의 첫당자로만 이루어져 있을 수 있다. 따라서 대표 글자 도출에서 구조 기준과 자모 기준은 순차적으로, 그리고 별도로 적용되어야 한다. 이는 이용제(2012)가 요구한 조형적 요건과 박재홍(2018)이 제시한 계량적 요건에 더해, 제3의 독립 기준이 필요하다는 뜻이다.

한편 세트 크기와 JCI의 관계는 [그림 6]과 같다. JCI는 30자 수준에서 0.38~0.41로 낮고, 100자 부근까지 급격히 상승한 뒤 150자에서 200자 사이 구간에서 0.69~0.71로 완만해진다. 이후 400자에 이르러도 0.79~0.82에 머문다. 즉 자형 수를 두 배 이상 늘려도 JCI는 0.1 정도만 상승한다. 커버리지 곡선이 156자 부근에서 완만해지는 것과 마찬가지로 JCI 역시 같은 구간에서 수확 체감을 보인다는 점은, 대표 글자의 임계 규모가 두 기준에서 공통적으로 나타남을 뜻한다.



[그림 6] 세트 크기에 따른 JCI

그러나 [그림 6]은 동시에 한계를 드러낸다. 세트를 400자까지 확대해도 JCI가 0.8 수준에 머무는 것은, 희소 자모가 저빈도 꼬리에 몰려 있어 빈도 기반 확장으로는 포괄되지 않기 때문이다. 세트 크기를 늘리는 방식은 자모 커버리지 문제의 해법이 되지 못한다.

다만 JCI 값의 상승이 곧 복원 결과물의 조형적 우수성을 함의하지는 않는다. 본 절의 결과는 세트 구성의 차이를 계량한 것이며, 구성의 차이가 결과의 차이로 이어지는지는 후속 검증의 대상이다.

4-7. 자모 보강: 최종 대표 글자 165자

중화 156자에서 포괄되지 않은 자모를 확인한 결과

는 [표 11]과 같다.

[표 11] 중화 156자의 미포괄 자모와 보강 자형

위치	미포괄 자모	종수	보강 자형 (코퍼스 내 최빈)
첫당자	ㅋ, ㅁ, ㅅ, ㅈ	4	꽤, 쫓, 쑤, 섹
홀자	ㅞ, ㅟ, ㅠ, ㅡ, ㅢ	4	재, 쾌, 꺾, 외
받침당자	ㄹ, ㄼ	2	늪, 뵤

※ ‘꽤’는 첫당자 ‘ㅋ’과 홀자 ‘ㅞ’를 겸하므로 실제 보강 자형은 9자이다.

코퍼스 내 출현 빈도가 극히 낮은 자모 10종은 포괄하지 않았다. 이들을 담은 자형의 빈도는 1회에서 5회 사이로 빈도 상위 156자 또는 400자 구간에 포함될 가능성이 원리적으로 낮다. 하지만 서체 설계의 관점에서 이들은 반드시 필요하다. 합용병서 계열 첫당자는 어필 서체의 시대적 표징이며, 겹받침 ‘ㄹ’과 ‘ㄼ’은 받침 공간 안에서 두 날자를 배분하는 조형 규칙을 결정하는 유일한 근거이다. 이 자형이 설계 대상에서 빠지면 해당 조형 규칙 자체를 세울 수 없다는 문제가 발생한다.

이에 미포괄 자모를 최소한의 자형으로 포괄하는 보강 절차를 적용하였다. 절차는 각 단계에서 미포괄 자모를 가장 많이 동시에 담고 있는 자형을 선택하고, 동수인 경우 코퍼스 내 빈도가 높은 자형을 선택하는 탐욕적 방식이다. 그 결과 <꽤, 꺾, 쑤, 늪, 재, 섹, 뵤, 외, 쫓>의 9자로 10종의 미포괄 자모를 모두 포괄할 수 있었다. <꽤>가 첫당자 ‘ㅋ’과 홀자 ‘ㅞ’를 동시에 담아 한 자를 절약한 결과이다.

[표 12] 최종 대표 글자 165자의 세 기준 평가

기준	빈도순 156자	중화 156자	중화+보강 165자	빈도순 400자
토큰 커버리지	80.0%	79.3%	79.5%	96.2%
모임꼴 최대 편차	6.9%p	0.4%p	1.6%p	2.2%p
첫당자 포괄	15/19	15/19	19/19	17/19
홀자 포괄	21/25	21/25	25/25	22/25
받침당자 포괄	7/9	7/9	9/9	8/9
JCI	0.691	0.689	0.828	0.792

중화 156자에 이 9자를 더한 165자 세트의 성능은 [표 12]와 같다. 165자 세트는 세 기준을 동시에 충족한다. 토큰 커버리지는 79.5%로 중화 156자보다 오

히려 0.2%p 상승하였다. 모임꼴 최대 편차는 1.6%p로 총화 156자의 0.4%p보다 확대되었으나, 빈도순 156자의 6.9%p에 비하면 4분의 1 수준이며 실무적으로 수용 가능한 범위이다. 편차 확대는 보강 자형 9자 중 3자가 석임모임 민글자에 속해 그 비중이 총화 156자의 7.7%에서 165자 세트의 9.1%로 높아진 데 따른 것이다.

가장 큰 개선은 자모 커버리지에서 나타난다. 165자 세트는 코퍼스의 모든 자모를 전수 포괄하여 세 위치 모두 다양성 항 ($D_p = 1.0$)을 달성하였고, JCI는 0.689에서 0.828로 0.139 상승하였다. 이는 자형 수를 5.8% 늘려 얻은 결과이다. 주목할 점은 이 값이 빈도순 400자 세트의 JCI 0.792를 상회한다는 사실이다. 빈도순 400자는 세트 규모가 2.4배에 이르지만 여전히 첫당자 2종, 홀자 3종, 받침당자 1종을 포괄하지 못한다. 즉 빈도 기반으로 세트를 확대하는 것보다, 미포괄 자모를 표적 보강하는 것이 자모 대표성 확보에 훨씬 효율적이다. [그림 6]에서 165자 지점이 두 곡선 위로 뚜렷하게 벗어나 있는 것을 시각적으로 보여 준다.

최종 도출된 165자의 구성은 [표 13]과 같다. 총별 자형 수는 코퍼스 내 각 총의 토큰 비중에 비례하여 배분되었으며 옛한글 자형 25자가 세로모임과 석임모임에 집중된 것은 아래야 계열이 두 모임꼴에 편중되어 나타난다는 4·3절의 결과와 일치한다.

[표 13] 대표 글자 165자의 구성 개요

총	자형 수	옛한글 자형 수	코퍼스 내 토큰 비중(%)	최빈 자형	최빈 자형 빈도(회)
가로모임 민글자	48	1	28.66	이	570
가로모임 받침글자	30	0	10.12	일	109
세로모임 민글자	37	10	22.65	ㅎ	840
세로모임 받침글자	25	7	10.77	을	256
석임모임 민글자	12	5	5.84	의	237
석임모임 받침글자	4	2	1.25	원	56
총화 소계	156	25	79.29		

※ 위 156자에 [표 15]의 자모 보강 자형 9자를 더한 165자가 최종 대표 글자 집합이다. 보강 자형은 회소 자모를 담기 위해 선정된 것이므로 토큰 비중 합산에서는 0.2% 미만에 해당한다. 토큰 비중은 통합 코퍼스 13,327 토큰을 분모로 한다.

총화 절차로 선정된 156자의 전체 목록은 [표 14]와 같다. 각 총안에서는 코퍼스 내 빈도 내림순으로 배열하였다.

[표 14] 총화 대표 글자 156자 목록

총	자형 수	옛한글 자형 수	대표 글자 (코퍼스 내 빈도 내림순)
가로모임 민글자	48	1	이, 니, 지, 야, 라, 예, 다, 디, 셔, 거, 아, 디, 가, 대, 하, 여, 러, 어, 리, 나, 히, 티, 지, 내, 미, 샤, 세, 며, 기, 태, 써, 마, 제, 게, 예, 러, 매, 제, 거, 저, 데, 비, 파, 사, 게, 허, 바, 서
가로모임 받침글자	30	0	일, 경, 상, 신, 만, 삼, 인, 병, 전, 안, 적, 진, 석, 양, 시, 건, 심, 면, 실, 평, 천, 장, 정, 당, 성, 년, 청, 텃, 명, 녕
세로모임 민글자	37	10	ㅎ, 고, 오, 로, 수, 조, 도, 으, 슈, 도, 보, 부, 낙, 르, 구, 드, 노, 브, 그, 후, 무, 우, 수, 호, 조, 드, 바, 박, 쏘, 모, 유, 토, 소, 그, 뉴, 쇼, 두, 모
세로모임 받침글자	25	7	을, 룰, 군, 난, 읍, 은, 흘, 똥, 공, 못, 동, 문, 음, 늘, 뽕, 7, 글, 글, 글, 글, 글, 글, 늘, 솔, 흘, 둔
석임모임 민글자	12	5	의, 다, 기, 와, 위, 화, 시, 과, 되, 지, 쥘, 니
석임모임 받침글자	4	2	원, 빅, 관, 힝

156자에 포함되지 않은 회소 자모를 담기 위해 추가한 9자는 [표 15]와 같다. 이 9자는 사용 빈도가 낮으나 해당 자모의 조형 규칙을 세우는 유일한 근거이므로 빈도 기준에 더해 삼중 기준의 필요성을 직접 제시한다.

이상의 절차를 정리하면 대표 글자 도출은 다음 세 단계로 정식화된다. 1단계는 빈도 기준으로, 누적 커버리지 목표에서 세트 규모를 결정한다. 2단계는 구조 기준으로, 여섯 모임꼴 토큰 비중에 따라 세트 내부를 총화 배분한다. 3단계는 자모 기준으로, 미포괄 자모를 담은 최소 자형을 보강한다. 규모가 정해지지 않으면 총화 배분을 할 수 없고, 총화 배분이 끝나지 않으면 무엇이 미포괄인지 확정할 수 없다는 점에서 세 단계는 순서대로 진행해야 한다.

[표 15] 자모 보강 자형 9자

번호	자형	보강 대상 자모	위치	선정 근거
1	ㅍ	ㅂ	첫당자	어필에서 ㅂ을 담은 최빈 자형
2	ㅍ	ㅅ	첫당자	겹서 계열 첫당자
3	ㅍ	ㅅ	첫당자	겹서 계열 첫당자 + 아래아
4	ㅍ	ㅅ	첫당자·홀자	첫당자 ㄱ과 홀자 ㅅ을 겹쳐 1자 절약
5	ㅍ	ㅅ	홀자	겹줄기 홀자
6	ㅍ	ㅅ	홀자	섞임모임 겹줄기 홀자
7	ㅍ	ㅅ	홀자	옛한글 겹줄기 홀자
8	ㄴ	ㄹ	받침당자	겹받침 ‘ㄴ’ 받침 공간 배분 규칙의 근거
9	ㄴ	ㄹ	받침당자	겹받침 + 아래아

4-8. 대표 글자 165자의 조형 변수 배분

대표 글자 165자는 설계 대상의 목록이면서 동시에 설계 순서의 규정이다. 대표 글자를 먼저 그린다는 것은 그 글자에서 결정되는 조형 변수를 먼저 확정한다는 의미이며, 총화 배분은 어떤 변수를 어느 시점에 확정할 것인지를 정해 놓는다. 따라서 165자의 총 구성은 조형 변수의 확정 순서로 볼 수 있다.

작업 순서는 먼저 대표 글자 중 민글자 85자에서 골격의 무게, 속공간의 크기, 줄기의 기울기를 확정한다. 그리고 받침글자 55자에서 공간의 배분 규칙을 확정하고, 섞임모임 16자에서 혼합 계열의 보정값을 확정한다. 마지막으로 보강 9자를 배치하여 희소 자모에 적용하는 검증을 진행한다. 이러한 과정을 통해 서체 디자인에서 대표 글자 적용의 순서를 체계화한다.

5. 결론

본 연구는 조선 후기 왕실 한글 필사 자료인 『정조 어필한글편지첩』과 『곤전어필』을 서체 연구용 코퍼스로 구축하고, 빈도, 구조, 자모 커버리지의 세 기준을 순차적으로 적용하여 역사 자료에 적합한 대표 글자를 도출하였다. 연구 문제에 대한 결과는 다음과 같다.

첫째, 낱자 사용 빈도는 현대 한글과 뚜렷한 차이를 보인다. 아래아가 정조어필 13.24%, 곤전어필 14.55%로 상위에 위치하며, 아래아를 포함하면 현대에 부재하는 홀자의 비중이 15.7~17.5%에 이른다. 또한 ‘ㅏ’와 ‘ㅑ’의 순위가 역전되고, ‘ㅑ’의 비중은 두 배 이상 높으며, 정조어필에서는 ‘ㅑ’ 받침이 6.95%로

4위를 차지한다. 이는 현대 코퍼스를 기반으로 도출한 대표 글자 목록을 역사 자료의 서체 복원에 그대로 적용하기 어렵다는 것을 보여 준다.

둘째, 모임꼴 구조 분포에서 어필 자료는 현대 한글과 차이를 보인다. 세로모임 비중이 39.1~40.2%로 현대 한글의 30.9%보다 8.2~9.3%p 높다. 이 차이는 어휘나 문체가 아니라 아래아라는 표기 체계의 차이에서 비롯된 구조적 결과로 해석할 수 있다. 조형적 대칭 처리가 요구되는 자형의 실사용 비중 역시 현대와 다르게 구성되며, 대표 글자 설계에서는 세로모임꼴과 섞임모임꼴에 더 많은 검증 기회를 배분할 필요가 있다.

셋째, 본 코퍼스에서 누적 사용량 80% 지점은 156자로 나타났으며, 이는 선행 연구가 제안한 109~181자와 같은 자릿수에 속한다. 다만 두 자료의 관측이므로 임계 규모의 일반적 수렴을 주장하기에는 사례가 부족하며, 자료 확대를 통한 재확인 필요하다.

넷째, 빈도만을 기준으로 선정한 대표 글자는 구조 대표성을 충분히 확보하지 못한다. 빈도순 156자는 여섯 모임꼴 구성에서 최대 6.9%p의 편차를 보였다. 반면 모임꼴 총화 기준을 결합한 156자는 커버리지 손실 0.7%p만으로 편차를 0.4%p 이내로 축소하였다.

다섯째, 구조 대표성을 확보한 대표 글자가 자모 수준의 대표성까지 확보하지는 못한다. 빈도순 156자와 총화 156자의 JCI는 0.691과 0.689로 사실상 동일하며, 자모 포괄 중수도 첫당자 15/19, 홀자 21/25, 받침당자 7/9로 완전히 같다. 이는 모임꼴 구성 편차와 자모 커버리지가 서로 독립적인 기준임을 뜻한다. 나아가 세트를 빈도순으로 400자까지 확대해도 JCI는 0.792에 머물러, 규모 확대로는 자모 커버리지 문제를 해소하기 어렵다. 미포괄 자모 10종을 담은 9자를 표적 보강한 165자 세트는 커버리지 79.5%와 모임꼴 편차 1.6%p를 유지하면서 자모를 전수 포괄하고 JCI를 0.828로 높였다. 이는 자형 수를 5.8% 늘려 빈도순 400자를 상회하는 자모 대표성을 얻은 결과이다.

이상의 결과는 서체 디자인 실무에 다음과 같이 적용할 수 있다. 첫째, 165자를 1차 설계 대상으로 삼아 골격의 무게, 속공간의 크기, 줄기의 기울기 등 서체 전체를 규정하는 조형 변수를 확정한다. 둘째, 모임꼴 별 파생 규칙을 세로모임과 섞임모임에 우선 배분한다. 어필 자료는 이 두 유형의 실사용 비중이 현대 한글보다 9.9~13.5%p 높으므로 현대 서체 개발의 관행대로 가로모임을 기준으로 규칙을 세우면 실제 지면에서 가장 자주 노출되는 자형이 검증에서 밀려난다. 또한, 세

로모임 민글자는 서체 인상에 미치는 영향 대비 설계 부담이 낮다. 셋째, 옛한글 자형을 선택 사항이 아닌 필수 항목으로 포함한다. 본 코퍼스에서 옛한글은 자형의 15.9%, 토큰의 18.3%를 점하므로 이를 제외한 서체는 원자료의 약 6분의 1을 조판할 수 없다. 넷째, 165자를 참조 세트로 활용하여 미포함 자모의 추정을 최소화한다. 이 네 항목은 서체 개발 착수 단계에서 우선적으로 고려할 사항이다.

본 연구의 의의는 세 가지로 정리할 수 있다. 자료적으로는 왕실 어필 두 종을 자형 단위로 계량한 코퍼스 규격을 제시하였다. 방법론적으로는 옛한글 첫가끝 연쇄를 음절 단위로 병합 계수하고 낱자 단위로 분해하여 여섯 모임꼴로 분류하는 절차를 정식화하였으며, 이를 통해 기존 유형화에서 발생하던 잔여 범주를 해소하였다. 아울러 선행 연구에서 개념 수준으로 제시된 자모 커버리지 지수에 모집단 설정, 받침 범주 처리, 계수 단위의 세 조건을 명시하고 이를 역사 자료 코퍼스에 실측 적용하였다. 응용적으로는 빈도·구조·자모 커버리지를 결합한 삼중 기준 표집 절차와 그 산물인 대표 글자 165자를 제안하였다. 나아가 본 절차는 왕실 어필에 한정되지 않는다. 이로써 본 연구는 문화유산 서체의 디지털 전환에서 착수 단계의 판단을 재현 가능한 절차로 기술할 수 있는 기반을 제공한다. 다만 절차의 적용 결과가 조형적으로 더 우수한 산물을 낳는지는 본 연구의 범위를 넘어선다.

본 연구는 다음과 같은 한계를 갖는다. 첫째, 분석 대상이 두 자료에 한정되어 왕실 한글 전반으로 일반화하기 어렵고, 모임꼴 분포의 차이가 왕실 한글의 특성인지 조선 후기 한글 표기 전반의 특성인지는 구분하지 못하였다. 둘째, 판독문을 저본으로 삼았으므로 판독의 이견이 있는 부분은 계수에 반영되었다. 셋째, JCI는 입력 세트의 구성을 기술하는 지표로만 사용되었으며, 지표의 값과 실제 복원 결과의 품질 간 관계는 검증하지 못하였다. 넷째, 도출된 165자 세트의 실효성 역시 아직 검증되지 않았다. 삼중 기준 세트가 빈도순 세트보다 실제로 우월한 조형 결과를 낳는지는 후속 실험을 통해 확인할 필요가 있다. 이 검증은 참조 세트 구성을 독립 변수로 두고 코퍼스 내 미포함 자형을 생성 대상으로 하여 구조 유사도(SSIM), 지각적 유사도(LPIPS), 옛한글 인식률과 전문가 조형 평가를 병용하는 설계로 수행할 수 있다.

후속 연구는 조선 후기 한글 자료를 확대하여 모임꼴 분포와 JCI의 수렴 여부를 검증하고, 본 연구의 165자와 동일 규모의 빈도순 세트를 참조 세트로 통해

실험을 실시하여 표집 기준의 실효성을 평가할 필요가 있다. 아울러 이용제(2012)가 제안한 좌우 대칭과 비대칭, 밀도의 고름과 불균등을 대비할 수 있는 구성 요건을 삼중 절차에 추가로 결합하는 방안을 검토할 필요가 있다.

참고문헌

1. 국립한글박물관, 『정조어필한글편지첩』, 국립한글박물관 소장자료총서 1, 2014
2. 김두식, 『한글 글꼴의 역사: 15세기~19세기 활자본 및 목판본을 중심으로』, 시간의 물레, 2008
3. 박재홍, ‘한글 글자디자인을 위한 21세기 세종계획 한글 빈도수 조사분석’, 조형미디어학, 제21권 제3호, 한국일러스트아트학회, 2018
4. 박정숙, ‘조선시대 (18세기) 왕실 한글 필사류 자료의 서예적 고찰’, 국립한글박물관 소장자료총서 1, 국립한글박물관, 2014
5. 안상수, 한재준, 이용제, 『한글 디자인 교과서』, 안그래픽스, 2009
6. 양효정, 임순범, 안병학, ‘고문헌 집자를 통한 옛한글 글꼴 디자인 토대 연구’, 기초조형학연구, 제23권 제5호, 한국기초조형학회, 2023
7. 우형식, ‘문자의 보수성과 아래아 표기의 변천’, 한국어문화교육, 제15권 제2호, 한국어문화교육학회, 2022
8. 이용제, ‘효율적인 한글 활자디자인을 위한 대표 글자 연구: 글자의 균형과 비례를 중심으로’, 글짜씨, 6호, 한국타이포그래피학회, 2012
9. 이은하, 남기춘, ‘세종 말뭉치에 나타난 한국어 음절의 빈도와 분포’, 언어과학연구, 제92호, 언어과학회, 2020
10. Hong, J. K., Kim, S. K., ‘Few-Shot Learning for Irregular Hangeul Typeface Expansion: A Comparative Study of GAN, VQGAN, and Diffusion Models’, Electronics, Vol.15, 2633, 2026