

기초 통계량을 이용한 저작자 진위 추론

이근무¹⁾, 이근우²⁾

요약

이 논문에서 문장특성을 파악하는 방법으로 주로 이용한 것은 특정문자의 출현율이다. 어떤 사람이나 글 속에는 자신의 개성이 들어있다. 문장의 길이를 비롯하여 문장의 구조나 어휘량, 유의어 중에서 선호하는 글자, 평서문이나 의문문의 사용, 품사의 사용, 문두나 문말에 오는 글자 등에서 각각의 개성이 드러난다. 그 중에서도 접속사나 조사, 접두어, 접미어 등 상대적으로 의미적인 요소보다는 형식적인 요소에 가까운 영역에서 문장의 특성이 두드러지는 것으로 보고되어 있다, 이런 특징을 이용하여 화랑세기의 저작자의 진위를 추론하고자 한다.

주요용어 : 문헌계량학, 유의어, 저작스타일, 어휘량

1. 서론

동서고금을 막론하고 세상에는 적지 않은 위작(僞作) 혹은 도작(盜作)들이 존재하였다. 예를 들어 셰익스피어의 희곡, 영국 조지 3세 때 유니우스(Junius)라는 필명으로 1769년부터 1772년에 걸쳐 당시의 지배층을 비판하기 위해서 신문에 투고된 유니우스 레터(Junius Letter), 미국 국민문학의 아버지라고 불리는 마크 트웨인(1835-1910)의 저작 중 전쟁모험담인 큐시에스 레터(Q.C.S Letter), 일련종(日蓮宗)의 개조(開祖) 니찌렌(日蓮)의 저술이라고 하는 『삼대비법품승사(三大秘法稟承事)』, 미국 헌법의 해설서로서 정치학상 중요한 문헌인 연방주의자(Federalist), 미국 신문왕의 딸 패트리시아 허스트의 유괴사건 때(1974)와 관련된 성명문 등 철학서 종교서 문학작품 역사서 신문기사 서한 등 어느 분야라고 할 것 없이 위작 혹은 위작의 가능성이 제시되어 있는 작품들이 적지 않게 있다[1].

특히 유명한 사상가나 종교가 소설가 등의 저술이라고 가탁한 글들이 적지 않으며, 그러한 위작들은 위대한 인물에 대한 이해나 학문 연구를 현저히 방해한다. 위작은 비록 그것이 아무리 뛰어나고 또 아무리 진작(眞作)과 유사하다고 하더라도 위작일 따름이다. 오히려 진작과 유사하면 유사할수록 사상 학문 종교의 연구를 혼란시킬 뿐만 아니라, 때로는 학문적 종교적인 대논쟁을 불러일으킬 수도 있다. 그러므로 많은 연구자들이 위작을 가려내는 데 적지 않은 노력을 기울여온 것이 사실이다. 그러나 그러한 노력에도 불구하고 여전히 해결을 보지 못한 문헌들도 적지 않다. 이 글에서 문제삼고자 하는 이른바 필사본 화랑세기도 우리 사회에서는 보기 드문 위작 논쟁을 불러일으켜 활발한 논의가 전개되었으나, 진위 여부에 대한 합의에 이르지 못한 채 쌍방의 논의가 평행선을 달리고 있는 상황이다.

1989년에 초략본 화랑세기가 발견된 직후부터 그 진위를 둘러싸고 서로 대립되는 견해가 제시되었다.

1) 경주대학교 컴퓨터 전자공학부 컴퓨터 공학전공 겸임교수

2) 부경대학교 역사학과 부교수

이후 초락본 화랑세기와 1995년에 공식적으로 확인된 필사본 화랑세기를 7세기 말의 김대문이 저술한 책으로 인정하려는 측과 이를 위작으로 보는 측은 팽팽히 대립하여 서로의 주장을 굽히지 않고 있다. 물론 양극단에 있는 주장 사이에는 진위판정에 좀더 신중한 태도를 가지고 있는 연구자들도 적지 않은가 하면, 진위 여부를 떠나서 필사본 화랑세기가 신라사회를 이해하는데 일정한 도움을 준다고 생각하여 적극적으로 자신의 연구에 반영하는 연구자들도 적지 않다[2].

그런데 화랑세기를 둘러싼 이러한 논쟁이 평행선을 달리고 있는 이유는 삼국사기나 삼국유사라고 하는 우리 고대사회에 대한 극히 제한된 사료만을 가지고 그 진위 여부를 분명하게 가려내기 어렵기 때문이라고 할 수 있다. 전통적으로 문헌의 진위감정은 1) 기재내용의 검토와 역사적 사실의 고증, 2) 필적의 감정, 3) 지질 먹 등의 화학적 분석에 의한 연대추정 등의 방법에 의존하였다.

문제가 되는 필사본 화랑세기는 20세기에 들어서 남당(南堂) 박창화(朴昌和)에 의하여 필사된 것이라고 하므로, 2)나 3)의 방법을 전혀 사용할 수 없으므로 지금까지는 주로 1)의 방법에 의존해왔다. 그렇지만 현실적으로 화랑세기에는 보이지만 삼국사기 등에 보이지 않는 사실관계를 여타의 사료를 통해서 직접적으로 확인할 수 없는 한계 때문에, 1)의 방법이 결정적인 논거를 제시하지 못하고 있는 형편이다. 더해서 신라사회를 보는 연구자들의 시각에 내재해있는 개인적인 편차도 화랑세기의 진위감정을 어렵게 하고 있다.

따라서 화랑세기의 진위감정에는 연구자 개인의 편차가 개재되지 않는 객관적인 기준에 따른 방법론이 필요하다고 하겠다. 그런 관점에서 최근에 학문의 한 분야로 자리잡고 있는 문헌계량학의 방법론을 이용하여 화랑세기의 진위감정을 시도해 보고자 한다. 문헌계량학의 방법은 문장의 길이, 단어의 길이, 품사의 출현율, 특수한 단어의 출현율 등 문장의 수량적인 성질을 조사함으로써, 집필자의 문장스타일이나 문장습관을 계량적으로 파악하는 것을 근간으로 한다. 그 진위가 의심받는 화랑세기와 같은 문제 저작들에 대한 진위판정에 서부터 도작 논쟁등에 대한 과학적 접근의 근거를 마련하고자 한다. 그런데 문헌의 진위감정을 계량분석을 통해서 시도할 때, 다음과 같은 문제점을 지적할 수 있다. 먼저 대상이 되는 문헌이 어느 정도의 길이를 가지고 있어야 한다는 것이다. 작성자의 문장습관을 통계학적으로 파악하기 위해서는 문헌의 분량이 적을 경우 거기에서 구한 통계량의 신뢰성이 문제가 될 수 있기 때문이다. 신뢰성이 높은 계량분석을 행하기 위해서는 문헌의 길이가 독립된 한 권의 책 분량인 것이 가장 바람직하다고 할 수 있다. 그렇지 못할 경우에는 일본어 문장의 예에 비추어 우리말 문장인 경우에 최소한 500-1000 단어의 길이가 필요한 것으로 판단된다. 화랑세기와 같은 한문 문장일 경우에는 구체적인 사례가 없으므로 어느 정도가 길이가 필요한지 단정하기 어려우나, 한 권의 책 분량이 화랑세기는 큰 문제가 없다고 하겠다.

둘째 계량분석에 이용가능한 문헌이 충분해야 한다는 점이다. 문헌의 계량분석에는 문헌의 길이에 관계없이 어떤 문헌이나 통계학적으로는 1개의 표본에 불과하다. 따라서 분석에는 가능한 한 많은 문헌을 사용할 수 있다면 보다 신뢰도가 높은 결과를 얻을 수 있으나, 현실적으로 한 개인의 문헌이 충분한 표본을 제공하는 경우가 그렇게 많지 않을 것으로 예측할 수 있다.

셋째, 한 권의 문헌 속에서 문장의 길이, 단어의 길이, 품사의 출현율, 단어의 출현율 등, 문헌의 수량적인 성질에 관한 다량의 정보(통계량)를 얻을 수 있지만, 이들 정보 중에서 문장특성을 발견해 내는 일은 그렇게 용이하지 않다는 점이다. 따라서 현재 단계에는 분석가능한 정보를 모두 조사하는 과정을 거칠 수밖에 없다.

이러한 한계에 불구하고, 문헌 내용의 검증을 통해서 그 진위를 명확히 하기 어려운 경우 혹은 합의에 도달하기 어려운 경우에는 객관적인 지표에 의거하는 통계학적인 방법을 원용할 수밖에 없다. 실제로 이러한 방법을 통해서 위작 논의가 있었던 문헌들에 대하여 신뢰할 만한 결론을 내린 경우가 적지 않기 때문이다.

2. 연구방법

계량문헌학의 방법론에 입각하여 필사본 화랑세기의 진위 여부를 가늠해 보려는 이 글에서는 다음과 같은 순서에 따라서 추론을 진행하고자 한다. 우선 필사본 화랑세기를 둘러싼 논의에서도 알 수 있는 바와 같이 필사본 화랑세기가 南堂 朴昌和에 의해서 위작된 것이라는 주장이 있다. 이러한 주장이 과연 성립될 수 있는지를 판단하기 위하여 박창화의 다른 저술과 필사본 화랑세기가 동일한 문장특성을 나타내는가를 살펴볼 것이다.

실제로 박창화는 金天大帝法興眞王紀와 같은 역사관련 저술을 비롯하여 桃紅紀, 於乙同紀 등 소설류 등 적지 않은 글을 남겼다. 필사본 화랑세기를 제외하면, 다른 문헌들은 박창화의 저술로 간주해도 좋을 것이므로, 이들 저술의 문장특성을 파악한 다음, 이를 필사본 화랑세기의 특성과 비교해 보면, 이들 문헌 사이의 문장 특성이 유사한지 다른지를 판별할 수 있다. 문장특성이 유사하다고 한다면, 필사본 화랑세기는 박창화가 위작하였거나 아니면 자신의 문체로 대대적으로 개서하였다고 볼 수 있을 것이다. 그 반대의 경우에는 박창화가 화랑세기를 위작하였거나 변개하였다는 혐의를 벗을 수 있을 것이다.

이러한 목적을 달성하기 위해서는, 박창화의 전체 저술을 데이터로 입력하여 우선 박창화의 문장특성을 확정한 다음에, 화랑세기의 문장특성과 비교하여야 한다. 그러나 제약된 시간 때문에 박창화의 전체 저술을 데이터로 입력할 수 없었으므로 金天大帝法興眞王紀라고 하는 저술을 필사본 화랑세기와 비교하는 자료로 삼았다. 또 이들을 비교한 결과가 유사성을 나타내는 것인지 이질성을 나타내는 것인지를 파악하는 참고 자료로 삼국사기와 삼국유사를 이용하였다.

3. 유의어 출현율

이 글에서 문장특성을 파악하는 방법으로 주로 이용한 것은 특정문자의 출현율이다. 어떤 사람이나 그 글 속에는 자신의 개성이 들어있다. 문장의 길이를 비롯하여 문장의 구조나 어휘량, 類義語 중에서 선호하는 글자, 평서문이나 의문문의 사용, 품사의 사용, 문두나 문말에 오는 글자 등에서 각각의 개성이 드러난다. 그 중에서도 접속사나 조사, 접두어, 접미어 등 상대적으로 의미적인 요소보다는 형식적인 요소에 가까운 영역에서 문장의 특성이 두드러지는 것으로 보고되어 있다.]

유의어 내부에서 출현율을 비교해 보기로 하였다. 예를 들어 이인칭 대명사인 爾와 汝, 어조사(전치사)인 于와 於, 문말 어조사인 也, 矣, 焉, 지시대명사인 是와 其 등을 각각 하나의 그룹으로 묶어 그 그룹 내에서 어떤 출현율을 나타내는가를 비교한 것이다. 네 문헌 속에서 유의어 내부의 사용빈도를 정리한 것이 <표1>이다.

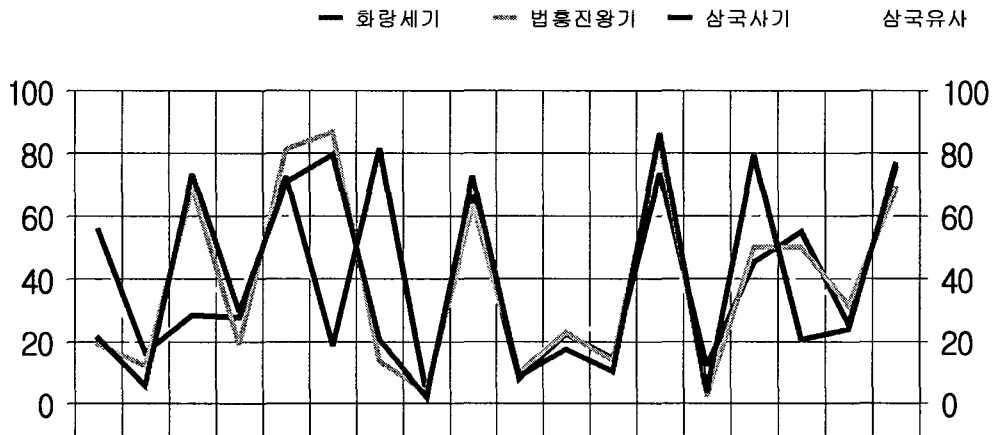
이를 차트로 나타낸 것이 <그림1>이다. 이들 비교에서는 특정문자의 단순비교에서 알 수 있는 것보다는 좀더 확실한 문장특성을 확인할 수 있었다. 삼국사기와 삼국유사의 유의어 비교에서 몇 군데 차이가 나고 있는 점이 분명히 드러나고 있을 뿐만 아니라, 삼국사기와 삼국유사의 차이보다 화랑세기와 법흥진왕기의 차이가 적음을 알 수 있다. 즉 화랑세기와 법흥진왕기의 비교에 있어서는 盖와 凡의 비율에서 다소 차이를 보이고 있을 뿐이며 다른 유의어의 사용에 있어서는 그 비율이 대단히 근접하고 있음이 확인된다.

특히 于와 於의 사용에 있어서 화랑세기와 법흥진왕기에서는 于가 80% 전후가 사용되었으나, 삼국사기와 삼국유사에서는 이와 반대로 於가 80% 정도 사용되고 있다. 又, 且, 亦의 사용에 있어서는 화랑세기와 법흥진왕기에서 20% 전후가 사용되었으나, 삼국사기와 삼국유사에서는 모두 55% 전후가 사용되었다. 또 矣, 也, 焉 중에서 焉은 화랑세기와 법흥진왕기 쪽이 그 사용빈도가 낮다.

기초 통계량을 이용한 저작자 진위 추론

<표 1> 유의어 출현빈도 비교

문자	화랑세기	법흥진왕기	삼국사기	삼국유사	비교
又	20.994	19.047	55.2584	55.9259	
且	5.524	11.904	16.3992	5.9259	
亦	73.480	69.047	28.3422	38.1481	
是	29.430	18.947	27.5247	34.9848	
其	70.569	81.052	72.4752	65.0151	
于	79.421	86.689	18.5239	20.0000	
於	20.578	13.310	81.4760	80.0000	
莫	1.6635	3.6269	3.1352	4.1425	
不	72.643	63.730	66.6354	62.8131	
非	8.3179	9.8445	7.7211	7.8998	
無	17.375	22.797	22.5081	25.1445	
矣	10.03861	13.698	14.3589	22.9188	
也	86.486	84.246	73.6410	66.2898	
焉	3.4749	2.0547	12.0000	10.7913	
盖	79.411	50.000	45.1612	65.6716	
凡	20.588	50.000	54.8387	34.3283	
卽	23.770	31.428	25.3444	36.8030	
則	76.229	68.571	74.7945	63.1970	



<그림 1> 화랑세기 법흥진왕기 삼국사기 삼국유사 유의어 출현빈도

삼국사기와 삼국유사, 화랑세기와 범홍진왕기는 대체로 비슷한 출현빈도를 보이고 있다. 이에 대하여 에서는 화랑세기와 삼국사기, 에서 화랑세기와 삼국유사의 경우에는 그 차이가 두드러지게 나타난다.

이로 보아 특정문자의 출현율보다는 유의어의 출현율이 이 글에서 다루는 4문헌의 문장특성을 비교하는데 보다 유효하다고 판단할 수 있다. 그 결과 화랑세기와 범홍진왕기 간의 문장특성상 유사성이 삼국사기와 삼국유사 사이의 유사성보다 훨씬 크다고 해야할 것이다

4. 결론

이상으로 몇 가지 용어들을 중심으로 해서 화랑세기와 범홍진왕기, 삼국사기와 삼국유사를 비교해 보았다. 가장 먼저 시도해본 한문문장을 형식적으로 구성하는 특정 문자의 출현율을 통한 문장특성에서는 변별적인 내용을 분명하게 확인하기 어려웠다. 두 번째로 시도한 유의어의 출현율에서는 화랑세기와 범홍진왕기의 유사성, 화랑세기 및 삼국사기 삼국유사의 차이점을 확인할 수 있었다. 그러나 유의어 집단으로 선택된 요소들이 반드시 의미가 동일하여 어느 쪽을 작자가 자유롭게 선택할 수 없는 측면도 가지고 있으므로, 유의어 출현율의 비교결과를 이용하는 데는 신중할 필요가 있다.

이러한 결과로 화랑세기가 위서인지를 단정하기는 어려우나, 박창화의 저술로 생각되는 범홍진왕기와 비교해 볼 때, 양 문헌이 상당한 정도로 유사성을 가지고 있는 것으로 판단된다.

참고문헌

- [1]村上征勝, 眞贋の科學-計量文獻學入門, 朝倉書店,1994
- [2]권덕영, 「필사본 화랑세기 진위논쟁 10년」 한국학보 일지사,1995
- [3]노태돈, 「필사본 화랑세기의 사료적 가치」 역사학보, 1995
- [4] 이근무, 저작자 진위 판별을 위한 계량적 지표에 관한 연구, 통계학회 2001 년 춘계개발포 논문
- [5] D*, Stremja Tichogo Dona(Zakadki romana). YMCA-prews, Paris,1974.
- [6] G. Kjesaa et al, The authorshio of Quiet Done, Humanities Press, 1984
- [7] R. D. Lord, Studies in the history of probability and statistics VIII. De Morgan and the statistical study of literature style. *Biometrika*,45, 1958
- [8] S. E De Morgan, Memoir of Augustus de Morgan by his wife Sophie Elizabeth De Morgan with selection from his letters. Longman,Green and Co.,1882.
- [9] G. U. Yule, The statistical study of Literacy Vocabulary . Cambridge University press, 1944
- [10] 藤退偉作, 未知の著作権の推定について。BASIC 數學 2 月号,1986
- [11] L. L. Adams and A. C. Rencher, The popular critical view of the Isaiah problem in light of statistical style analysis. *Computer Studies in the Humanities and Verbal Behavior* . 7(3-4),1973.
- [12] 村上征勝, 眞贋の科學, 朝倉書店, 1994.
- [13] 種村秀弘, 贋作者列傳, 青土社, 1992.