대한한의학원전학회지
2018년 5월 31권 2호

pISSN 1229-8328
J Korean Med Classics 2018:31(2) 071-082
https://doi.org/10.14369/jkmc.2018.31.2.071

JKMC
Journal of Korean Medical Classics

의학 사상의 유사성은 계량 분석 될 수 있는가

-『동의보감』과『의학입문』,『경악전서』를 중심으로 -

한국한의학연구원
오준호*

Can Similarities in Medical thought be Quantified?

- Focusing on *Donguibogam*, *Uihagibmun* and *Gyeongagjeonseo* -

Oh Junho*
Korea Institute of Oriental Medicine

Objectives : The purpose of this study is to compare the similarities among *Donguibogam(DO)*, *Uihagibmun(UI)*, and *Gyeongagjeonseo(GY)* in order to examine whether the medical thoughts embedded in the texts can be compared in a quantitative way.

Methods : Under an empirical assumption that medical thoughts can be reduced to the frequency of major key words within the text, we selected the fourteen words of the four categories that are commonly used to describe physiology and pathology in Korean medicine as key words. And the frequency of these key words was measured and compared with each other in the three important medical texts in Korea.

Results : As a result of quantitative analysis based on $\chi 2$ statistic, the key words in the books were distributed most heterogeneously in *DO* and distributed most homogeneously in *UI*. In comparison of the similarity analyzed by the same method, *DO* and *UI* were significantly more similar than those of *DO* and *UI*. The results of the word frequency pattern and the similarities of the book contents(CBDF) show that *DO* is influenced by *UI*, and the differences between standardized residuals and homogeneity tells us that internal context of both books are constructed differently.

Conclusions : These results support the results of traditional research by experts. With the above, we were able to confirm that medical thoughts can be reduced to the frequency of major key words within the text, and compared through the frequency of such key words.

Key Words : Korean Medicine, quantitative analysis, word frequency, Chi-Square Statistic, Donguibogam.

* Corresponding Author : Oh Junho.
Korea Institute of Oriental Medicine, 1672 Yuseong-daero, Yuseong-gu, Daejeon, 34054
Tel : +82-42-868-9317, E-mail : junho@kiom.re.kr
Received(April 20, 2018), Revised(May 1, 2018), Accepted(May 1, 2018)

Copyright ⓒ The Society of Korean Medical Classics. All rights reserved.
ⓒ This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

오늘날 과거에 만들어진 전통의학 서적을 탐구하는 것은 책을 지은 의가(醫家)의 '의학 경험'을 알고 싶기 때문이다. 어떤 병증을 가진 환자를 보았고, 그 환자를 치료하기 위해 어떤 사유를 하였으며, 그 사유 끝에 어떤 치료 방법이 탄생하였는지를 알 수 있다면 오늘날 우리가 직면한 의학적 문제에 대해서도 많은 단초를 얻을 수 있다. 그러나 한 의가가 경험한 '현실 세계'는 이미 오래 전에 역사 속으로 사라졌기 때문에 그의 '의학 사상'을 통해 그것을 재구성해야 한다. 의학 사상은 그의 '저서'에 자연스럽게 스며들어 있다. 따라서 오늘날 우리는 '저서'로부터 의가의 '의학 사상'을 추출하고, 그 의학 사상을 통해 그가 겪었을 '의학 경험'을 재구성해야 한다.

동아시아 전통의학을 연구하는 전통의학 전문가들은 이를 위해 의학 서적을 대상으로 저자들의 사승관계나 핵심 내용을 분석하여 의학 사상이 전해지고 변화해 가는 모습을 포착하려고 노력해 왔다. 이런 작업은 매우 전문적인 영역으로 대부분 오랜 시간 동안 숙련된 지식을 바탕으로 이루어졌다. 이러한 정성적인 평가 방법이 학술적으로 매우 중요함에도 불구하고, 그 결과를 검토하거나 경쟁하는 주장을 평가하는 작업은 거의 이루어지지 못하였다. 그것은 현실적으로 선행 연구의 근거와 논리를 따져보는 것 이외에 다른 연구 방법이 존재하지 않기 때문이다.

전문가들의 지식과 경험에 기반을 둔 기존 연구 방법과 별개로, 이를 보완하거나 검토할 새로운 접근 방법이 개발된다면 동아시아 전통의학 서적에 내재된 의학이론을 고찰하거나 비교하는 연구에 도움이 될 것이다. 또 기존의 전문가 고찰 방법으로 도출해 내지 못했던 결과를 얻거나 새로운 탐구 소재를 발견하는 경우에도 기여할 수 있다. 본 연구에서는 이러한 접근 방법의 하나로 디지털 텍스트를 계량 분석하는 정량적 방법을 검토하고자 한다[1].

디지털 텍스트를 통한 계량 분석 방법은 일찍이 정보 검색(Information retrieval), 전산언어학(computational linguistics) 등에서 대량의 데이터를 조작하거나 그 속에서 의미를 발견하기 위한 방법으로 발전되어 왔다. 디지털 텍스트를 통한 방법은 단어의 빈도(frequency)를 기반으로 이루어지는데, 텍스트가 가지는 의미 손실이 크다는 비판에도 불구하고 비교적 간단히 만들 수 있고 작업을 자동화 할 수 있어서 수작업에 비해 시간과 노력이 절약되는 장점 때문에 보편적으로 사용되는 방식이다.

본고에서는 디지털 텍스트에 대한 계량 분석 방법의 하나인 카이제곱 값을 계량 지표로 하여 『동의보감』이 『의학입문』 및 『경악전서』와 가지는 관계를 정량화하고 그 차이를 비교함으로써 한의학 분야에서 의학 사상을 탐구하는 경우에도 이 방법이 유용한지 검토해 보고자 한다.

2. 대상 및 방법

계량적 분석 방법이 의학 사상을 살펴보는데 도움이 될 수 있는지 검토하기 위해 본고에서는 다음과 같은 기준에 따라 분석 대상 서적을 도출하였다. 첫째, 계량 분석 결과를 신뢰하기 위해 분량이 가능한 많은 의서일 것[2]. 둘째, 의학 사상의 비교를 위해 다루고 있는 질병의 범위가 유사한 의서일 것. 셋째, 기존 연구 성과와의 비교를 위해 학계에 익숙한 서적으로 정성적 분석 연구에서 다루어진 의서일 것.

분과 서적은 예외적인 경우를 제외하고는 분량이

1) 의서의 특성을 정량적으로 비교하려는 시도는 이미 몇 차례 이루어졌다. 이 연구들은 치료 방법에 주안점을 두어 본초 구성 비율을 이용한 것으로 본고의 연구 방법론과는 차이가 있다.
김기욱, 김태열, 이병욱. 본초 비율의 순위를 이용한 문헌의 특징 분석 방법 - 태평혜민화제국방, 난실비장, 소문선명론방을 중심으로. 대한한의학원전학회지. 2014. 27(4). pp.73-84.
오월한 外 3인. 방제구성을 이용한 유하간(劉河間) 및 이동원(李東垣)의 저작과 『화제국방(和劑局方)』의 특성 비교. 대한한의학원전학회지. 2015. 28(1). pp.55-69.

2) 텍스트 기반의 계량분석은 텍스트를 용어집(types)의 집합으로 보고 이들의 빈도를 측정한다. 이때 측정된 빈도가 작으면 그로부터 도출된 결과를 신뢰하기 어렵다. 따라서 희소한 용어까지도 빈도가 어느 정도 측정될 수 있도록 텍스트의 크기가 커야 한다.

www.kci.go.kr

많지 않고 학계의 연구도 충분하지 않다고 판단하여 첫 번째와 두 번째 기준에 따라 종합의서를 우선 고려하였다. 종합의서 가운데 학계에 관심과 연구가 많은 서적으로는 단연 『동의보감』을 들 수 있으며, 서적의 분량과 의학 사상에 있어서 『동의보감』과 비견할 만한 서적으로 『의학입문』과 『경악전서』를 추가로 선정하였다.

텍스트 사이의 관계를 유사도(similarity) 혹은 거리(distance)[3]로 환원하는 방법은 텍스트를 대상으로 하는 계량 분석에서 일반적으로 사용되는 방법이다. 이러한 분석은 용어 사용 빈도가 해당 텍스트의 내용을 반영한다는 가정 위에서 작동한다. 텍스트에 대한 계량분석이 보편화된 오늘날에도 이 가정에 대한 옳고 그름은 여전히 논쟁거리로 남아 있다. 경험적으로 어느 정도 수긍할 수 있기도 하지만, 다종다양한 텍스트들에 이러한 가정이 모두 그대로 적용된다고 증명하기 어려운 면도 있다. 본고는 이러한 가정이 한의학 고문헌에도 적용될 수 있는지에 대한 탐색적인 연구를 지향하고 있다. 따라서 의서의 의학 사상은 용어 사용을 통해 드러난다는 가정을 바탕으로 논의를 진행한 뒤, 그 결과와 기존 연구 결과를 비교하여 이 가정에 대해 검토해 보고자 한다.

한의학 고문헌 텍스트를 계량적으로 분석하기 위해서는 먼저 ①디지털 텍스트 전자파일 구축, ②빈도의 기준이 되는 용어집 마련, ③빈도 측정 및 정규화, ④유사성(혹은 거리) 계측의 과정을 거치게 된다. 본고에서도 이와 같은 과정을 수행하였다.

먼저 『동의보감』, 『의학입문』, 『경악전서』 각각의 전자파일(digital file)을 마련하였다[4](이하 '디지털 텍스트' 혹은 '전자 텍스트'로 지칭). 이 디지털 텍스트에서 한자 이외의 기호들을 모두 제거하였고, 분석 오류를 줄이기 위해 다중코드자와 이체자를 통합하였다[5].

빈도의 기준이 되는 용어집(types)은 연구 목적에 따라 정해진다. 본 연구에서는 의학 사상을 검토하는 것을 목적으로 하고 있기 때문에 한의학에서 의학 사상을 가장 잘 드러낸다고 여겨지는 용어를 선택하고자 하였다. 이에 대한 선행연구가 존재하지 않기 때문에 본고에서는 다음과 같은 기준을 수립하였다. 첫째, 한의학 고문헌에 보편적으로 나타나는 용어일 것. 둘째, 생리·병리 현상 등 의학 사상과 관련이 깊은 용어일 것[6]. 셋째, 출현 빈도가 높은 용어일 것. 이상의 기준에 부합하는 용어로 오장(五臟 : 肝·心·脾·肺·腎), 육기(六氣 : 風·寒·暑·濕·燥·火), 정기신혈담울(精·氣·神·血·痰·鬱), 팔강(八綱 : 陰·陽·表·裏·寒·熱·虛·實)[7]의 용어들 24가지(肝·心·脾·肺·腎·風·寒·暑·濕·燥·火·精·氣·神·血·痰·鬱·陰·陽·表·裏·熱·虛·實)를 선정하였다(이하 '핵심어'로 지칭)[8].

본고에서 핵심어를 한 글자(uni-gram)로 상정한 것은 현재 학계에 신뢰할만한 용어집이 존재하지 않는다는 현실적인 이유도 있지만, 한자(漢字)가 표음문자와 달리 하나의 글자로 뜻을 표현할 수 있는 표의문자라는 특수성을 고려한 결과이다. 일반적인 텍스트 계량 분석에서 한 글자로 용어집을 구축하는 경우는 없다. 그것은 표음문자에서 한 글자는 의미를 지니지 못하기 때문이다. 예를 들어 알파벳의 "L"이나 한글에서의 "간"이라는 글자는 어떤 의미를 갖지 못한다. "간(肝)"이라고 한자를 병기해 주거

3) 유사도(similarity)와 거리(distance)는 반대 개념으로 사용된다. 즉, A와 B 두 인자의 유사도가 높으면 거리는 가깝다고 할 수 있으며, 유사도가 낮으면 거리가 멀다고 할 수 있다. 이 두 개념은 함께 사용되는 경우가 많기 때문에 미리 관계를 밝혀 둔다.

4) 전자 텍스트는 한국한의학연구원에서 제공하고 있는 한의학고전DB(mediclassics.kr)로부터 추출하였다.

5) 이정현. 유니코드 한자 검색의 문제점 및 개선방안. 정보화정책저널. 2012. 19(3). pp.50-63.

6) 의서를 치료 데이터로 보았을 때 '병증', '본초', '혈위'는 가장 핵심적인 용어이다. 이들 용어는 종류가 많고 같은 의미를 다르게 표현하는 경우가 많다. 이들 용어를 연구에 활용하기 위해서는 이들 각각의 용어집 구축이 선행되어야 한다. 이와 관련해서는 후속 연구가 필요하다.

7) 팔강(八綱)의 경우, 이들 서적들이 성립될 당시에는 팔강이라는 개념이 존재하지 않았지만 팔강을 구성하는 음양 표리 허실 한열의 개념들은 새로운 것이 아니다.

8) 단어(글자)는 문맥에 따라 다른 의미를 지닐 수 있으나, 선정된 핵심어는 의서 내에서 빈도가 충분히 높게 나타날 것을 기대되기 때문에 이러한 영향을 적게 받을 것으로 판단하였다. 따라서 빈도 측정 시 문맥에 따른 의미 변화는 고려하지 않았다. 단, 고유명사로서 핵심어의 의미를 가지지 않는 "丹溪心法", "脾胃論"의 단어들을 예측 가능한 오류이므로 디지털 텍스트에서 처음부터 제거하였다.

www.kci.go.kr

나 “간장”이라고 하나의 단어를 구성하여 비로소 의미를 구체적으로 전달할 수 있다. 그러나 한문에서의 한 글자는 표음문자에서 하나의 단어에 준한다. 한자는 “肝”이라는 글자 하나로 이미 의미를 표현할 수 있기 때문이다. 물론 “肝虛”, “抑肝” 등 연속되는 글자에 따라 의미가 확장되고 변형되기도 하지만 “肝”이라는 본질적인 의미를 완전히 벗어나는 경우는 거의 없다. 따라서 한자로는 비록 한 글자에 불과하지만 표음문자로 이루어진 텍스트 정량 분석에서 하나의 단어를 기준으로 진행하는 분석과 유사하다고 할 수 있다.

빈도를 통해 유사성(similarity)을 측정하는 방법은 매우 다양하다. 그러나 절대적으로 우월한 방법이 있다고 할 수 없으며 대상 자료의 성격과 분석의 목적에 따라 달라질 수 있다[9]. 따라서 본 연구에서는 다음의 기준을 통해 χ^2값(chi-squared statistic)을 선택하였다. 첫째, 빈도를 변수로 하여 보편적으로 널리 사용되는 방법일 것, 둘째, 공식이 의미하는 바를 직관적으로 파악하기 용이할 것, 셋째, 텍스트를 이루는 언어나 텍스트의 내용에 의존적이지 않을 것. χ^2값[1]은 해당 공식은 관찰된 값(O)과 기댓값(E) 사이의 차이를 의미하며, 범주형 자료로 이루어진 데이터를 분석하는데 보편적으로 사용되는 방법이다. 또 텍스트의 언어나 내용 특성과 무관하며 빈도 데이터에 적합한 방법이다.

χ^2값은 자유도(df ; degree of freedom)에 따라 차이가 날 수 있으므로 본 연구에서는 이것을 자유도로 나누어 정규화(normalization)한 CBDF(Chi By Degrees of Freedom) 값을 사용하였다. 이 값을 통해 텍스트 내의 동질성과 텍스트 사이의 유사성을 측정하였다. CBDF는 Corpus 비교를 위해 Kilgarriff가 제안한 방법[10]으로 0보다 큰 값만을 가지며 보통 1에서 50 사이의 값을 보인다[11].

$$\chi^2 = \sum \frac{(O-E)^2}{E} \qquad [1]$$

3. 결과

3.1. 단순 빈도 비교 결과

우선 핵심어의 빈도로 해당 서적을 환원시켰다. 이들 핵심어의 빈도의 패턴을 통해 의서를 먼저 비교하였다. 모든 글자의 총 빈도수는 『동의보감』이 868,411자, 『의학입문』이 696,152자, 『경악전서』가 862,441자로 나타났다[12]. 『동의보감』과 『경악전서』는 거의 동일한 수이며 『의학입문』은 상대적으로 적었다. 분석 과정에서 나타날 수 있는 왜곡을 방지하기 위해 핵심어 빈도를 셈한 뒤에 이 세 서적의 글자 수 평균으로 정규화한 상대관찰빈도를 이후 분석에 이용하였다.

1) 전체 핵심어 빈도

의서에서 관찰된 빈도의 분포를 서적별로 비교해 보면 핵심어 빈도의 순위가 서적별로 유사하게 나타났다(Fig. 1a, 1b). 상관계수(Correlation coefficient)[13]를 보면 『동의보감』-『의학입문』이 0.98, 『동의보감』-『경악전서』가 0.93로 양자 모두 높은 상관관계를 보였다. 한 서적에서 높은 빈도로 나타난 핵심어가 다른 서적에서도 높게 나타난다는 의미로 풀이된다.

핵심어 빈도를 보면, 고빈도 5개 핵심어 가운데 『동의보감』과 『의학입문』은 氣・血・熱・風 등 4개가 동일하였고, 『동의보감』과 『경악전서』는 氣・熱 2개만 동일했다. 氣와 熱은 세 서적 모두 높은 빈도를 보였다(Table. 1 참조). 특히 氣의 경우 2순위 핵심어와의 빈도 차이가 컸다.

9) Pavel Pecina. Lexical association measures and collocation extraction. Lang Resources & Evaluation. 2010. 44. 137-158.

10) Adam Kilgarriff. Using Word Frequency Lists to Measure Corpus Homogeneity and Similarity between Corpora. In Language Engineering for Document Analysis and Recognition. Proceedings, AISB Workshop(Falmer). 1997. pp.231-245.

11) Adam Kilgarriff, Raphael Salkie. Corpus similarity and homogeneity via word frequency. Proceedings, EURALEX(Gothenberg, Sweden). 1996. pp.121-130.

12) ‘寒’은 육기와 팔강에 모두 포함됨. 핵심어가 포함된 고유명사(“丹溪心法”, “脾胃論”)에 포함된 빈도는 배제하였음.

13) 두 변수 사이의 상관관계를 표현하는 값으로 -1에서 1사이의 값으로 나타난다. 0에서 멀어질수록 두 변수 사이의 상관관계가 높다고 해석한다.

www.kci.go.kr

Fig. 1a. Keyword word frequency distribution between Donguibogam(DO) and Uihagibmun(UI)

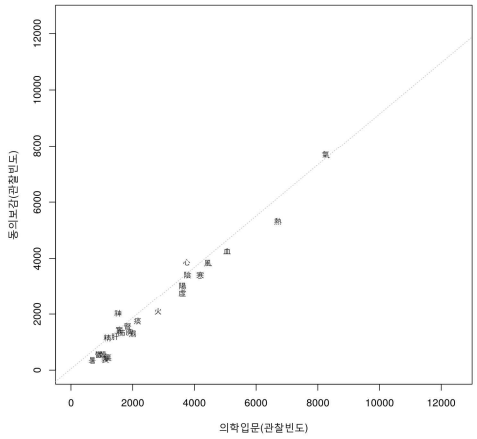


Fig. 1b. Keyword word frequency distribution between Donguibogam(DO) and Gyeongagjeonseo (GY)

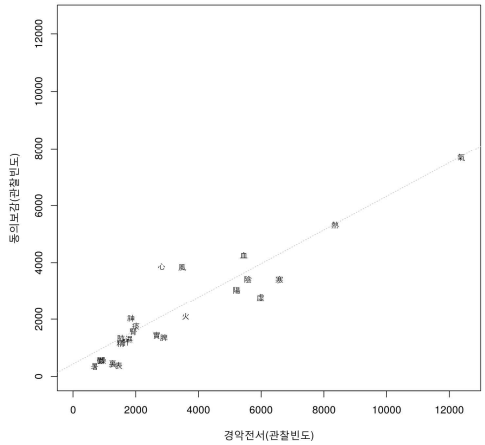


Table. 1. Keyword word frequency among Donguibogam(DO), Uihagibmun(UI), and Gyeongagjeonseo(GY).

핵심어	관찰빈도			상대관찰빈도					
	동의보감	의학입문	경악전서	동의보감		의학입문		경악전서	
肝	1164	1100	1359	1084.37	(1.34)	1278.32	(1.58)	1274.79	(1.576)
心	3835	3418	2489	3572.64	(4.416)	3972.07	(4.91)	2334.77	(2.886)
脾	1322	1563	2565	1231.56	(1.522)	1816.37	(2.245)	2406.06	(2.974)
肺	1299	1298	1185	1210.13	(1.496)	1508.41	(1.865)	1111.57	(1.374)
腎	1523	1496	1587	1418.81	(1.754)	1738.51	(2.149)	1488.66	(1.84)
精	1111	847	1181	1034.99	(1.279)	984.30	(1.217)	1107.82	(1.369)
氣	7649	7928	12045	7125.72	(8.808)	9213.16	(11.388)	11298.65	(13.966)
神	1991	1186	1512	1854.79	(2.293)	1378.26	(1.704)	1418.31	(1.753)
血	4220	4726	5105	3931.30	(4.859)	5492.11	(6.789)	4788.68	(5.919)
痰	1723	1827	1662	1605.13	(1.984)	2123.17	(2.624)	1559.02	(1.927)
鬱	527	576	549	490.95	(0.607)	669.37	(0.827)	514.98	(0.637)
陰	3386	3451	5246	3154.36	(3.899)	4010.42	(4.957)	4920.94	(6.083)
陽	3001	3280	4883	2795.70	(3.456)	3811.70	(4.712)	4580.43	(5.662)
表	352	774	1117	327.92	(0.405)	899.47	(1.112)	1047.79	(1.295)
裏	407	867	924	379.16	(0.469)	1007.54	(1.245)	866.75	(1.071)
寒	3373	3850	6241	3142.25	(3.884)	4474.10	(5.53)	5854.29	(7.236)
熱	5279	6371	8021	4917.85	(6.079)	7403.77	(9.152)	7523.99	(9.3)
虛	2749	3277	5654	2560.94	(3.166)	3808.22	(4.707)	5303.66	(6.556)
實	1404	1233	2332	1307.95	(1.617)	1432.88	(1.771)	2187.50	(2.704)
風	3805	4107	3159	3544.69	(4.382)	4772.76	(5.9)	2963.26	(3.663)
暑	313	352	352	291.59	(0.36)	409.06	(0.506)	330.19	(0.408)
濕	1265	1671	1448	1178.46	(1.457)	1941.88	(2.4)	1358.28	(1.679)
燥	519	701	598	483.49	(0.598)	814.64	(1.007)	560.95	(0.693)
火	2058	2485	3257	1917.21	(2.37)	2887.83	(3.57)	3055.19	(3.776)

※ 음영 부분은 의서 별 빈도 상위 5개 값.
상대관찰빈도는 3종 의서의 텍스트 길이 평균값으로 정규화한 결과. (괄호 안은 1,000자 당 출현 빈도)

www.kci.go.kr

2) 주제별 핵심어 빈도

각각의 핵심어를 그룹에 따라 비교한 결과이다(상대관찰빈도 기준, 노란색은『동의보감』, 빨간색은『의학입문』, 초록색은『경악전서』). 전체적으로 오장 용어의 빈도가 상대적으로 낮고 팔강 용어의 빈도가 높았으나『경악전서』의 경우에는 그 차이가 더 두드러졌다(Fig. 2a). 오장 용어를 보면,『동의보감』과『의학입문』에서 心의 빈도가 높았던 반면,『경악전서』에서는 脾의 빈도가 높았다(Fig. 2b). 육기의 경우에는 熱의 빈도가 공통적으로 높았고,『경악전서』에서는 寒의 빈도가 두드러지게 높았던 반면『동의보감』과『의학입문』에서는 寒의 빈도가 상대적으로 낮고 風의 출현이 눈에 띈다(Fig. 2c). 정기신혈담울 용어에서는 세 서적 모두 氣의 쓰임이 압도적이었고 특히『경악전서』에서 이런 현상이 더 뚜렷했다. 또『경악전서』에서 血의 쓰임이 상대적으로 적었다(Fig. 2d). 팔강 용어의 경우, 세 서적 모두 비슷한 형태를 보이고 있으나『경악전서』에서 특징적으로 寒이 자주 출현하였다(Fig. 2e, Fig. 2f).

전반적으로『의학입문』과『동의보감』이 유사한 패턴으로 빈도가 표현된 반면,『경악전서』에서는 이와 다른 모습으로 나타났다.

Fig. 2a. Frequency pattern by topic

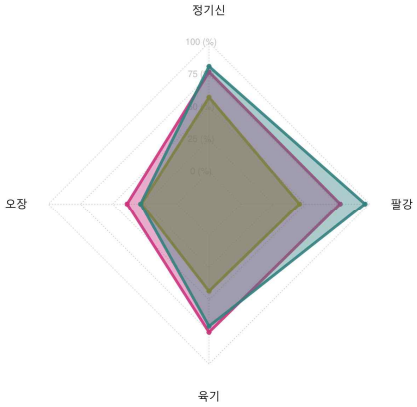


Fig. 2b. Frequency pattern by Five viscera

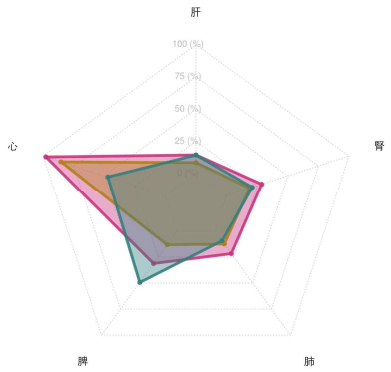


Fig. 2c. Frequency pattern by Six qi

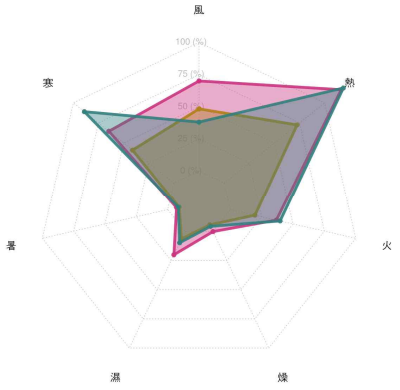
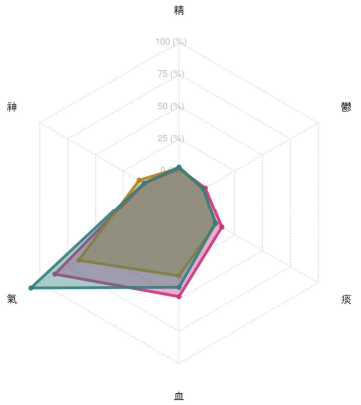


Fig. 2d. Frequency pattern by fluid and humor



www.kci.go.kr

Fig. 2e. Frequency pattern by Eight principles 1

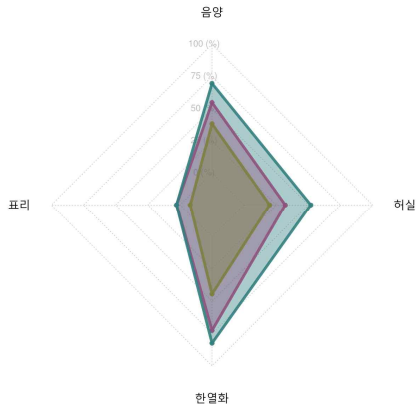
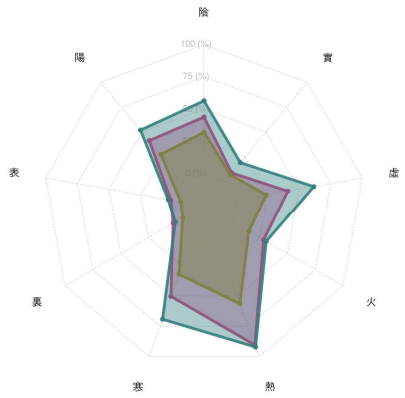


Fig. 2f. Frequency pattern by Eight principles 2



3) 의서 간 핵심어 빈도 차이

빈도의 차이를 상대적으로 비교하기 위해 χ^2분석으로 도출한 표준화 잔차(standardized residual) 값을 비교하였다. 잔차(R)는 기대빈도(O)[14]와 관찰빈도(E)의 차이[2]이며, 표준화 잔차(SR)는 잔차(R)를 기대빈도의 제곱근으로 나눈 값[3]이다. 이 값은 기대빈도에 비해 얼마나 차이가 나는지를 보여준다. 관찰빈도와 기대빈도 사이에 차이가 크다면 우연히 나타나기 어려운 결과로, 해당 변수 값이 어떤 힘에 의해 억제되었거나 강화되었다고 할 수 있다.

Fig. 3은 이러한 표준화 잔차와 기대빈도 값으로 그려진 그래프로, x축은 기대빈도의 제곱근 값이며 y축은 χ^2의 표준화 잔차(SR)이다. 따라서 x값과 y값으로 만들어지는 사각형의 면적이 잔차(R) 즉, 관찰빈도와 기대빈도의 차이라고 할 수 있으며, 중심선에서 위와 아래로 멀어질수록 상대적인 차이가 더 크다고 이해할 수 있다.

『의학입문』과 비교하였을 때 『동의보감』에서는 대부분 0을 중심으로 분포하고 있었고 차이가 있다 하더라도 神을 제외하고 ±10 안에 관찰되었다(Fig. 3a). 대부분의 핵심어들이 『의학입문』과 『동의보감』에서 비슷한 비율로 출현하였음을 의미한다. 차이가 있는 핵심어로는 神이 『동의보감』에서 더 두드러지게 나타났고 表·裏는 『의학입문』에서 상대적으로 더 사용되었다.

『경악전서』와 비교했을 때 『동의보감』의 용어들은 차이가 상대적으로 더 크게 나타났다. 대부분의 핵심어들이 0에서 벗어나 있어 두 서적에서 서로 다른 비율로 사용되었음을 보여준다(Fig. 3b). 특히 『동의보감』에서 心·風·神이 더 많이 사용되었고, 虛·寒·氣와 같은 핵심어들은 덜 사용되었다.

이와 같이 표준화 잔차(SR)로 보았을 때, 『동의보감』은 『의학입문』 보다 『경악전서』와 비교했을 때 상대적으로 핵심어의 빈도가 더 크게 차이가 났다.

$$R = O - E \quad [2]$$

$$SR = \frac{O - E}{\sqrt{E}} \quad [3]$$

14) 기대빈도는 확률적으로 기대되는 결과이다. 동전 던지기의 예를 들면, 100번을 던진다고 하였을 때 동전 앞면이 나올 확률이 0.5이므로 이 가운데 50회 앞면이 나올 것을 기대할 수 있다.

www.kci.go.kr

Fig. 3a. Keyword frequency characteristic of Donguibogam(DO) contrast Uihagibmun(UI)

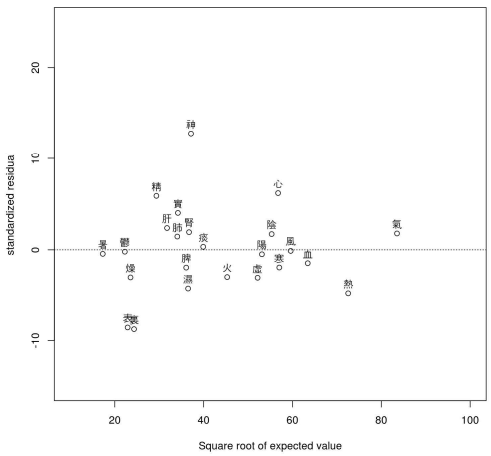
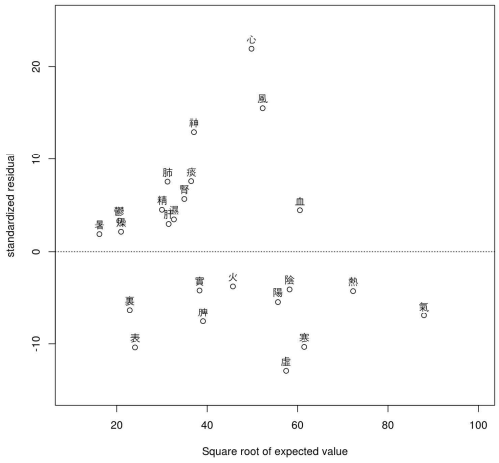


Fig. 3b. Keyword frequency characteristic of Donguibogam(DO) contrast Gyeongagjeonseo(GY)



3.2. 동질성과 유사성 측정 결과

본 연구에서는 두 텍스트의 의학 사상을 핵심어 빈도를 기반으로 CBDF를 통해 비교해 보았다.

1) 의서 간 동질성 비교

의서 간 비교에 앞서 한 서적 내에서 핵심어들이 얼마나 균등하게 분포하였는지 검토해 보았다. 동질성(homogeneity)은 서적 내에 핵심어가 얼마나 균질하게 사용되었는가를 나타낸다. 예를 들어 핵심어가 부분마다 사용된 비율이 크게 차이 난다면 동질성은 낮아지게 되고, 반대로 모두 비슷하다면 동질성은 높아지게 된다. 동질성은 유사성을 의미하고 CBDF는 이질성을 나타내는 지표이기 때문에 CBDF 값이 높을수록 동질성이 낮다고 할 수 있다.

서적의 동질성은 하나의 서적을 반으로 나누어 둘 사이에 CBDF를 구하는 방법으로 이루어졌다. 서적을 반으로 나누기 위해 글자 수를 균등하게 하여 텍스트를 100개의 조각으로 나누고, 이를 임의로 50개씩 묶어 1쌍으로 만들었다. 이렇게 만들어진 텍스트 사이에 CBDF를 구하였다. 이를 1만 번 반복하여 서적마다 1만개의 CBDF 값의 분포를 비교해 보았다.

동질성 평가 결과 『의학입문』이 가장 동질성이 높고, 『동의보감』이 가장 낮은 것으로 나타났다(Fig. 4). 평균값(괄호 안은 표준편차)은 『동의보감』이 23.28(7.95)로 가장 컸고, 『경악전서』가 19.03(7.16), 『의학입문』이 14.23(6.66) 순이었다[15]. 『동의보감』에서 부분마다 사용된 핵심어의 비율 격차가 『의학입문』이나 『경악전서』보다 더 크다는 의미이다. 앞서 핵심어 빈도 패턴과 빈도의 상대적 차이 비교에서 『동의보감』과 『의학입문』이 상당히 유사한 결과를 보인데 반해, 동질성에 있어서는 차이를 보인 셈이다.

2) 의서 간 의학 사상의 차이

다음으로 『의학입문』, 『동의보감』, 『경악전서』을 둘씩 짝지어 CBDF를 구하였고, 이 수치를 바탕으로 다차원척도법을 통해 상대적인 분포를 추상화 하였다. 다차원척도법(multi-dimensional scaling ; MDS)은 개체(node) 사이의 거리가 주어지는 경우 이들 사이의 상대적인 분포를 2차원 평면에 나타내는 대표적인 방법이다. 왜곡이 있기는 하지만 거리를 통해 개체의 분포를 추상화하는데 도움을 준다.

CBDF 값은 『동의보감』과 『의학입문』은 37.72, 『

15) 여기서 CBDF 값은 확률적인 의미를 가지지 못하며, 상대적인 척도로서의 의미만 지닌다. 값의 크기가 상대적으로 클수록 두 개체 간의 차이가 '크다'고 해석된다. 거리 관점에서는 '멀다'고 할 수 있다. 이하 같다.

www.kci.go.kr

동의보감』과 『경악전서』는 130.30, 그리고 『의학입문』과 『경악전서』는 99.66이었다. 이를 거리로 보고 다차원척도법으로 2차원 평면에 나타낸 그래프(Fig. 5)에서 『동의보감』은 『경악전서』보다 『의학입문』에 더 근접하게 나타났다.

Fig. 4. Homogeneity of three kinds of books

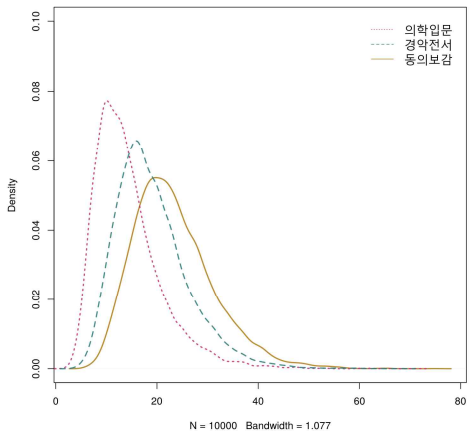
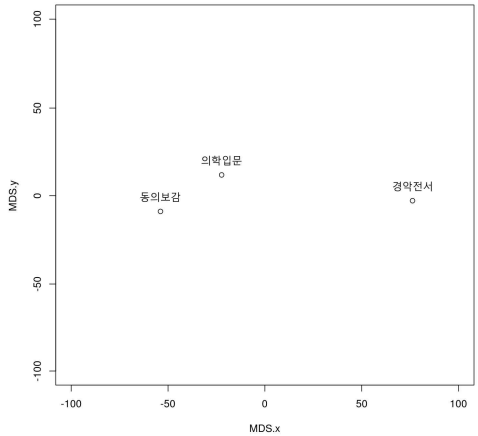


Fig. 5. Relative position of three kinds of books



4. 고찰

『동의보감(東醫寶鑑)』(1610), 『의학입문(醫學入門)』(1575), 『경악전서(景岳全書)』(1636-1640?)는 조선에서 가장 영향력 있었던 의서이다. 한국의학사를 깊이 연구했던 김두종은 『동의보감』을 중심으로 임상을 펼쳤던 '보감파(寶鑑派)'와 『의학입문』을 기준으로 삼았던 '입문파(入門派)'의 존재를 언급하며 조선시대 학풍을 설명한 바 있다[16]. 또 20세기 초에 전통적인 방식으로 한의학을 전수받고 활약했던 청강 김영훈(1882-1974)은 "이론적 근거는 『의학입문』에서 찾고 방약(方藥)은 『동의보감』을 따르라"고 후학들을 지도하였다고 한다[17]. 『경악전서』는 비록 조선 후기에 유입되었으나 『의문보감(醫門寶鑑)』, 『제중신편(濟衆新編)』, 『의종손익(醫宗損益)』, 『방약합편(方藥合編)』 등 18세기 이후 조선 의서들에 적지 않은 영향을 주었다.

이들 서적에 대하여 전문가 주도로 정성적 분석 방법에 따라 이루어진 선행 연구 결과는 다음과 같이 요약된다. 첫째(①), 『동의보감』은 『의학입문』으로부터 영향을 크게 받았으나 편찬 의도와 집필 방향은 차이가 있다. 『동의보감』에 표시된 출전 표기를 기준으로 하였을 때 『의학입문』은 『동의보감』에서 『본초』에 이어 두 번째로 많이 인용된 서적이다[18]. 세부적인 내용에서도 그러한데, 조학준은 요통(腰痛)에 대한 『동의보감』의 의론이 『의학입문』의 영향을 크게 받았다고 하였다[19]. 『동의보감』이 『의학입문』의 영향을 받았으나 전문가들은 두 서적의 차이도 분석하였다. 차웅석은 "『동의보감』은 당대 중국의학의 성과를 체화하는 과정에서 최신 의서인 『의학입문』을 다수 인용했기 때문에 기본적인 의학 개념은 거의 같다고 할 수 있으나 『의학입문』이 음양으로 인체의 생리병리를 이해하려 하였던 반면 『동의보감』은 정기신의 관점이 근간을 이룬다."[20]고 지적하였고, 홍세영 등은 두 의서의 중요성을 재확인하면서도 "『동의보감』과 달리 『의학입문』은 기초

16) 김두종. 한국의학사(전). 서울. 탐구당. 1981. p.264.
17) 김영훈 저, 이종형 편. 晴崗醫鑑. 성보사. 1984. p.476.
18) 김중권. 『동의보감』의 문헌적 연구 – 인용문헌을 중심으로. 서지학연구. 1995. 11. pp.207-243.
19) 조학준. 『의학입문』 요통의 문헌 근거 –『동의보감』 요통과 비교–. 대한한의학원전학회. 2014. 27(2). pp.119-135.
20) 차웅석. 중국의 『醫學入門』이 한국의 『東醫寶鑑』에 미친 영향. 한국의사학회지. 2000. 13(1). pp.111-127.

www.kci.go.kr

이론이 착실하게 정리되어 있어 이론서로서의 측면이 두드러진다."[21]고 평가하였다. 요컨대 두 서적은 조선에서 가장 애독되었던 의서로서, 『동의보감』은 『의학입문』에 영향을 다분히 받았으나 양자는 서로 다른 편찬 의도와 집필 방향을 가지고 있었다.

둘째(②), 『의학입문』은 단계학파(丹溪學派)의 영향을 받아 성립되었고, 『경악전서(景岳全書)』(1636-1640?)는 온보학파(溫補學派)를 대표하는 서적으로 의학사상에서 차이가 있다. 『의학입문』은 자음(滋陰)을 중시하는 단계학파의 영향을 받아 성립된 저작이다[22]. 그에 반해 『경악전서』의 저자 장개빈(張介賓, 1563-1640)은 기존의 자음(滋陰) 학설들을 강하게 비판하고 온보(溫補)를 강조한 대표적인 인물이다[23]. 이를 통해 『경악전서』는 『동의보감』과 의학사상의 차이가 상대적으로 클 것으로 추측할 수 있다. 『경악전서』는 1713년에서 1724년 사이에 조선에 유입되어[24] 『동의보감』과는 서로 직접적인 영향을 주고받을 수 없었으나, 『의문보감』, 『제중신편』 등 18세기 이후 조선 의서들은 『동의보감』의 내용을 근간으로 하면서도 『경악전서』의 내용을 추가로 인용하였다. 이러한 역사적 사실은 『경악전서』의 내용이 『동의보감』과 어느 정도 차별성을 지닌다는 점을 방증한다.

셋째(③), 세 서적은 서술 방식에 차이를 가진다. 『의학입문』과 『경악전서』는 단일 저자에 의해 서술된 서적이고, 책의 구성에 있어서 의론 부분과 처방 부분이 나뉘어져 있다. 이에 반해 『동의보감』은 허준의 저작이기는 하지만 낱낱의 문장들은 근거 문헌의 것을 보존하려고 노력하였고, 구성에 있어서도 앞의 책과 달리 의론 중간에 관련된 처방을 함께 수록하였다.

본 연구는 의서의 의학 사상이 용어 사용 양상을 통해 드러난다는 가정을 확인하고자 수행되었다. 이러한 가정 아래 의학 사상은 핵심어 빈도의 유사도(similarity)로 유비된다. 따라서 기존 연구결과(①, ②)를 통해 『동의보감』과 『의학입문』의 의학 사상의 유사도는 『동의보감』과 『경악전서』의 그것에 비해 높을 것으로 예상된다.

본고에서 수행한 계량 분석 결과는 이러한 예상과 부합한다. 단순 빈도 비교 결과 핵심어 빈도의 상관계수는 『동의보감』-『의학입문』이 0.98, 『동의보감』-『경악전서』가 0.93로 비록 차이는 적지만 전자가 후자에 비해 높게 나타났다(Fig. 1). 주제별 핵심어 빈도에서는 전자의 분포 패턴이 후자에 비해 현저히 유사하게 나타났다(Fig. 2). 의서 간 핵심어 빈도 차이를 통한 분포에서도 전자가 후자에 비해 유사하게 나타났다(Fig. 3). CBDF를 이용한 서적 사이의 유사도 비교에서도 『동의보감』과 『의학입문』은 37.72, 『동의보감』과 『경악전서』는 130.30, 그리고 『의학입문』과 『경악전서』는 99.66로 나타났다(Fig. 5).

한편 의서의 구성, 집필 의도 및 방법은 사용된 핵심어의 편중으로 나타날 수 있다. 본고에서 수행한 계량 분석 결과 이러한 차이를 포착할 수 있었다. 동질성(homogeneity) 평가 결과 『의학입문』이 14.23(6.66), 『경악전서』가 19.03(7.16), 『동의보감』이 23.28(7.95) 순으로 『동의보감이』이 가장 동질하지 않게 나타났다. 이러한 결과는 『의학입문』과 『경악전서』의 경우 한 명의 저자에 의해 상대적으로 일관되게 서술된 반면, 『동의보감』의 경우 허준이 편찬을 하였지만 낱낱의 기록들은 다양한 서적에서 인용하는 방식을 택했다는 사실(③)과 부합한다[25]. 다만 『의학입문』과 『경악전서』에서 처방을 한데 모아 둔 것에 반해 『동의보감』에는 처방이 본문 중간 중간 섞여 있어 텍스트가 상대적으로 균질할 것으로 예상 되었으나 이에 대한 영향을 크지 않았다.

21) 홍세영 外 3인. 임상학습서로서의 醫學入門 의 가치. 대한한의학원전학회지. 2016. 29(1). pp.125-137.
22) 차웅석. 李梴 醫學思想의 學術系統 및 特徵에 대한 硏究(박사학위논문). 서울. 경희대학교. 2001. pp.1-61.
23) 陳大舜 외 저, 맹웅재 외 역. 各家學說(中國篇). 서울. 대성의학사. 2001. pp.322-324.
24) 하기태. 『경악전서』가 조선후기 한국의학에 미친 영향에 대한 연구. 대한한의학회지. 1999. 20(2). pp.12-26.
25) 이러한 추정은 『동의보감』과 유사한 방식으로 서술된 서적과의 비교를 통해 확인해 볼 수 있으나 본고의 주제를 넘어선 것이기 때문에 후속 연구로 남겨 둔다.

www.kci.go.kr

5. 결론

지금까지 의서 속에 내재된 의학 사상을 계량적인 방법으로 비교할 수 있는지 검토하기 위해『동의보감』과『의학입문』,『동의보감』과『경악전서』의 유사성을 비교해 보았다. 이를 위해 의학 사상이 텍스트 내에서 주요 핵심어의 빈도로 환원될 수 있다는 가정 아래 한의학에서 생리와 병리를 설명하기 위해 보편적으로 사용되는 정기신혈담울(精氣神血痰鬱), 오장(五臟), 육기(六氣), 팔강(八綱)의 4가지 주제 24가지 용어를 핵심어로 선정하였다. 그리고『동의보감』,『의학입문』,『경악전서』에서 이 핵심어들의 빈도를 측정하여 서로 비교하였다.

핵심어는『경악전서』에서 전체 글자 수 대비 8.63%으로 가장 높게 나타났으며,『의학입문』 8.39%,『동의보감』 6.25%로 나타났다. 이들 빈도를 각 주제별로 나누어 패턴을 살펴보면,『동의보감』이『의학입문』과 상당히 유사한 패턴을 보이는데 반해『경악전서』와는 心, 風, 神, 虛 등에서 전혀 다른 패턴을 보였다. 이러한 사실은 관찰빈도와 기대빈도의 차이를 계측하는 표준화 잔차(standardized residual) 비교를 통해서도 확인할 수 있었다.

χ^2값(Chi-Square Statistic)을 기반으로 계량 분석한 결과에서, 서적 내 핵심어들은『동의보감』에서 가장 이질적으로 분포하였고『의학입문』에서 가장 고르게 분포하였다. 같은 방법으로 분석한 유사도 비교에서『동의보감』과『의학입문』은『동의보감』과『경악전서』에 비해 월등히 유사하게 나타났다.

이상을 통해 의학 사상이 텍스트 내에서 주요 핵심어의 빈도로 환원될 수 있다는 가정 하에 계량적 분석을 통해 도출된 결과와 기존 전문가들이 도출해 낸 정성적 연구 결과가 서로 부합한다는 점을 확인할 수 있었다. 단순 빈도 패턴과 서적 간 유사도 결과는『동의보감』과『의학입문』이 유사하게 나타났으며, 동질성에서는 차이가 나타났다. 이러한 결과는『동의보감』이『의학입문』의 영향을 받았으나 두 서적의 구성과 집필 방식이 다르다는 전문가들의 정성적 분석 결과와 부합한다. 또『동의보감』과『경악전서』의 유사도가 상대적으로 멀리 나타났다. 이는『경악전서』가『동의보감』 이후에 조선에 영향을 주었다는 역사적 사실, 그리고『동의보감』이 단계학파의 저작으로 구분되는『의학입문』의 영향을 많이 받은 반면『경악전서』는 온보학파의 저작으로 구분된다는 기존 연구 결과와 부합한다.

본 연구는 제한된 서적만을 대상으로 하였으며, 경험적 판단에 의해 핵심어를 도출하였고, χ^2값 이외의 방법을 검토하지 못했다는 점 등 분명한 한계를 지니고 있다. 그러나 이상의 결과는 의학 사상이 텍스트 내에서 주요 핵심어의 빈도로 환원될 수 있다는 가정을 지지해 준다. 단, 일반화된 결론에 이르기 위해서는 계속적인 후속 연구가 필요하다.

감사의 말씀

본 연구는 한국한의학연구원 주요사업 “한의 고문헌 데이터 구축 및 분석 도구 개발(K18201)”의 지원을 받아 수행되었습니다.

References

1. Kim DJ. Korean Medical History. Seoul. Tamgudang. 1981.
김두종. 한국의학사(전). 서울. 탐구당. 1981.
2. Kim YH original work. Lee JH ed.. Cheonggang-uigam. 1984.
김영훈 저, 이종형 편. 晴崗醫鑑. 성보사. 1984.
3. Chen DS original work. Meng WJ et al. trans.. Various Schools (China). Seoul. Daeseong-uihagsa. 2001.
陳大舜 외 저, 맹웅재 외 역. 各家學說(中國篇). 서울. 대성의학사. 2001.
4. Kim KW, Kim TY, Lee BW. Analysis of Prescriptions from Taepyeonghyeminhwajegukbang, Somunsunmyungronbang and Nansilbijang based on Herb weight ratio grade. The Journal Of Korean Medical Classics. 2014. 27(4).

www.kci.go.kr

김기욱, 김태열, 이병욱. 본초 비율의 순위를 이용한 문헌의 특징 분석 방법 - 태평혜민화제국방, 난실비장, 소문선명론방을 중심으로. 대한한의학원전학회지. 2014. 27(4).

5. Kim JG. A literary research of Dong-uibogam. The Journal Of Bibliographic Research. 1995. 11.
김중권. 『東醫寶鑑』의 文獻的 硏究 - 인용문헌을 중심으로. 서지학연구. 1995. 11.
6. Oh WH et al. Feature Comparison by Prescription Configuration Analysis among Liuhejian's and Lidongyuan's Books and『Hejijufang』. The Journal Of Korean Medical Classics. 2015. 28(1).
오월한 外 3인. 방제구성을 이용한 유하간(劉河間) 및 이동원(李東垣)의 저작과 『화제국방(和劑局方)』의 특성 비교. 대한한의학원전학회지. 2015. 28(1).
7. Lee JH. Problems with Chinese Ideographs Search in Unicode and Solutions to Them. Informatization policy. 2012. 19(3).
이정현. 유니코드 한자 검색의 문제점 및 개선방안. 정보화정책저널. 2012. 19(3).
8. Jo HJ. Reference Research on Low Back Pain in Uihagimmun - Compared with Donguibogam -. The Journal Of Korean Medical Classics. 2014. 27(2).
조학준. 『의학입문』 요통의 문헌 근거 -『동의보감』 요통과 비교 -. 대한한의학원전학회. 2014. 27(2).
9. Cha WS. A Study of Yi Chenon's Medical ideology and Research System. Seoul. Kyunghee Univ. 2001.
차웅석. 李梴 醫學思想의 學術系統 및 特徵에 대한 硏究(박사학위논문). 서울. 경희대학교. 2001.
10. Cha WS. Comparative studies on『Eui Hak Ip Mun』and『Dong Eui Bo Gam』. Korean Society of Medical History. 2000. 13(1).
차웅석. 중국의 『醫學入門』이 한국의 『東醫寶鑑』에 미친 영향. 한국의사학회지. 2000. 13(1).
11. Ha KT. A Study on the Influence of Jingyuequanshu to Korean Medicine in the late Chosun dynasty. Journal of Korean Medicine. 1999. 20(2).
하기태. 『경악전서』가 조선후기 한국의학에 미친 영향에 대한 연구. 대한한의학회지. 1999. 20(2).
12. Hong SY et al.. Examining Distinctive Points of Introduction to Medicine(Yixuerumen) through the Clinical Window. The Journal Of Korean Medical Classics. 2016. 29(1).
홍세영 外 3인. 임상학습서로서의 醫學入門의 가치. 대한한의학원전학회지. 2016. 29(1).
13. Adam Kilgarriff, Raphael Salkie. Corpus similarity and homogeneity via word frequency. Proceedings, EURALEX(Gothenberg, Sweden). 1996.
14. Adam Kilgarriff. Using Word Frequency Lists to Measure Corpus Homogeneity and Similarity between Corpora. In Language Engineering for Document Analysis and Recognition. Proceedings, AISB Workshop(Falmer). 1997.
15. Pavel Pecina. Lexical association measures and collocation extraction. Lang Resources & Evaluation. 2010. 44.
16. KIOM. Mediclassics. Daejeon. KIOM. 2013 [cited on June 08, 2017]: Avaiable form: https://mediclassics.kr/
한국한의학연구원. 한의학고전DB. 대전. 한국한의학연구원. 2013 [cited on June 08, 2017]: Avaiable form: https://mediclassics.kr/

www.kci.go.kr