

第 125 回人文科学とコンピュータ研究会発表会
テキストマイニングを用いた古典漢詩イメージと語義分析
—中国・朝鮮『「鸚鵡洲」詩歌』を対象として—

林政君 (リン ブンクン)

論文要旨

「鸚鵡洲」とは、中国の地名、湖北省漢陽区の西南、揚子江中の中州。漢陽区は中国湖北省武漢市に位置する市轄区。『三国志』の時代、即ち、後漢末以降で、政治的にも軍事的にも重要な舞台となり、後の詩人たちの関心を集めるようになる。昔は荊州の劉表と孫呉との争い、狂士・禰衡の故事は、この間出来事である（『後漢書』「文苑列伝」第七十）。当初、黄祖は禰衡を歓待し、禰衡も彼の長子・黄射と親密に交際した。ある時、宴会の席上（『太平御覧』巻六十九所引「江夏記」では鸚鵡洲でのこととする）で、鸚鵡を献上する者があったため、射の命ずるままに、「鸚鵡賦」を詠む。「文無加點、辭采甚麗」という出来栄であった。後に、文学作品において鸚鵡が詠じられる際、その多くがこの「鸚鵡賦」を踏まえている。唐の崔顥の詩に「晴川歴歴漢陽樹、芳草萋萋鸚鵡洲」は一つ例である。なお、韓国において李氏朝鮮時代、平安道の三登に中国の「黄鶴楼」を模倣と建立し、また、対向の山壁の上に「鸚鵡洲」を題名することである。

本研究のタイトルなどでも用いている『「鸚鵡洲」詩歌』ということばは、上記の中国と韓国「鸚鵡洲」を詠む詩歌を総称するものであるが、「鸚鵡洲」この地名と典故を古典漢詩に引用されたことは多い、周知の事実であるが、古典漢詩の研究分野を振り返って、『「鸚鵡洲」詩歌』という呼称は用い具体的な事例がないし、「鸚鵡洲」に関する研究事例も従来あまりなかった。しかし、本論を通して、『「鸚鵡洲」詩歌』というのは、中国で成立したことば・典故・概念を借用した便利的な呼称である。また、中国と朝鮮の漢詩の中に、そのような事例が存在することを明らかにすることが出来た。そして、デジタル人文学も本研究の目指す目標とする。

本論の分析対象は中国古典籍と韓国文集叢刊の中に、「鸚鵡洲」を主題する詩歌を集まる、データを整備する。テキストマイニング (text mining) 分析には KH Coder (樋口, 2014) を使用した。最初にデータセットを形態素解析の分かち書き処理により単語に分解した。KH Coder ではこの操作を前処理と呼ぶ。前処理により総抽出語数と異なり語数、そして助詞・助動詞を除いた総抽出語数と異なり語数が明らかとなる。続いて前処理を実行されたデータから頻出 100 語の抽出語リストを作成した。このリストで示される数値は、本研究の場合、抽出語の出現回数 (term frequency : TF) と抽出語を含む文書数 (document frequency : DF) を分析し、データを整理して、中国古典詩から朝鮮漢詩までの「鸚鵡洲」この地名・典故広範の使用と転換にも複数の実例があると確かめられたことで、「鸚鵡洲」のイメージと意義の爆発的な隆盛がどのようにもたらされたかについての解明が進んだ。その他、本論で使用した分析方法は、視覚的なデータ分類を可能にする対応分析 (コレスポンデンス分析)、共起関係にある語彙元素を線で結ぶ共起ネットワーク、特徴的な語をリストアップする Jaccard・Cosine・Euclid 三つ分析結果を解析する。分析方法の詳細については樋口 (KH Coder 3 リファレンス・マニュアル,

2014) を参照されたい。

さらに、『「鸚鵡洲」詩歌』について、中国と韓国資料の収集と調査を行ってきた筆者の経験に照らして、古典漢詩の中に、「鸚鵡洲」のイメージと意義、「詩の古跡」と典故の生成などを考察と構築して、人文学の直接調査することと情報学の立場からそれを解決するソフトウェアの重要性およびそれに伴うデジタル人文学の可能性について、具体的なテキストマイニング研究事例をもとに報告したい。

キーワード：デジタル・ヒューマニティーズ(デジタル人文学、人文情報学)、テキストマイニング、古典漢詩、鸚鵡洲、「鸚鵡洲」詩歌、詩のイメージ

1. はじめに

周知の通り、現在の「武漢」(1949以降)は、もともと市であった武昌區(長江南岸)、漢口區(長江北岸漢水以東)、漢陽區(長江北岸漢水以西)の三地区を合わせて、一大都市を形成している。「鸚鵡洲」とは、中国の地名、湖北省漢陽区の西南、揚子江中の中州。漢陽区は中国湖北省武漢市に位置する市轄区。『三国志』の時代、即ち、後漢末以降で、政治的にも軍事的にも重要な舞台となり、後の詩人たちの関心を集めるようになる。昔は荊州の劉表と孫呉との争い、狂士・禰衡の故事は、この間出来事である(『後漢書』「文苑列伝」第七十)。当初、黄祖は禰衡を歓待し、禰衡も彼の長子・黄射と親密に交際した。ある時、宴会の席上(『太平御覽』卷六十九所引「江夏記」では鸚鵡洲でのこととする)で、鸚鵡を献上する者があったため、射の命ずるままに、「鸚鵡賦」を詠む。「文無加點、辭采甚麗」という出来栄であった。後に、文学作品において鸚鵡が詠じられる際、その多くがこの「鸚鵡賦」を踏まえている。唐の崔顥の詩に「晴川歴歴漢陽樹、芳草萋萋鸚鵡洲」は一つ例である。なお、韓国において李氏朝鮮時代、平安道の三登に中国の「黄鶴楼」を模倣と建立し、また、対向の山壁の上に「鸚鵡洲」を題名することである。

本研究では、中国古典籍と韓国文集叢刊の中に、「鸚鵡洲」を詠んだ詩歌作品を収める、データを整備し、テキストマイニング(text mining)分析にはKH Coder(樋口, 2014)を使用した。この際、古典漢詩の複雑なテーマ・イメージ・歴史典故が問題となったが、KH Coderを用いた既に研究は触れない分野を発見した。

デジタル技術を用いることで、中国・朝鮮『鸚鵡洲詩歌』の抽出語リスト、KWIC コンコーダンス、関連語検索、対応分析、多次元尺度構成法、頻出語・共起ネットワークという形でテキスト情報を可視化し、文字情報のみ得られない視点からテキスト解釈を行うことができ、こうした分析と、古典漢詩の精読の間を行き来することで、これまでよりも深く、また多角的な人文学と情報学と対話することが可能になるだろう。

2. 「鸚鵡洲」詩歌」テキストマイニング研究に必要なこと

2.1 分析対象と研究方法

まず、中国・韓国『「鸚鵡洲」詩歌』データの整備をする。本論で対象とした詩歌は、中国と朝鮮の古典漢詩から、「鸚鵡洲」を詠んだ詩歌である。資料の来源は書籍『明刻黄鶴樓集校注』を選択と電子データ入力した、また中央研究院歴史語言研究所を開発する『漢籍電子文獻資料庫』と北京愛如生数字化技術研究センターを開発する『中國基本古籍庫』と KRpia Korean Database『韓国文集叢刊』からキーワード「鸚鵡洲」検索し、テキストコピーした。全て資料を参照しながら入力と整理し、中国詩 80 首と韓国李氏朝鮮時代の漢詩 32 首データベースを作成した、分析対象とした。

また、プロジェクトを作成した際に(Excel形式)、入力と言語単位認定に際して以下の点に留意した:

- (1) 1行目に列の名前を入力(編次、書籍番号、朝代、作者、書目、巻次、題名、詩歌内容、版本)。

- (2) 2行目からデータを入力する。
- (3) 1つ目のシートに入力する。
- (4) データ量が多くて KH Coder でエラーになる場合、CSV 形式で保存してから KH Coder へ処理する。
- (5) Excel で 1つの列にテキストを入力し、ほかの列にテキスト以外の情報を入力。テキスト以外の情報のことを「外部変数」と呼ぶ。KH Coder には、分析対象のテキストとその外部変数とのリスト (Excel や CSV) をテキスト形式に自動変換して、外部変数と共に入力する。
- (6) 1つプロジェクトの中に、1つ種類の言語を入力だけ、例え、『韓国文集叢刊』のデータベースは漢語 (漢字) と韓国語二つである。テキスト自動変換と前処理で順調に、プロジェクトの中に漢語を入力しか。そして、中国語の場合について、Stanford POS Tagger を使用した、文の分割を行うために句点「。」で文を区切りとした。
- (7) 普通の電子データについて、バージョンの概念がないので、一つ詩歌はほとんど一つ詩例であるだけ、本論は漢籍書誌学と中国文献研究の特徴を考えており、詩歌の違うバージョンを含めて、データベースを作成した。

2.2 前処理の実行

上述した指定作業の後、KH Coder で「前処理の実行」を行った。分析対象ファイルの文章から語を切り出し、その結果をデータベース化する。中国・韓国『鸚鵡洲詩歌』データについて、分析に使用する語は、抽出された語数を下記表 1 に示す。

表 1. 中国・韓国鸚鵡洲詩歌抽出された語数

	総抽出語数 (使用)	異なり語数 (使用)	文書の単純集計 (文/段落/H5)
中国	4786 (3247)	2660 (2509)	572/572/80
韓国	2289 (1488)	1354 (1247)	241/241/32

3. テキストマイニング分析結果

中国・韓国『鸚鵡洲詩歌』で抽出語の合計は 4786・2289 (中・韓)、異なり語数で 2660・1354 (中・韓) であった (表 1)。原始資料の上位 100 位までの頻度順位を表 2 に示す。

3.1 抽出語リスト

中国・韓国『鸚鵡洲詩歌』の抽出語リストで、KH Coder の解析結果を元に、誤解析の部分を修正し、例え、普通名詞と固有名詞の重複 (鸚鵡洲、鸚鵡など) 計算を排除し、特に実行の前に、KH Coder の古典中国語の解析状況には不明な点であるので、「頻度別抽出語」の計算方法は以下の過程を通りで行う：「KH Coder で最初の抽出語リストを作る (表 2)」→「人工検査」→「疑問な語彙が KWIC コンコーダンスを使って計算する (図 1)」→「関連語検索で語彙の分析を検討する (図 2・図 3)」→「使用しない語を入力する・語の取捨選択」→「抽出語リストを再計算」→「語の抽出結果確認する」、最後の結果は表 3 に示す。「使用しない語を入力する」について、この部分に入力された語は (例え、鸚鵡)、単に分析の対象から除外され

るだけでなく、存在しないものと見なされる。すなわち、文書に含まれる語の数を数えるといった処理の際にも無視される。そうすると新しい抽出語リストが計算できる。

この抽出語リストを検視し、中国と韓国の『鸚鵡洲詩歌』の共同の名詞が多く、「洲」「鸚鵡洲」「黃鶴樓」「月」「水」「醉」「愁」「仙人」と続いた、この状況を古典漢詩の判断し、韓国鸚鵡洲詩歌において、中国詩の典故やイメージや主題などの摂取例は枚挙に暇ありませんので、この中に「擬古・模倣」という可能性であるそうです。

表 2. 中国・韓国鸚鵡洲詩歌における頻度別抽出語・原始資料（上位 100）

中国・原始資料				韓国・原始資料			
抽出語	文書数 (h5)	抽出語	文書数 (h5)	抽出語	文書数 (h5)	抽出語	文書数 (h5)
不	33	到	4	江	8	兩地樓	2
洲	32	却	4	洲	8	南	2
*鸚鵡	22	又	4	*鸚鵡	8	*原隰陣流	2
已	13	古	4	使	6	古	2
人	10	在	4	水	6	古木	2
去	10	坐	4	不	5	名	2
未	10	大	4	亦	5	君	2
行	10	天地	4	去	5	君山	2
君	9	寒	4	川	5	吹	2
更	9	小	4	晴	5	回	2
月	9	尚	4	更	5	地	2
江	9	抱	4	能	5	城	2
*鸚鵡洲	9	故	4	入	4	墨	2
使	8	橫	4	又	4	壁	2
只	8	欲	4	可	4	天公	2
如	8	洞庭	4	坐	4	太白	2
愁	8	碧	4	山	4	如	2
秋	8	竟	4	白	4	孤	2
舟	8	芳草	4	花	4	家	2
黃鶴樓	8	赤壁	4	人	3	寒山寺	2
明月	7	路	4	但	3	寒江	2
水	7	頭	4	何	3	少	2
知	7	*鸚鵡洲	4	分	3	已	2
醉	7	乘	3	到	3	平生	2
亦	6	乾坤	3	却	3	得	2
共	6	事	3	客	3	悄然	2
武昌	6	令	3	日	3	悠悠	2

江上	6	但	3	月	3	悠然	2
留	6	何	3	江流	3	情	2
*鸚鵡洲	6	再	3	流	3	愁	2
前	5	出	3	深	3	扁舟	2
可	5	分	3	碧	3	揮	2
天	5	動	3	秋	3	斗	2
山	5	勢	3	秋色	3	日日	2
平生	5	南	3	舟	3	曾	2
悠悠	5	南山	3	*芳草	3	木	2
文章	5	古	3	*芳草	3	杏花	2
江南	5	埋	3	醉	3	水落	2
浪	5	堪	3	*鸚鵡洲	3	汀州	2
狂	5	多	3	黃鶴樓	3	江上	2
白	5	夜深	3	一	2	波	2
看	5	子	3	上天	2	浪	2
真	5	客	3	乃	2	海上	2
能	5	岸	3	也	2	渺然	2
酒	5	州	3	人生	2	皆	2
*鸚鵡	5	得	3	仙人	2	相逢	22
世	4	情	3	伏	2	禰衡洲	2
仙人	4	才	3	住	2	*縣郭孤烟 暝色浮	2
休	4	投	3	佳句	2	美人	2
作	4	故人	3	倒	2	與	2

表3. 中国・韓国鸚鵡洲詩歌における頻度別抽出語・整理後結果（上位100）

中国・資料整理後結果				韓国・資料整理後結果			
抽出語	文書数 (h5)	抽出語	文書数 (h5)	抽出語	文書数 (h5)	抽出語	文書数 (h5)
不	33	古	4	江	8	南	2
洲	32	在	4	洲	8	古	2
鸚鵡洲	29	坐	4	鸚鵡洲	7	古木	2
已	13	大	4	芳草	6	名	2
人	10	天地	4	使	6	君	2
去	10	寒	4	水	6	君山	2
未	10	小	4	不	5	吹	2
行	10	尚	4	亦	5	回	2
君	9	抱	4	去	5	地	2

更	9	故	4	川	5	城	2
月	9	橫	4	晴	5	墨	2
江	9	欲	4	更	5	壁	2
使	8	洞庭	4	能	5	天公	2
只	8	碧	4	入	4	太白	2
如	8	竟	4	又	4	如	2
愁	8	芳草	4	可	4	孤	2
秋	8	赤壁	4	坐	4	家	2
舟	8	路	4	山	4	寒山寺	2
黃鶴樓	8	頭	4	白	4	寒江	2
明月	7	乘	3	花	4	少	2
水	7	乾坤	3	鸚鵡	3	已	2
知	7	事	3	人	3	平生	2
醉	7	令	3	但	3	得	2
亦	6	但	3	何	3	悄然	2
共	6	何	3	分	3	悠悠	2
武昌	6	再	3	到	3	悠然	2
江上	6	出	3	却	3	情	2
留	6	分	3	客	3	愁	2
前	5	動	3	日	3	扁舟	2
可	5	勢	3	月	3	揮	2
天	5	南	3	江流	3	斗	2
山	5	南山	3	流	3	日日	2
平生	5	古	3	深	3	曾	2
悠悠	5	埋	3	碧	3	木	2
文章	5	堪	3	秋	3	杏花	2
江南	5	多	3	秋色	3	水落	2
浪	5	夜深	3	舟	3	汀洲	2
狂	5	子	3	醉	3	江上	2
白	5	客	3	黃鶴樓	3	波	2
看	5	岸	3	一	2	浪	2
真	5	州	3	上天	2	海上	2
能	5	得	92	乃	2	渺然	2
酒	5	情	3	也	2	皆	2
世	4	才	3	人生	2	相逢	2
仙人	4	投	3	仙人	2	禰衡洲	2
休	4	故人	3	伏	2	美人	2
作	4	春風	3	住	2	與	2

到	4	時	3	佳句	2	船	2
却	4	本	3	倒	2	芙蓉	2
又	4	歌	3	兩地樓	2	萋萋	2
★鸚鵡	1	100位入っていない					

図1~15：中国・韓国鸚鵡洲詩歌における可視化の図

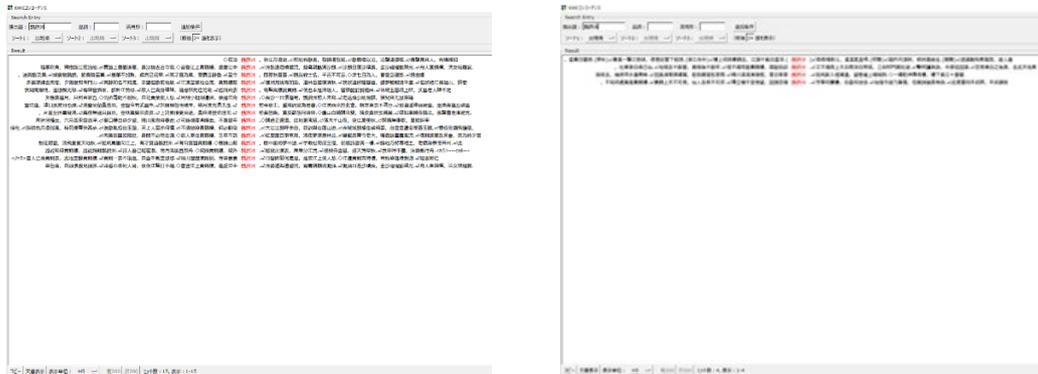


図1. 中国・韓国鸚鵡洲詩歌における KWIC コンコーダンス

N	抽出語	品詞	全体	共起	Jaccard
1	便	Verb	8 (0.100)	4 (0.267)	0.2105
2	前	JJ	5 (0.063)	2 (0.200)	0.1765
3	不	Adv	23 (0.413)	7 (0.467)	0.1707
4	巴	Adv	13 (0.163)	4 (0.267)	0.1667
5	江上	Noun	6 (0.075)	2 (0.200)	0.1667
6	愁	Verb	8 (0.100)	3 (0.200)	0.1500
7	舟	Noun	8 (0.100)	3 (0.200)	0.1500
8	黃鶴樓	Noun	8 (0.100)	3 (0.200)	0.1500
9	樓	Noun	9 (0.113)	3 (0.200)	0.1429
10	江	Noun	9 (0.113)	3 (0.200)	0.1429
11	人驚採擷	Noun	2 (0.025)	2 (0.133)	0.1333
12	光	Noun	2 (0.025)	2 (0.133)	0.1333
13	城	Noun	2 (0.025)	2 (0.133)	0.1333
14	管管	ProperNoun	2 (0.025)	2 (0.133)	0.1333
15	花雲	Noun	2 (0.025)	2 (0.133)	0.1333
16	江中	Noun	2 (0.025)	2 (0.133)	0.1333
17	秀女	Noun	2 (0.025)	2 (0.133)	0.1333
18	龍虎	Noun	2 (0.025)	2 (0.133)	0.1333
19	漢長	Noun	2 (0.025)	2 (0.133)	0.1333
20	精確紙	Noun	2 (0.025)	2 (0.133)	0.1333
21	花白	Verb	2 (0.025)	2 (0.133)	0.1333
22	碧梧	Noun	2 (0.025)	2 (0.133)	0.1333
23	芙蓉	Noun	2 (0.025)	2 (0.133)	0.1333
24	遠閣	Noun	2 (0.025)	2 (0.133)	0.1333
25	采采	Verb	2 (0.025)	2 (0.133)	0.1333
26	雲	Adv	2 (0.025)	2 (0.133)	0.1333
27	舟	Verb	3 (0.038)	2 (0.133)	0.1250
28	雲	Noun	3 (0.038)	2 (0.133)	0.1250
29	州	Noun	3 (0.038)	2 (0.133)	0.1250
30	故人	Noun	3 (0.038)	2 (0.133)	0.1250
31	沙	Noun	3 (0.038)	2 (0.133)	0.1250
32	雲	Verb	3 (0.038)	2 (0.133)	0.1250
33	雲	Verb	3 (0.038)	2 (0.133)	0.1250
34	雲	Verb	3 (0.038)	2 (0.133)	0.1250
35	黃鶴樓	ProperNoun	3 (0.038)	2 (0.133)	0.1250
36	作	Verb	4 (0.050)	2 (0.133)	0.1176

図2. 中国鸚鵡洲詩歌における関連語検索

N	抽出語	品詞	全体	共起	Jaccard
1	云	Verb	5 (0.156)	3 (0.750)	0.5000
2	名	Noun	2 (0.063)	2 (0.500)	0.5000
3	體	Noun	2 (0.063)	2 (0.500)	0.5000
4	愁	Adv	2 (0.063)	2 (0.500)	0.5000
5	愁	Verb	2 (0.063)	2 (0.500)	0.5000
6	美人	Noun	2 (0.063)	2 (0.500)	0.5000
7	樓	Noun	2 (0.063)	2 (0.500)	0.5000
8	愁	Adv	2 (0.063)	2 (0.500)	0.5000
9	入	Verb	4 (0.125)	2 (0.500)	0.3333
10	又	Adv	4 (0.125)	2 (0.500)	0.3333
11	自	JJ	4 (0.125)	2 (0.500)	0.3333
12	花	Verb	4 (0.125)	2 (0.500)	0.3333
13	不	Adv	5 (0.156)	2 (0.500)	0.2857
14	亦	Adv	5 (0.156)	2 (0.500)	0.2857
15	能	Verb	5 (0.156)	2 (0.500)	0.2857
16	一	Noun	1 (0.031)	1 (0.250)	0.2500
17	三	ProperNoun	1 (0.031)	1 (0.250)	0.2500
18	上	Verb	1 (0.031)	1 (0.250)	0.2500
19	下	Adv	1 (0.031)	1 (0.250)	0.2500
20	不	Verb	1 (0.031)	1 (0.250)	0.2500
21	不	ProperNoun	1 (0.031)	1 (0.250)	0.2500
22	言	Adv	1 (0.031)	1 (0.250)	0.2500
23	雲	Noun	1 (0.031)	1 (0.250)	0.2500
24	雲	ProperNoun	1 (0.031)	1 (0.250)	0.2500
25	雲	ProperNoun	1 (0.031)	1 (0.250)	0.2500
26	舟	ProperNoun	1 (0.031)	1 (0.250)	0.2500
27	舟	Noun	1 (0.031)	1 (0.250)	0.2500
28	舟	Noun	1 (0.031)	1 (0.250)	0.2500
29	舟	Verb	1 (0.031)	1 (0.250)	0.2500
30	亦	Verb	1 (0.031)	1 (0.250)	0.2500
31	雲	Noun	1 (0.031)	1 (0.250)	0.2500
32	雲	Noun	1 (0.031)	1 (0.250)	0.2500
33	雲	Verb	1 (0.031)	1 (0.250)	0.2500
34	雲	ProperNoun	1 (0.031)	1 (0.250)	0.2500
35	雲	Verb	1 (0.031)	1 (0.250)	0.2500
36	雲	ProperNoun	1 (0.031)	1 (0.250)	0.2500

図3. 韓国鸚鵡洲詩歌における関連語検索

3.2 対応分析

KH Coder 抽出語を用いた『鸚鵡洲詩歌』対応分析を行い、その結果を二次元の散布図に示すコマンドである。このコマンドは、データをいくつかの部分に分けて、それぞれの部分ごとの特徴を見るのに適している。古典漢詩分析をすることに際して、様々な詩歌主題を見える。例えば、中国『鸚鵡洲詩歌』を用いたチュートリアルでは(図4)、原点(0,0)付近にはこれといって「鸚鵡洲」「鸚鵡」というの語彙が集まる、上の方向は「江上」「白」「浪」風光明媚な景色のことであった。左下には「黃鶴樓」「明月」など詩と建築物のことを示す。また、右下のは「武昌」「芳草」「仙人」「愁」地名・発想・言語感覚の特徴語を見える。朝鮮詩歌の場合は(図5)、上の方向は「芳草」「芙蓉」「杏花」「邊草」綺麗な景勝の言葉であった。左下には「扁舟」「黃鶴」など漢詩からのことを表示する。右下のは「秋色」「禰衡洲」「寒江」「愁」漢詩からのイメージ・地名、また朝鮮の風景を描いた特徴語を見える。そして、朝鮮漢詩について、

同じ詩歌の中の用語「瀾衡洲」「寒江」「古木」「水落」(朝鮮後期・金燾)を集まる状況である。データを検視し、その結果はこの作品が違う版本、二つ以上の詩例を含める為に、群集の状況を産生した。

そうすれば対応分析によって、それぞれの詩歌主題・内容に特徴的な語を探索できる。また、これらの著作年代と言語発想がお互いに近くにプロットされたかどうかを見ることで、どの年代・地域とどの特徴の語彙が似通っていたのかを同時に探索できる。

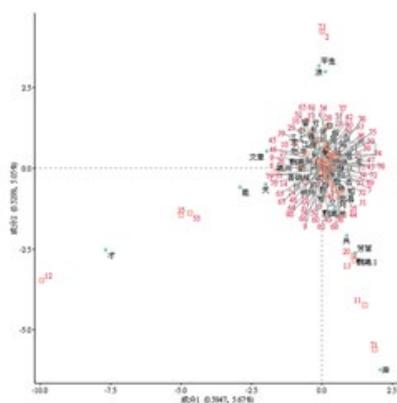


図 4. 中国鸚鵡洲詩歌における対応分析

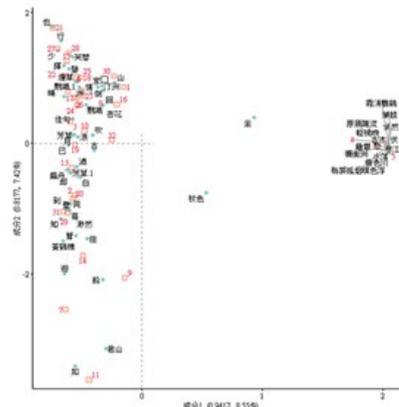


図 5. 韓国鸚鵡洲詩歌における対応分析

3.3 多次元尺度構成法

抽出語を用いて多次元尺度構成法(図7・8・9)を行い、結果が以下の散布図に示すコマンドである。このコマンドは、例えば同じ段落によく一緒に出現する語の組み合わせ、すなわち共起する語の組み合わせをさがしできる。なお、2語の組み合わせに限らず、共起する語のグループを探せば、そこからデータ中に多くあらわれたテーマないしはトピックを読み取れる。基本的には、散布図で近くに布置された語ほど、データ中でよく共起していた、あるいは出現パターンが似通っていたと解釈できる。また、本論の計算方法は Kruskal を選ぶ、そこでプロットを解釈する助けとして、語のクラスター分析を行い、結果を色分けによって示す。

更に、本論の漢詩テーマ分析は「多次元尺度構成法」が主な根拠資料を解析し、共起ネットワークは参考になる。既にテキストマイニング研究について、共起ネットワークを主要な可視化の結果とするが、共起ネットワークは実際に足りないことである。(1) 頻度(フリークウェンシ)があるけど、結ぶ線の関係と方向性が標識できない。(2) 群集間に破線であるけど、具体的の語意・語彙関係には、データを返って検視することが必要、分かりにくい。(3) 台湾のデジタル人文学研究場合は、Gephi を使っている社会ネットワーク分析(Social Network, SNA)に人物関係を示すものが多い、共起ネットワークも丸いバブル語彙群集との線を結ぶ、それでは、KH Coder の特性が表現できない。以上の理由を考えると、本論は「多次元尺度構成法」のプロットに基づいて、漢詩テーマ分析を行う。

以下は Jaccard・Cosine・Euclid 三つ結果を比較して、中国と韓国の『鸚鵡洲詩歌』の主題特徴を解釈する。

3.3.1 中国『鸚鵡洲詩歌』(図7)

(1) 番号 01 「鸚鵡」「洲」「月」「江」「不」「黃鶴樓」「舟」「秋」、このプロットに唐代白居易詩「夜泊鸚鵡洲、秋江月澄澈」、または孟浩然詩「昔登江上黃鶴樓、遙愛江中鸚鵡洲...舟人牽錦纜、浣女結羅裳。月明全見蘆花白、風起遙聞杜若香」の例を取り上げてみる、「鸚鵡洲」「月」「秋」「江」の雰囲気などについては、詩の中に、巧妙表現な運用を目立つ。

(2) 番号 02 「山」「洞庭」「赤壁」「寒」「老」、このプロットに唐代劉長卿「漢口夕陽斜渡鳥、洞庭秋水遠連天。孤城背嶺寒吹角、獨樹臨江夜泊船。」宋代樓鑰「洞庭一螺浮具區、赤壁但歌前後賦...歸老故山吾已矣」の語彙を見えてきた、「洞庭」「赤壁」の景色と周囲環境の感覚を描いた。

(3) 番号 03 「天」「平生」「浪」「行」「坐」「欲」、このプロットに唐白居易「白花浪濺頭陀寺、紅葉林籠鸚鵡洲。恁是平生未行處、醉來堪賞醒堪愁。」鸚鵡洲また近い風景地であることを詠んだ。

(4) 番号 04 「醉」「酒」「更」「留」、このプロットに宋梅堯臣「得錢留酒家、醉卧江蕪綠。」同じ宋王遽「滿甕尚留桑落酒、登盤今有武昌魚。」また、宋章甫「折梅置酒齋堂幽、醉狂於披紫綺裘...有酒如川容拍浮」などのことを見られるように、詩人はお酒と酔っ払い気持ちを描いたことである。

(5) 番号 05 「悠悠」「江南」「天地」「未」「狂」「大」「横」、このプロットに宋李綱「行盡江南山、始次湖外州。吳越天杳杳、風煙晚悠悠。」長江が極めて清澄、遠い目の感じを歌う。

(6) 番号 06 「武昌」「知」「仙人」「芳草」「鸚鵡」「岸」、このプロットに宋劉放「黃鶴仙人去不留、武昌官柳傍層樓。前山欲近滄浪水、芳草遙分鸚鵡洲。」「黃鶴仙人」の典故・伝説を運用することである。

(7) 番号 07 「愁」「鸚鵡洲」「明月」「南」「碧」、このプロットに唐孟浩然「昔登江上黃鶴樓、遙愛江中鸚鵡洲。洲勢逶迤繞碧流、鴛鴦鸚鵡滿沙頭。」元郝經「去年燕南醉明月、黃金臺上秋風發。」詩人の「愁思」を出された。

(8) 番号 08 「只」「故」「休」「在」「看」、このプロットに宋釋居簡「致身須黃閣、看鏡方黑頭。故家邁種德、台輔不足酬。古道復大明、殊俗皆蒙休。」「擬古・復古」の主題を書いた。

以上の多次元尺度構成法の結果を見たとき、古典漢詩と文学史的に言えば、詩歌内容の模範・典型を発見された、此の上とも、中国古典漢詩としての「鸚鵡洲」を考える場合においても、「黃鶴樓」「鸚鵡洲」「明月」「醉」「酒」などイメージは後世の於いて、よく用いられました、決定的な作用を及ぼした作品と言える。つまり、デジタル可視化の図表通り、中国古典詩を詠む場合は、このイメージと語彙が無視できない存在となっておりということである。

3.3.2 韓国『鸚鵡洲詩歌』(図7)

(1) 番号 01 「鸚鵡」「洲」「江」「水」「舟」「山」「川」「客」、このプロットに朴胤源「不有鸚鵡洲、難爲黃鶴樓。舟兼蒼壁轉、江與白雲流。」徐必遠「深追山水子長遊、豈學蓴鱸張翰舟。晴川漠漠漢陽樹、落日亭亭鸚鵡洲」の詩例を取り上げて見ると、朝鮮漢詩については中国「黃鶴樓」と「鸚鵡洲」を模倣し建立したことである。そして、中国『鸚鵡洲詩歌』から語彙と典故を継承する事実である。

(2) 番号 02 「黄鶴樓」「去」「人生」「芳草」「扁舟」「遊」「白」「美人」、このプロットに朴準源の詩を連想できて、「去年伯氏湖西遊、扁舟一上黄鶴樓。今年匹馬關西出、黄鶴樓上弄明月。兩地樓名偶相同、隨緣一遊無西東。二樓眞假強欲分、仙人一去空白雲。更指鸚鵡洲中景、謂言關西是眞境。...但記晴川與芳草、祇今悅對風景好。」この一連のイメージ、中国の『鸚鵡洲詩歌』は朝鮮漢詩に「鸚鵡洲」関わる主題・典故・用語に多大な影響を与えてくれたことを示す。

(3) 番号 03 「日」「江流」「得」「地」「芳草」「壁」、このプロットに朴胤源「獨立欄干空悵望、回頭忽得鸚鵡洲。芳草何萋萋、白雲何悠悠。...飄然脚下生雲氣、濯足俯臨滄江流。...願君須及秋風至、丹楓翠壁月幽幽。」この詩の一連イメージから、中国唐詩の語彙を模倣した、朝鮮の景色を描いたことを明らかにする。

(4) 番号 04 「秋色」「古木」「更」「君山」「船」、このプロットに金燿「古木霜清鶴鶴愁」「鴈挈秋色向陽流」、洪爽周「洞庭秋色更茫然」「君山月落蒼蒼樹」「夢中歸去五湖船」の詩例の語彙とあります。朝鮮の地名を中国の地名に換えたり、あるいは漢語風の地名にすることは、朝鮮中後期の漢詩壇においてひろく行われたことである。このような技法が、漢詩を作る上にどのような意味を持ったか、それは文学研究の分野であるけど、可視化のプロットから「地名の転換」と「唐詩の模倣」などのことが確かめられた。

(5) 番号 05 「上天」「浪」「吹」「倒」「悠然」「好」「古」、このプロットに申景濬「杏花深處餘村酒、長咏滄浪一曲回」徐居正「白鷗落照心悠然」というの詩例もあります。唐詩の模倣を強く意識する朝鮮漢詩は、実は「鸚鵡洲」以外の典故「杏花」「滄浪」「白鷗」など中国の古典を詩に詠み込むことがあった。例え、申景濬の詩について、実は中国唐代杜牧の詩で「借問酒家何處有？牧童遙指杏花村」とあります。徐居正の詩は唐杜甫「舍南舍北皆春水、但見羣鷗日來」というイメージと語彙を改めて組み合わせと利用した。言い換えて、このプロットから、朝鮮漢詩を創作する際、唐詩の原典を意識し、重ねて理解と使用していたと考えられる。

(6) 番号 06 「江上」「深」「杏花」「城」「家」「汀洲」、このプロットに金載瓚「能城江上之南津、兩山夾出水西東。」「自言家住清江上、少小捉魚清江曲。林樾春深桃杏花、汀洲秋入蒲菰葉。」このプロットから分かることが、朝鮮漢詩は、中国古典詩を典拠にしつつ、唐詩などの言葉を使用したのである。このような表現の手法は、朝鮮文人と漢の人情が同じという同情の思想に基づくものと言えよう。「朝鮮」「漢文化」同情という文学観が前提とすれば、理論上、漢詩という枠組みの中に漢の言葉を用いて、朝鮮の人情を十分に表現することが可能になると信じたのである。

(7) 番号 07 「不」「晴」「可」「何」「病」「海上」「行」、このプロットに金祖淳「晴沙綺壁暎秋暉」、徐居正「雲卷芙蓉海上城」、李安訥「葺山病守非崔顥、醉墨還題最上頭」の語彙とあります。さらに、このプロットから「朝鮮-中国」同情という観点からみると、朝鮮漢詩が唐詩の模倣のような詩を作ることは、ある詩例の正当性を持っているようにも解釈できる。朝鮮も漢（中国）も、その根底にある人情が同じであれば、人情を表現するための手段である模倣は、形式通りの漢の人情だけではなく、朝鮮の人情をも表現することが可能になる。「朝鮮-中国」同情という文学観は、朝鮮漢詩の詩を考える上で、必要不可欠の概念

であると言えよう。

(8) 番号08「醉」「木」「平生」「情」「碧」、このプロットに徐居正「鸚鵡洲邊木蘭棹、誰將醉墨語江風」「山石平生犖确行、此軒贏得十年情」、李玄錫「一嘯乾坤獨倚樓、樓下長江十里碧」の詩例とあります。漢詩という枠組みに朝鮮の人情をも取り入れる例として、朝鮮漢詩の作品を「朝鮮-中国」同情の観点から検討してきた。朝鮮漢詩における漢詩文の摂取など、語彙の枠組みの中に朝鮮の人情を表現しようとする動きもその時代にあった。

以上の分析結果を準拠して、朝鮮漢詩は、中国の古典の見聞も詩の素材として用いていたことである。意外にも詩に多様な朝鮮風土の要素までを取り入れていたのである。唐詩の模倣にこだわった朝鮮漢詩が、それとともに、当時の朝鮮を表現しようとした証拠であると指摘できる。本論は、テキストマイニングデジタル分析方法を用いた、その中国・韓国『鸚鵡洲詩歌』の文学表現の再評価を試みる。中国詩を目指したはずのその詩に朝鮮的な要素が見られることに着目し、それが朝鮮漢文学としての性格をもつことを論じる。そして、その「朝鮮-中国」漢字文化圏の背景に「朝鮮-中国」同情という思想があったことを指摘する。それにより、中国と韓国漢詩との重要な受容の例が明らかにされる。

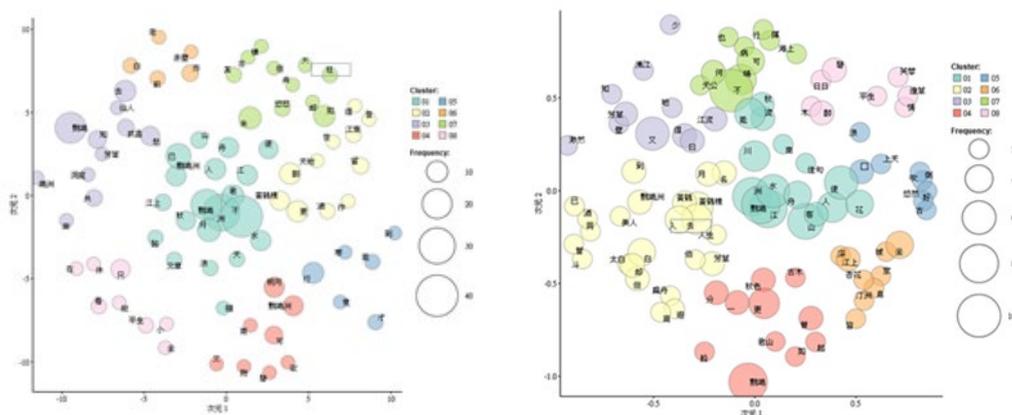


図7. 中国・韓国鸚鵡洲詩歌における多次元尺度構成法 (Jaccard)

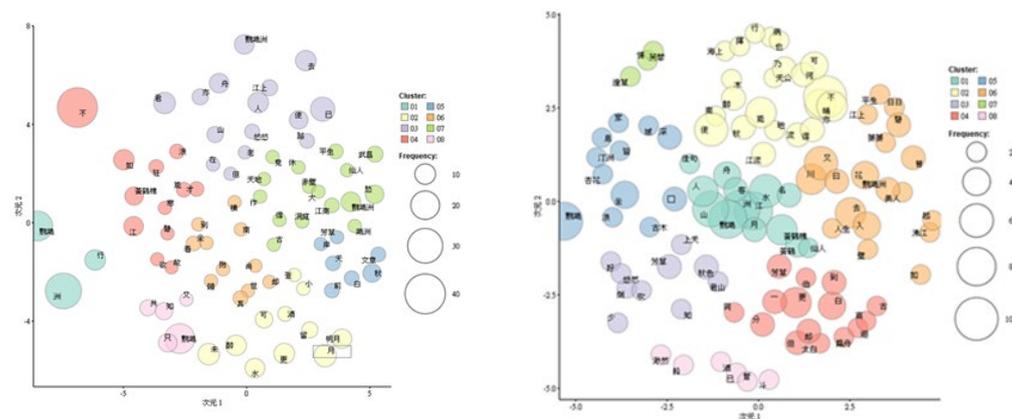


図8. 中国・韓国鸚鵡洲詩歌における多次元尺度構成法 (Cosine)

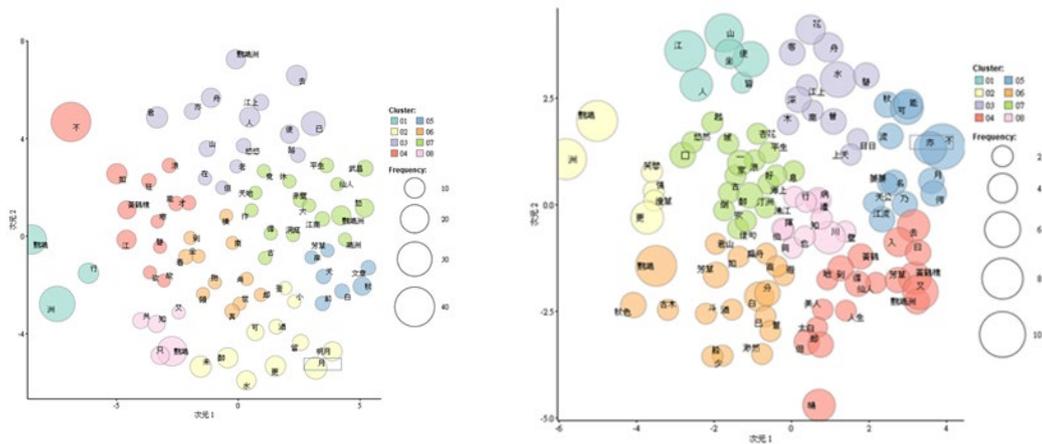


図9. 中国・韓国鸚鵡洲詩歌における多次元尺度構成法 (Euclid)

3.4 頻出語・共起ネットワーク

表3「中国・韓国鸚鵡洲詩歌における頻度別抽出語・整理後結果」上位100位の頻出語を図10から15までには頻出語の共起ネットワークを示した。中国鸚鵡洲詩歌の頻出語の上位を占めるは「不」「洲」「鸚鵡洲」「黃鶴樓」「明月」「舟」「愁」「秋」「醉」で、韓国鸚鵡洲詩歌の頻出語の上位を占めるのは「江」「洲」「鸚鵡洲」「芳草」「川」「晴」「江流」となった。これら頻出語の他に、共起ネットワーク分析から明らかとなる「中国鸚鵡洲詩歌」の代表的なテーマは「黃鶴樓と鸚鵡洲の景色」「李白・崔顥・孟浩然の詩例」「地名・懐古・述景」という一連の連想を有機的に結び付けたということである。「韓国鸚鵡洲詩歌」の代表的なテーマは情景を歌いあげていることなどである。

3.4.1 中国『鸚鵡洲詩歌』

中国『鸚鵡洲詩歌』については、共起ネットワークから「黃鶴樓」「舟」「酒」「醉」「留」(図11左図)を見ると、テーマの特徴を示す。中国『鸚鵡洲詩歌』では、実際に禰衡『黃鶴樓賦』・李白『黃鶴樓詩』・崔顥『黃鶴樓詩』の典故を多く用いる。また、ほかのネットワーク、例え、「鸚鵡洲」「使」「人」「愁」「悠悠」「天地」「江南」「横」(図11左図)という用語を観察し、中国古典詩のテーマの特徴、また、詩と名勝・古跡歌い必要条件は、次のように集約できるものと思われる。(1)ある具体的な空間・物景(例え、「黃鶴樓」「黃鶴樓」)が存在することと、(2)詩に歌われること(著名な作品、歴史典故など)と、(3)空間・景物に関する発想のイメージと(4)詩人の感想・感覚言語。つまり、(1)～(4)の言葉などのことを繰り返し反復利用し、中国古典詩の創作手法を解明できて、中国『鸚鵡洲詩歌』の重要なイメージと意義また思想に基づくものだと考えられる。

3.4.2 韓国『鸚鵡洲詩歌』

韓国『鸚鵡洲詩歌』については、共起ネットワークを検討して、テーマの特徴を簡単に解明できる。例え、「禰衡洲」「蒹葭」「寒江」「鼃鼃」「愁」「古木」「悄然」というのは、朝鮮漢詩には、確かに中国の古典詩の素材として用いていた。それに、詩に多様な朝鮮の要素までを取り入れていたのである。中国詩の模倣にこだわった朝鮮漢詩が、それとともに、当時の朝鮮を

表現しようとした証拠であると指摘できる。それ以上、「黄鶴楼」「仙人」「寒山寺」「兩地樓」という言葉は、建築名と地名を中国の「黄鶴楼」などの模倣であると、その中国風の地名からの連想語を詩中にちりばめるのは、朝鮮漢詩の技法表現にも通じる方法であった。その結果については、中国古典文学史上に朝鮮漢詩を位置づけるための新たな視点を提示するものとして注目に価すると思われる。

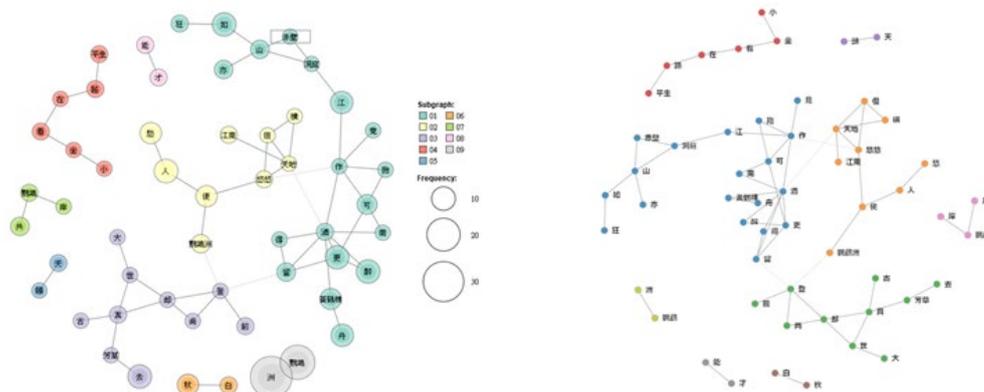


図 10. 中国鸚鵡洲詩歌における抽出語・共起ネットワーク (Jaccard)

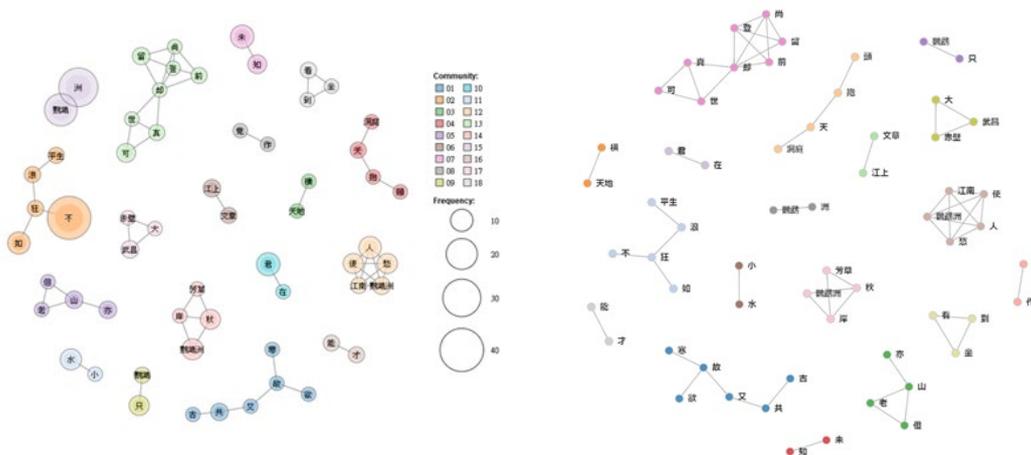


図 11. 中国鸚鵡洲詩歌における抽出語・共起ネットワーク (Cosine)

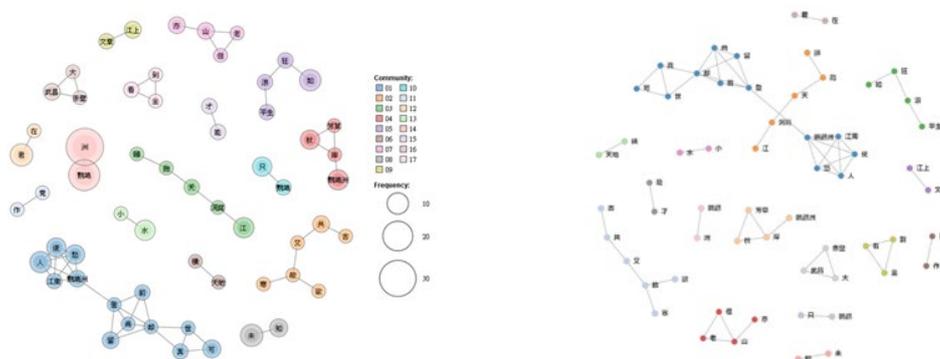


図 12. 中国鸚鵡洲詩歌における抽出語・共起ネットワーク (Euclid)

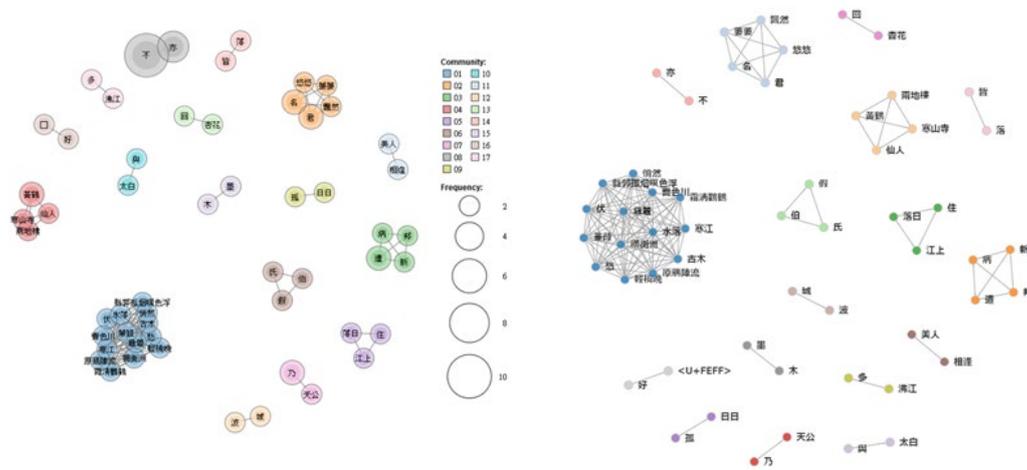


図 13. 韓国鸚鵡洲詩歌における抽出語・共起ネットワーク (Jaccard)

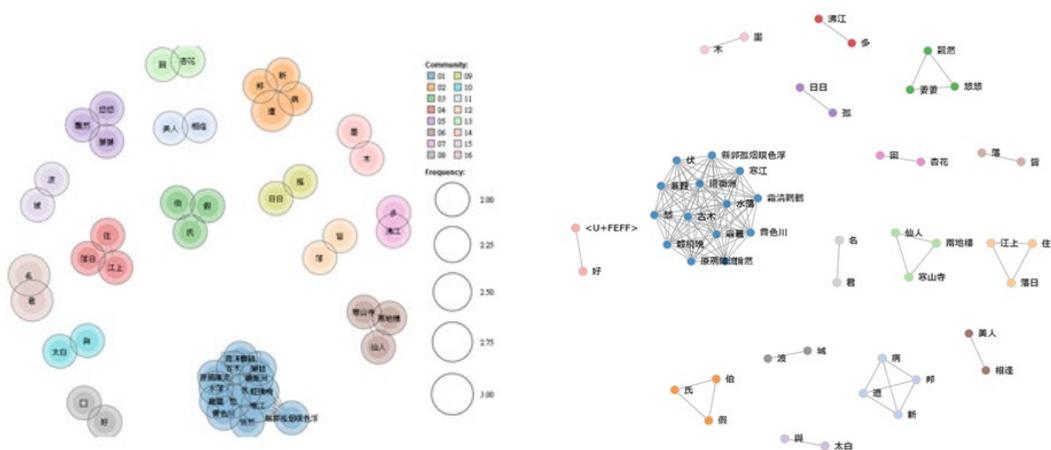


図 14. 韓国鸚鵡洲詩歌における抽出語・共起ネットワーク (Cosine)

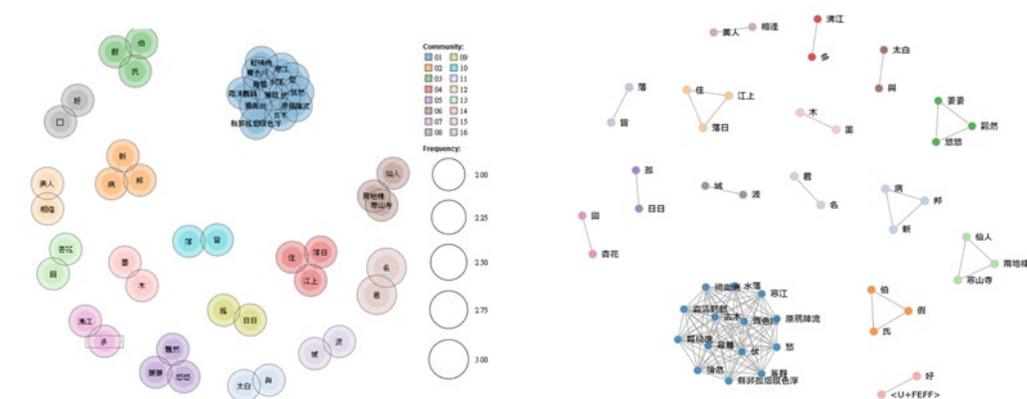


図 15. 韓国鸚鵡洲詩歌における抽出語・共起ネットワーク (Euclid)

4. まとめと意義

4.1 まとめ

本研究はテキストマイニング (KH Coder) の手法を導入して、中国と韓国の古典漢詩『鸚鵡洲詩歌』のイメージと意義について検討した。その結果、中国と朝鮮『鸚鵡洲詩歌』のイメー

ジ特徴、典故と語彙間の関係、語彙で思惟を表現すること、漢語風の地名表現的創造と変換が明らかとなった。但し、本論では分析結果の全てについて詳述することはせず、部分の結果だけに注目した偏った説明していた。言及に至らなかった結果の詳細については、各図表を基づいて参考されたい。

2と3で紹介した通り、「KH Coder」をはじめとする古典漢詩の主題・典故・イメージへのデジタル技術と可視化の活用には、従来のテキストマイニングの運用と研究、及び古典漢詩の分析と解読に既に文学史の制約から解放される可能性、新しい視点を開拓したのが本研究における第一の成果である。

特に台湾のデジタル人文学の研究分野において、Gephiで社会ネットワーク分析と語彙共起ネットワークを作る事例が多いし、「KH Coder」を運用する事例がないので、今回の研究通り、デジタル人文学と古典文学を結ぶ可能性として提供する道筋を示した。

語彙共起ネットワークについて、今回は「多次元尺度構成法」で方向性を示すことに基づいて、台湾の学界がないと新しい文学主題・典故の可視化を提供しているが、今後は古典文学においてもテキストマイニング分析を行う研究が増えていくと思われる。テキストマイニングが古典文学の研究でどのように利用され、どのような展開を見せるのか注視していきたい。

4.2 本研究の意義

本研究の第一意義は、日本開発するテキストマイニングソフトウェア KH Coder を運用し、台湾のデジタル人文学界既に研究に対して、新しいテキストマイニング資源を紹介と使用し、公開する点にある。また、従来の KH Coder 研究を遡って、古典漢詩に用いる事例が少し、このような状況を考えて、本論は KH Coder で中国・朝鮮漢詩を分析を行う、深い記述が可能となる、デジタル人文学の研究という点で、研究者また KH Coder ユーザーの試みは一つモデルとなるだろう。

なお、第二の意義は、中国古典文学研究にデジタル元素とテキストマイニングの手法を導入して、中国・韓国『鸚鵡洲詩歌』のテーマ・典故・イメージと意義について検討した。今までに中国古典漢詩における研究には、伝統的なテキスト解釈・文義を分析、詩歌主題・内容・イメージを解読、詩歌用語の典故を考察することである方面の三種があったとされる。しかし、東アジア漢詩、またデジタル人文学の研究には、古典漢詩とデジタル人文学の境界線上にあるような事例が少なくない。本研究ではそうした中国・朝鮮『鸚鵡洲詩歌』事例を集めて精査することによって、テキストマイニング (KH Coder) で古典漢詩におけるテーマの特徴・イメージ・意義それぞれの実態を把握し、それらが互いにどのように絡み合っていたのかを解明することであり、これこそ、人文学が取り組むことに相応しいデジタル研究だと言える。

謝辞

本研究で用いた中国・朝鮮漢詩『鸚鵡洲詩歌』のテーマ発想について、文学史と歴史的な研究観点にし、それをもとにデジタル人文学の研究手法など丁寧に指導して下さった台湾中央研究院歴史語言研究所教授研究員・ASCDC デジタル文化センター召集人陳熙遠先生にこの場を借りて心より感謝申し上げます。尚、当然のことながら本論における分析の一切の責任は私にあることを申し添えておきます。

参考文献

- 王啟興, 張虹, 陳順智: 明刻黃鶴樓集校注, 中国: 湖北人民出版社, 2019
- 後藤真(編集), 橋本雄太(編集), 国立歴史民俗博物館(監修): 歴史情報学の教科書: 歴史のデータが世界をひらく, 東京: 文学通信, 2019, <https://bungaku-report.com/metaresource.html>
- 小風尚樹: 歴史研究とデジタル・ヒューマニティーズの相関, 「東京大学大学院横断型教育プログラム デジタル・ヒューマニティーズ」, 研究報告書, pp. 1-8, 2016
- 小風尚樹, 中村 覚, 纒田宗紀, 山王綾乃, 小林拓実, 清原和之, 金甫榮, 福田真人, 山崎翔平, 槇野翔, 小川潤, 橋本雄太, 宮本隆史, 菊池信彦, 後藤真, 崎山直樹, 元ナミ, 加藤諭: デジタルヒストリー入門: 2018 Spring Tokyo Digital History Symposium 開催報告, 「2018 Spring Tokyo Digital History Symposium」シンポジウム, 研究報告書, pp. 1-57, 2018
- 陳熙遠: 人去樓坍水自流——試論座落在文化史上的黃鶴樓, 中國的城市生活: 十四至二十世紀, 臺北: 聯經, pp. 367-416, 2005
- 陳熙遠: 長江圖上的線索: 自然地理與人文景觀的歷史變遷, 中央研究院歷史語言研究所集刊, No. 85.2, pp. 269-358, 2014
- 寺尾剛: 李白における武漢の意義—「詩的古跡」の生成をめぐって, 中國詩文論叢, No. 11, pp. 68-94, 1992
- 樋口耕一: 社会調査のための計量テキスト分析—内容分析の継承と発展を目指して, ナカニシヤ出版, 2014
- 臺灣中央研究院・歴史語言研究所「漢籍電子文獻資料庫」:
<http://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm>
- 韓国電子書籍学術論文サービス KRpia: <https://www.krpia.co.kr/>
- 北京愛如生数字化技術研究センター: 『中國基本古籍庫』

Classical Chinese Poetry Image and Content Analysis using Text Mining Software KH Coder-- Take Chinese and Joseon “Parrot Isle Poetry” As a Research Subject

Lin, Wenjiun

Abstract

“Hanyang(漢陽)” is a place name in China, a sandbar in the Yangtze River(揚子江), southwest of Hanyang County(漢陽縣), Hubei Province(湖北省). Hanyang County is a district located in Wuhan City, Hubei Province, China. In the era of “*Record of the Three Kingdoms*(三國志)”, which is after the end of the Eastern Han Dynasty, it became an important political and military stage and attracted the attention of later poets. In the past time, the battle between Liu Biao(劉表) and Son Wu(孫吳) in Jingzhou(荊州), and there was a story of Mi Heng(禰衡), a madman, which also happened during this period.

Initially, Huang Zu(黃祖) welcomed Mi Heng, who also had an intimate relationship with his firstborn son, Huang Ye(黃射). At that time, there was a person who presented a parrot at a banquet, and Mi Heng wrote a wonderful and well-known literature work, “*Parrot Fu*(鸚鵡賦),” and this article went down in history. Later, when “parrot Isle” were sung in literary works, many of them were based on this “*Parrot Fu*(鸚鵡賦).” One example of the poem by Cui Hao(崔顥) in the Tang Dynasty is “The Tower of Yellow Crane” In South Korea, during the Joseon Dynasty, the Chinese “Yellow Crane Tower” was erected in Santo(三登, 삼등), Pyongan Province(平安道, 평안도), and this building was also given the name “Yellow Crane Tower”, on the opposite mountain wall was titled “Parrot Isle(鸚鵡洲)”.

The word “Parrot Isle Poetry” used in the title of this research is a general term for the poems that describe the above-mentioned Chinese and Korean Joseon “Parrot Isle Poetry”, on the top of that “Parrot Isle” is a classical allusion of this place name and origin. It is a well-known fact that is often cited in Chinese poetry, but looking back on the research field of classical Chinese poetry, there is no concrete example of using the name “Parrot Isle Poetry.” In view of the mentioned reasons, this research uses the convenient name that borrows the words, allusions, and concepts that were established in China. It was also possible to clarify that such cases exist in Chinese and Korean poetry. Digital humanities are also the goal of this research.

The subject of this paper's analysis is to prepare data that collects poetry on the theme of “Parrot Isle” in the Chinese classical books and the Korean literature collection. KH Coder (Higuchi, 2014) was used for text mining analysis. First, the dataset was decomposed into words by the morphological analysis. In KH Coder, this operation is called preprocessing. The preprocessing reveals the number of words, which is different from the total number of extracted words, and the number of words, which is different from the total number of extracted words excluding particles and auxiliary verbs. Subsequently, a list of 100 frequently occurring words was created from the preprocessed data.

In the case of this study, the numerical values shown in this list analyze the number of occurrences of extracted words (term frequency: TF) and the number of documents including extracted words (document frequency: DF), organize the data and Chinese classical poetry. From Chinese to Korean

classical poetry, it was confirmed that there are multiple examples of this place name and deceased widespread use and conversion, and the explosive rise of the image and significance of "Chinese-Korean Parrot Isle Poetry" The elucidation of what was done has progressed. Other analysis methods used in this paper are correspondence analysis that enables visual data classification, co-occurrence network that connects co-occurrence vocabulary elements with lines, and the lists characteristic words are analyzed by "Jaccard," "Cosine" and "Euclid" three analysis calculation formulas.

Furthermore, in light of the experience of the author who has collected and researched Chinese and Korean materials on "Parrot Isle Poetry," the image and significance of "Parrot Isle Poetry," are included in the classical Chinese poetry. A concrete text mining study on the importance of software that directly investigates the humanities and solves it from the standpoint of informatics, and the possibility of digital humanities that is also an important goal on this research.

Keywords: Digital Humanities, Text Mining, Classical Chinese Poetry, Parrot Isle, Parrot Isle Poetry, Poetry Image