政大日本研究【第十八號】
2021 年 1 月，頁 253～271
10.30379/CJJS.202101_(18).0010

從文字探勘的角度來看漢字文化圈的〝異〞與〝同〞

落合由治*

中文摘要

語言類型學自建立以來，一直是語言學領域的一個問題，漢語和日語被認為屬於對立的語言類型。但是，當我們從實際使用的社會文本來比較，就會發現不同的視角。在此，筆者想從文字探勘的角度來考慮中日文本在體裁上的差異，文字探勘已被應用於語言材料的定性分析。

如果以文字探勘為方法，用同樣的方法分析同一體裁的文本表達，中日兩國的結果是完全一樣的，可以認為某一體裁的語言表達超越了語言的差異。由此可見，語言表達的功能超越了個別語言的差異，在語言比較中，有必要設置一個不同於傳統語言類型學的研究課題。

關鍵字：語言類型學，中文，日文，文字探勘，體裁

* 淡江大學日本語文學系特聘教授

"Different" and "Same" in Kanji Cultural Areas From the perspective of text mining

Yuji Ochiai*

Abstract

Linguistic typology has been an ongoing issue in the field of linguistics since its inception, and Chinese and Japanese are considered to belong to contrasting linguistic types. However, when we compare them in terms of the social genres in which they are actually used, we can find a different perspective. Here, I would like to consider the differences between Chinese and Japanese text genres from the perspective of text mining, which has been applied to the qualitative analysis of linguistic materials.

Using text mining as a method, the results of analyzing text expressions in the same genre using the same method are exactly the same results in both Chinese and Japanese, and it can be considered that linguistic expressions in a particular genre transcend linguistic differences. This indicates that there is a level of linguistic expression that functions beyond the differences between individual languages, and that there is a need to set up a different research agenda from the conventional linguistic typology in language comparison.

Keywords: linguistic typology, Chinese, Japanese,.text mining, genre

* Distinguished Professor, Department of Japanese Language, Tamkang University

漢字文化圏における「異」と「同」
―テキストマイニングの観点から―

落合由治*

要旨

言語類型論は、言語学の分野でその成立から現在まで継続している課題であり、その中で中国語と日本語は対照的な言語類型に属すると考えられている。しかし、実際に言語が運用されている社会的ジャンルで比較してみると、また異なる観点を取り出すことができるようである。ここでは、言語資料の質的分析に応用されるようになったテキストマイニングを一つの観点として、中国語と日本語の文章ジャンルでの異同を考えてみたい。

テキストマイニングを方法とすると、同じ方法で同じジャンルの文章表現を分析した結果、中国語でも日本語でも結果はまったく同一であり、特定のジャンルでの言語表現は言語の差異を超えていると考えることができる。こうした個々の言語の差異を超えて機能している言語表現のレベルが存在していることをこれは示しており、従来の言語類型論とは別な研究課題を設定していく必要が言語比較においても存在していると考えられる。

キーワード：言語類型論、中国語、日本語、
テキストマイニング、ジャンル

* 淡江大学日本語文学科特別教授

1.はじめに

言語の類型論は、言語学の分野でその成立から現在まで継続している課題であり、中国語と日本語は対照的な言語類型に属すると考えられている。言語類型論では、言語の見かけ上の形態の差異から文法形式の異同など、主として文レベルの構造の比較が中心的視点であった。しかし、実際に言語が運用されている社会的ジャンルで比較してみると、また異なる観点を取り出すことができるようである。文章レベルでの各言語の比較ここでは、言語資料の質的分析に応用されるようになったテキストマイニングを一つの観点として、中国語と日本語の文章ジャンルでの異同を考えてみたい。本論文では、以下の順番で考察を進めていきたい。

（1）19 世紀からの言語類型論の代表的な論点を紹介し、20 世紀後半に盛んになった文章談話研究の中での対照分析の視点について主な論点を述べる。

（2）異なる言語の比較の視点として文章の社会的ジャンルの観点から、テキストマイニングを方法として中国語と日本語の新聞社説を比較し、言語間での文章の社会的ジャンルの異同について考察する。

2.言語類型論の視点

以下では、まず、言語の分類では最も伝統的な観点を提示する。

2.1 形態論的類型論

こうした論の紹介は多数あるが、郡司隆男(2011:2)によれば、言語類型論の考え方は、19 世紀には確立し、古典的観点として、言語の見かけ上の形態的特徴による Schleicher(1850)による 3 分類が知られ、現在でもこうした基本分類が使われている。

a。孤立語−個々の形態素が語としての独立性を保ち、語形変化や接辞を伴ったりしない。語と語の文法的関係は語順によって示される。中国語・チベット語など。

b。膠着語−内容をあらわす語幹に接辞をつけて、さまざまな文法機能をあらわす。日本語、韓国語など。

c。屈折語-主に語形や語尾の変化（屈折）によって、さまざまな文法的機能をあらわす。英語、ドイツ語など多くのヨーロッパ言語。

2.2 統語論的類型論

郡司隆男(2011:2-3)によれば、20 世紀後半になって、世界の各地域の言語の統語構造の研究が進み、文法の観点から言語のタイプ分けが言われるようになった。古典的研究は、Greenberg (1963)による一般化で、現在も広く用いられている。文の最も基本的な要素である、主語(S)、目的語(O)、動詞(V)という 3 要素の配列で 3 パターンで大半が構成されるとした。世界の言語の分布は以下のような比率になる。

a。SOV-日本語、韓国語など（45%）

b。SVO-英語、中国語など　（37%）

c。VSO-アラビア語、アイルランド語など（18%）

現在までの言語類型論は、以上、二つの観点をパラダイムに研究が進められて、世界各国、地域、民族の言語の基本的記述のフレームになっている。

2.3 パラメータ類型論

郡司隆男(2011:3-4)によれば、1980 年代以降の生成文法では、世界中の言語に共通の普遍文法を仮定し、付随するパラメータ値の設定で個別言語の文法が生まれると考えている。今までに提案されたパラメータの例は、以下のようになる。

a。主要部パラメータ:主要部先行 vs。主要部後行―SVO 言語のように重要部分が先に出る言語と、SOV 言語のように陳述決定要素が最後に来る言語の対比である。

b。一致パラメータ:一致のある言語 vs。ない言語―一致とは S や O などの文法項の数・人称・性などに応じて動詞の語尾形が変わる現象である。欧米語は前者であり、日本語や中国語は後者である。

c。空主語:主語などの文法要素の省略を許さない言語 vs。許す言語―S や O のない文でも文法的な文になるか、省略できないかという要素の省略から見た分類である。

以上、郡司隆男(2011)によって見てきたように形態論的類型論、統語論的類型論を元にして、さらに細部を比較する生成文法などで、こうした比較の観点が出されて各国語の比較が行われている。文法形式の違いなどから各国語を見る場合、有効な方法と言える。日本語教育の場合も、各学習者の母国語のとの対照を考えて、教授ポイントを工夫する必要がある。

2.4 文章談話対照分析の視点

第二言語教育の観点では、この 40 年あまり各種言語類型論に基づいて、学習者の書いた作文や話している談話を資料として、母国語使用者との比較対照が行われてきた。

泉子・K・メイナード(1977:74-80)によれば、対照談話分析の始まりは Gleason(1968)で、ニューギニアの言語と英語の物語展開の比較がされている。さらに、Hartmann(1980)が対照テキスト論を提唱し、各言語で名詞で表現されるか動詞で表現されるかなど、表現方法の比較が行われた。有名な説は、Kaplan(1972)で、世界各国語の論理展開パターンを以下のように整理している。

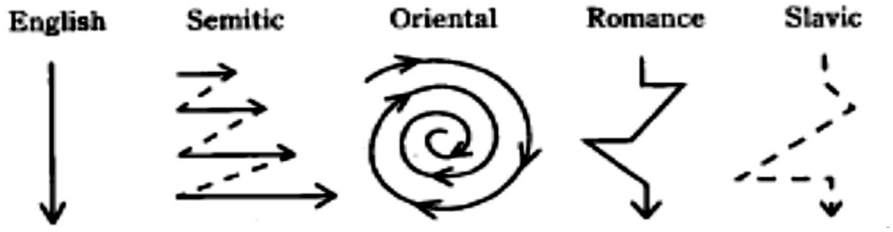


図 1　カプランによる各国語の論理展開モデル

さらに、日本語の対照分析の先駆は、Hinds(1990)で日本語、中国語、タイ語、韓国語では帰納的レトリックが使われ、英語圏では演繹的レトリックが使われるとしている。

こうした対照談話分析の流れは、日本でも試みられ、本名(1989)は日本語の文体と英語の文体を比較し、英語は「直線的な文章」であり、日本語は「遠回しな文章」であるとしている。また、西原(1990)は、日本人が書いた英作文を元にして、英語では以下への予告として一般的な事柄が述べられるのに対して、日本語では最後に結論を述べる起承転結の展開が多いと述べている。

泉子・K・メイナード(1977)によって対照談話分析の初期成果を紹介したが、こうした研究が元になって現在では日本語教育な

ど第二言語教育の分野で、文法的形式から文章・談話展開まで同じ観点で各種の比較が行われるようになっている。

以上述べてきた、比較言語学、対照談話分析は、各言語の差異に注目して、一定の基準で類別を行い、相互の特徴を記述する研究である。以下では、第二世代 AI ブームの中で発展し、現在、広く言語の特徴分析に使用されるようになったテキストマイニングを基準に使って各国語の対照をおこなってみたい。李在鎬編(2017:10-11)は、AI 技術分野の一領域である自然言語処理の発展で可能になった言語の特徴を取り出す技術であるテキストマイニングについて、以下のように説明している。

- 第 1～第 2 世代 AI の時代には計量言語学、計量国語学、計量文献学、計量文体論などの領域で品詞量の統計的分析による特徴の抽出が行われ、テキストマイニングの技術が発展した。
- 第 3 世代 AI の時代である現在、計量的方法や機械学習で擬似的に言語の意味を抽出し、特徴を可視化することが可能になり、言語表現の質的分析手法として社会科学分野やビジネス分野で応用されるようになった。

本論では言語比較の方法として、テキストマイニングを用いて考察を試みる。

3. 文章ジャンルによる言語比較

文章ジャンルとは、社会的に言語非言語の表現が使用される場合、その使用目的に応じて異なる表現構造が用いられることを意味し、一般的には評論、論説、説明などある対象について話し手の思いを述べる場合と、小説、物語などのように対象となる現象を経時的に描写する場合、詩、歌詩などのように反復する一定の形式を用いて話し手の思いを述べる場合など、さまざまな表現場面が考えられる。文章ジャンルは、質的な表現分類である。

文章ジャンルに関する研究は、文章の基本的類型などのテーマで日本でテクスト研究が始まった。1960 年代～80 年代にかけてさまざまな考察が行われた課題の一つである。日本で初めてテクス

ト研究（文章研究）を提唱したのは時枝誠記であり、その時から言語研究の語、文、文章という基本的単位が認定されるようになった。ここでは、文章の基本的類型のモデルとして、文法形式との対応がはっきりしている永尾章曹の二種三類の文章の基本的類型を取り上げる。永尾章曹(1975、1992)は、文章を質的統一体として捉え、以下のように二種三類の文章の基本的類型を提示している。この類型は、意味解釈や文、段落の形式を基準にしている他の研究者の類型と異なり、テクスト全体の特徴から対応する言語形式の特徴を見出しており、読解、作文等に応用できる言語モデルである。

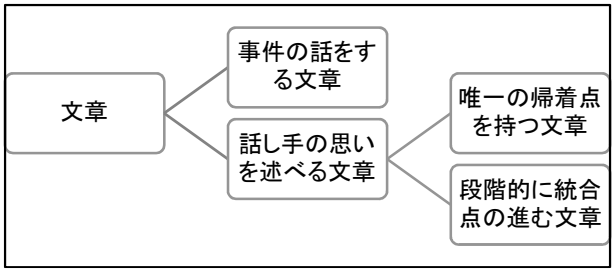


図 2　文章の基本的類型

「事件の話をする文章」は小説、童話、事件の報告に見られ、「話し手の思いを述べる文章」は、詩の形を取ることが多い「唯一の帰着点を持つ文章」と論説、評論、随筆として実現することが多い「段階的に統合点の進む文章」に分けられる。テクストの読解において、このモデルを適用し、文章構成を明らかにすることで、基本的には語彙の相互関係の数量的分析であるテキストマイニングの結果と対照させ、どの程度、特徴が帰納できるかを確認してみることにする。

3.1 文章の基本的類型とテキストマイニング[1]

以下では、「事件の話をする文章」の事例として、芥川龍之介「羅生門」、「話し手の思いを述べる文章」の事例として芥川龍之介「大川の水」を取り上げて、文章の基本的構成に基づいた文章構成の分析を行い、次に資料をテキストマイニングで分析した結果と比較して質的分析（意味的な内容理解）との関係から、相互関係を考察

[1] この節の内容は落合由治(2020a)、落合由治(2020b)で発表した内容を基にしている。

する。

3.2「羅生門」と「大川の水」の文章構成

3.2.1「羅生門」の文章構成

典型的な「事件の話をする文章」である「羅生門」は以下の表1のような文章構成を採っている。

表1　「羅生門」の文章構成

段落	内容	機能
Ⅰ	**羅生門で雨止みを待つ下人の登場**	**描写Ⅰ**
Ⅱ	門の周りに誰もないことの説明	説明
Ⅲ	誰もない理由の説明	説明
Ⅳ	羅生門の様子の説明	説明
Ⅴ	下人が羅生門にいる理由の説明	説明
Ⅵ	**ある日の暮れ方の羅生門の情景**	**描写Ⅱ**
Ⅶ	下人が羅生門にいる理由の説明	説明
Ⅷ	**下人の動き＋羅生門の情景**	**描写Ⅰ＋Ⅱ**
Ⅸ	**下人が羅生門の上にあがる動き**	**描写Ⅰ**
Ⅹ	**羅生門に上がった下人の動き**	**描写Ⅰ**
ⅩⅠ	**羅生門に上がった下人の動き**	**描写Ⅰ**
ⅩⅡ	羅生門の中の様子	説明
ⅩⅢ	**鼻をおおう下人の動き**	**描写Ⅰ**
ⅩⅣ	下人が老婆を見つける	説明
ⅩⅤ	老婆の様子の説明	説明
ⅩⅥ	下人の心理の説明	説明
ⅩⅦ	下人の心理の説明	説明
ⅩⅧ	**下人と老婆のやり取り（以下同じ）**	**描写Ⅰ**
ⅩⅩⅦ	**下人と老婆のやり取り**	**描写Ⅰ**
ⅩⅩⅧ	下人の喪失	説明

「羅生門」は、全部で28段落に分かれているが、特定の時の持続の中で、ある登場者が登場し、その動きを時の経過に従って描写するⅠ、Ⅷ、Ⅸ、Ⅹ、ⅩⅠ、ⅩⅢ、ⅩⅧ～ⅩⅩⅦの描写Ⅰの段落と、第Ⅵ段落、第Ⅷ段落の一部に見られる、特定の時の持続を止めてそ

の時空にある何らかの登場者を取り上げて気分を示す、いわゆる情景描写に当たる描写Ⅱが、「羅生門」の事件の話をしているストーリーである。一方、その他の各段落は、必要に応じてストーリーに様々な説明を加え、プロットを構成している説明の段落である。「事件の話をする文章」は、このように大きくはストーリーの表現とプロットの表現で構成されている。

3.2.2「大川の水」の文章構成

一方、「話し手の思いを述べる文章」である「大川の水」は、表2のように「羅生門」とはまったく違った文章構成をしている。

表 2　「大川の水」の文章構成

段落	内容	機能
Ⅰ	大川端に近い町に生まれ、「親しく思い出す」ことの説明	話題Ⅰ
Ⅱ	大川を「懐かしい思慕」のゆえに愛する理由の説明	話題Ⅱ
Ⅲ	川の眺めにこころを「おののかせた」思い出の説明	話題Ⅲ
Ⅳ	今でも「さびしい、なつかしい」大川に行く説明	話題Ⅳ
Ⅴ	大川の季節の光景に「さびしさ」を感じる説明	話題Ⅴ
Ⅵ	大川を「慕わしく」思い出す説明	話題Ⅵ
Ⅶ	「なつかしく」思い出す大川の様々な光景の説明	話題Ⅶ
Ⅷ	「なつかしく、さびしさを感じる」大川の渡しの説明	話題Ⅷ
Ⅸ	自分を「魅する」大川の光の説明	話題Ⅸ
Ⅹ	大川の「なつかしい」水の説明	話題Ⅹ
ⅩⅠ	大川の日暮れに「涙を流した」説明	話題ⅩⅠ
ⅩⅡ	大川と東京を「愛する」理由の説明	話題ⅩⅡ
ⅩⅢ	大川の渡しが消えて行く説明	話題ⅩⅢ

「大川の水」は、全部で 13 段落に分かれているが、それぞれ別々な話題を取り上げているわけではなく、すべて大川に関する自分の思いと思い出に関する内容である。それぞれの話題で取り上げている内容は異なるが、それらに抱く思いは「親しさ」、「懐かしさ」「さびしさ」「慕わしさ」など自分にとって極めて身近な存在に向ける愛情で共通しており、全体として故郷の東京の象徴である失われつつある大川の原風景への郷愁を語っていると言える。この

ように、「話し手の思いを述べる文章」は、異なる素材であっても共通点のある話題を並べて、それらに共通する焦点を浮かび上がらせる文章構成である。「事件の話をする文章」と「話し手の思いを述べる文章」の文章構成は、大きく異なっている。

3.3　「羅生門」と「大川の水」のテキストマイニング

以上の質的分析に対して、両作品にテキストマイニングを行った結果は以下のようになった。ここでは、共起ネットワーク分析の結果を示す。[2]まず、語彙の共起ネットワーク分析であるが、以下の図 3 のように、作品中での主要部分が抽出された(以下、ローマ数字は作品の段落、色は図の該当色を示す)。

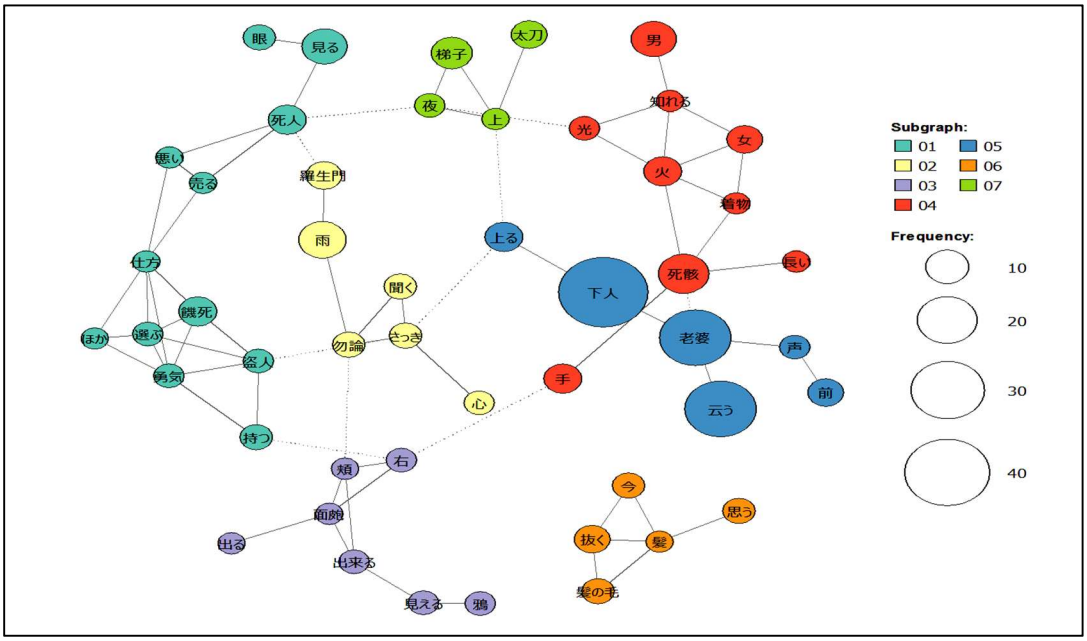


図 3　「羅生門」の共起ネットワーク分析の結果

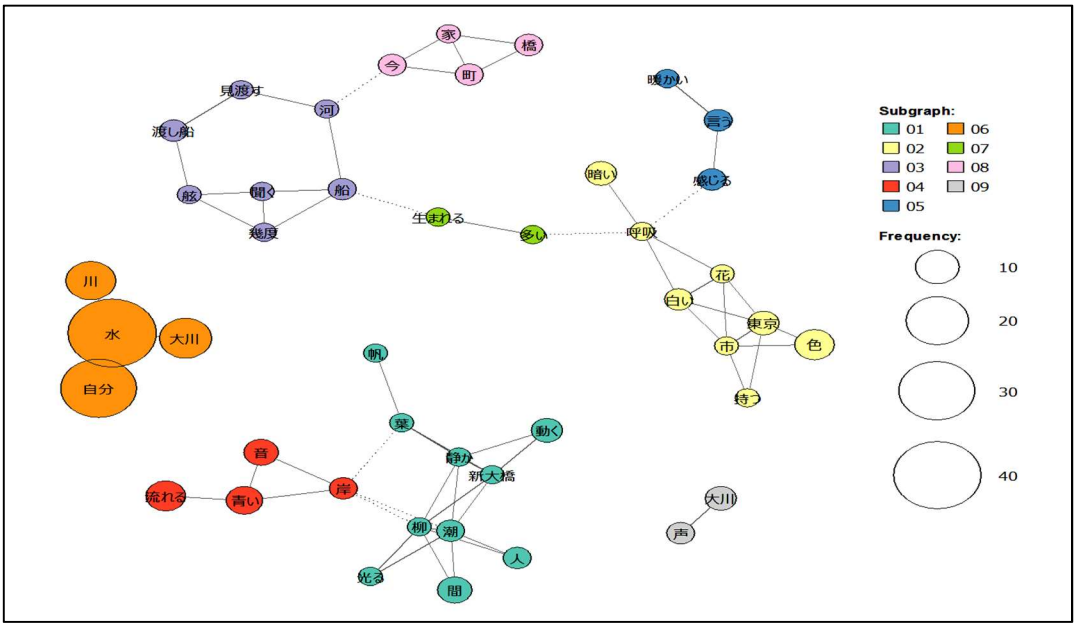
「羅生門」では、7 クラスターに分かれ下人の心理(左上:Ⅶ、XⅥ、XⅦ)、羅生門の情景(左中:Ⅴ)、下人の様子(左下:X)、羅生門に上がった様子(右上:XⅡ)、下人と老婆の対話(右中:XⅧ～XXⅦ)、死人から髪を抜く様子(右下:XⅤ)、階段を上る下人の様子(上中:Ⅸ)に主に関わる語彙が 7 クラスターに分かれ、各場面のキーワードの大きなグループが抽出された。しかし、これらは実際に作品を読んで各場面を理解していないと意味が理解できない、過

[2] テキストマイニングは、既成のプログラムを用い、共起ネットワーク分析には KH Coder を使用した。

渡的な分析である。

図４　「大川の水」の共起ネットワーク分析の結果

一方、「大川の水」では、大川と自分が各段落に共通する語彙(左中)として大きなクラスターになり、新大橋などⅧの内容(下中)、水に関わるＸの内容(右中)、渡し船に関わるＸＩの内容(左上)、青いなど川の水に関わる各段落の内容(左下)、暖かさを感じる各段



落の内容(右上)、大川に近い町で生まれたことに関わる町、橋、声などの各語彙(中、上中、右下)が 9 クラスターに分類された。これらは、各話題の中心的なキーワードで、これらの出ている部分を実際に読んでいなくても、主な内容は何か理解できる。

他の多くの作品と比べる必要があるが、「事件の話をする文章」の場合は、段落単位で見られる比較的内容の大きなまとまりが抽出されるが、その意味は作品との関係の中でしか理解できない。一方、「話し手の思いを述べる文章」の場合は、関連性のある主な語彙のグループがクラスターの中心になって取り出されやすいように思われる。つまり、文章構成によって、テキストマイニングで取り出される内容は、大きく異なるということである。また、内容の要点抽出に有効にテキストマイニングが機能するのは、「話し手の思いを述べる文章」の場合である。

では、こうした結果は、言語を超えて共通して見られるであろうか？

4.テキストマイニングを用いた日本語と中国語での文章構成の対照分析

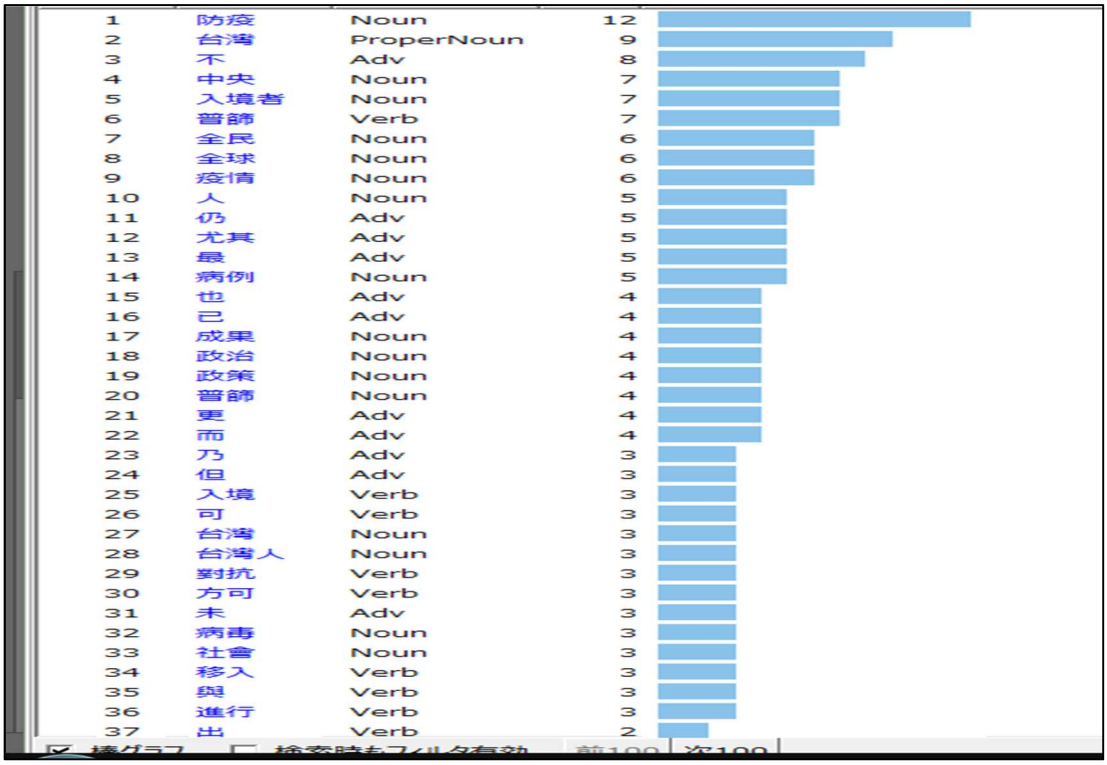
以下では、３節で用いたテキストマイニングツールとして社会科学分野の質的研究に広く使用されるようになっている樋口浩一(2020)の「KH Coder」を用いて、台湾の『自由時報』と日本の『読売新聞』の社説を資料にして、結果を比較した。テキストマイニングで最も有効なのは、「話し手の思いを述べる文章」であるため、論説文の例として社説を選び、分析をおこなった。

4.1 出現語彙の頻度による比較

まず、テキストマイニングの基本である、出現語彙の頻度からどの程度内容の要点が抽出できるか、分析をおこなった。

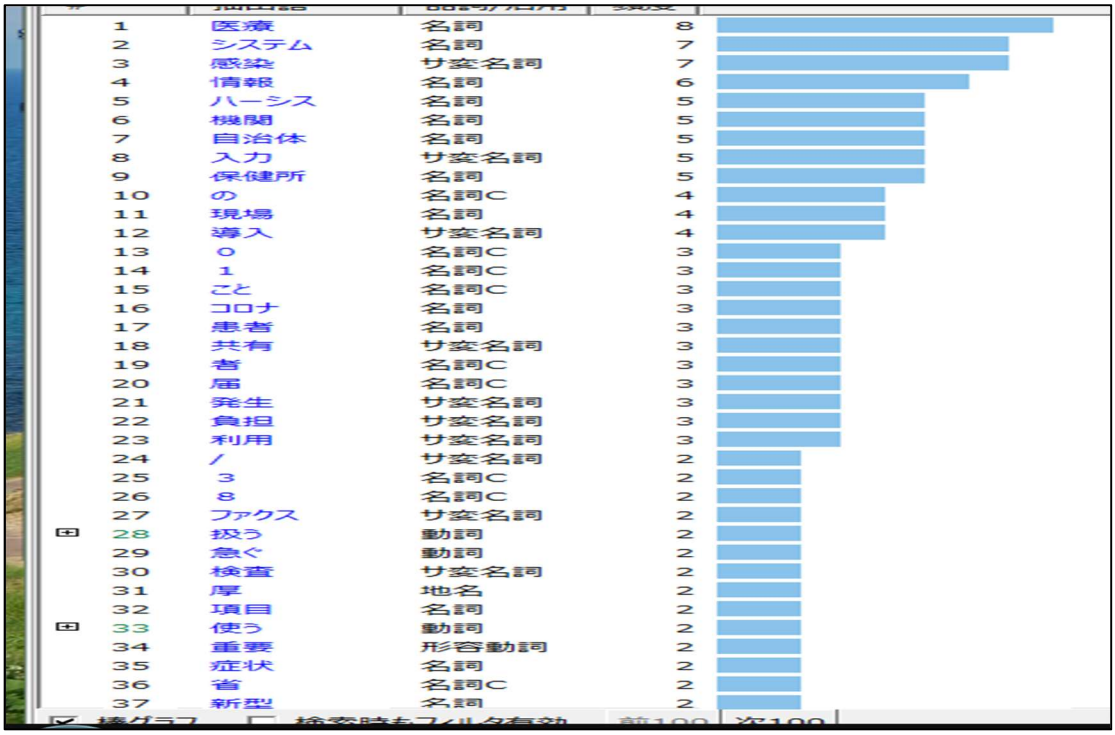
『自由時報』の語彙の出現頻度で見ると、「防疫」「台湾」から台湾の防疫に関するテーマで、「入場者」「普篩」「全民」「全球」等の語彙から海外からの入国者や台湾の全市民に PCR 検査を実施すべきかどうかが問題であり、「政治」「政策」でどうするのかを決める必要があるという、社説の主要話題が見出せる。また、「入境」「移入」等の語から、海外からの入国者や台湾の全市民に PCR 検査を実施すべきかどうかを論じていることが伺え、話題の問題の中心を出現語彙の頻度から推測できる。

『読売新聞』の場合も同じで「医療」「システム」「自治体」「保健所」等の語から新型コロナに対する政府自治体の検査体制を問題にしていることが分かる。また、「患者」「共有」「負担」「利用」などの語から、情報の共有や費用負担の問題など現場に解決すべき個々の問題が多くあることが窺える。また、「検査」「重要」などの語彙からは、現場の問題として検査態勢に関する細部の問題が多くあることが分かる。



1	防疫	Noun	12
2	台湾	ProperNoun	9
3	不	Adv	8
4	中央	Noun	7
5	入境者	Noun	7
6	普師	Verb	7
7	全民	Noun	6
8	全球	Noun	6
9	疫情	Noun	6
10	人	Noun	5
11	仍	Adv	5
12	尤其	Adv	5
13	最	Adv	5
14	病例	Noun	5
15	也	Adv	4
16	已	Adv	4
17	成果	Noun	4
18	政治	Noun	4
19	政策	Noun	4
20	普師	Noun	4
21	更	Adv	4
22	而	Adv	4
23	乃	Adv	3
24	但	Adv	3
25	入境	Verb	3
26	可	Verb	3
27	台湾	Noun	3
28	台湾人	Noun	3
29	對抗	Verb	3
30	方可	Verb	3
31	未	Adv	3
32	病毒	Noun	3
33	社會	Noun	3
34	移入	Verb	3
35	與	Verb	3
36	進行	Verb	3
37	出	Verb	2

図 5　『自由時報』社説の出現語彙の頻度



1	医療	名詞	8
2	システム	名詞	7
3	感染	サ変名詞	7
4	情報	名詞	6
5	ハーシス	名詞	5
6	機関	名詞	5
7	自治体	名詞	5
8	入力	サ変名詞	5
9	保健所	名詞	5
10	の	名詞C	4
11	現場	名詞	4
12	導入	サ変名詞	4
13	0	名詞C	3
14	1	名詞C	3
15	こと	名詞C	3
16	コロナ	名詞	3
17	患者	名詞	3
18	共有	サ変名詞	3
19	者	名詞C	3
20	届	名詞C	3
21	発生	サ変名詞	3
22	負担	サ変名詞	3
23	利用	サ変名詞	3
24	/	サ変名詞	2
25	3	名詞C	2
26	8	名詞C	2
27	ファクス	サ変名詞	2
28	扱う	動詞	2
29	急ぐ	動詞	2
30	検査	サ変名詞	2
31	厚	地名	2
32	項目	名詞	2
33	使う	動詞	2
34	重要	形容動詞	2
35	症状	名詞	2
36	省	名詞C	2
37	新型	名詞	2

図 6　『読売新聞』社説の出現語彙の頻度

以上、中国語と日本語の「話し手の思いを述べる文章」である新聞社説を比較した場合、言語の差異を超えてテキストマイニン

グの結果は、その文章中での中心的キーワードを抜き出すことができており、両者に差異は見られない。つまり、文章レベルでの文章構成は言語の差異を超えたレベルで共通した構成を持っていることが窺える。

4.2 共起ネットワーク分析による比較

続いて、テキストマイニングの中で、出現語彙の近さを測定することで資料中での主要キーワードをグループ別に抽出できる共起ネットワーク分析で、どの程度内容の要点が抽出できるか分析をおこなった。

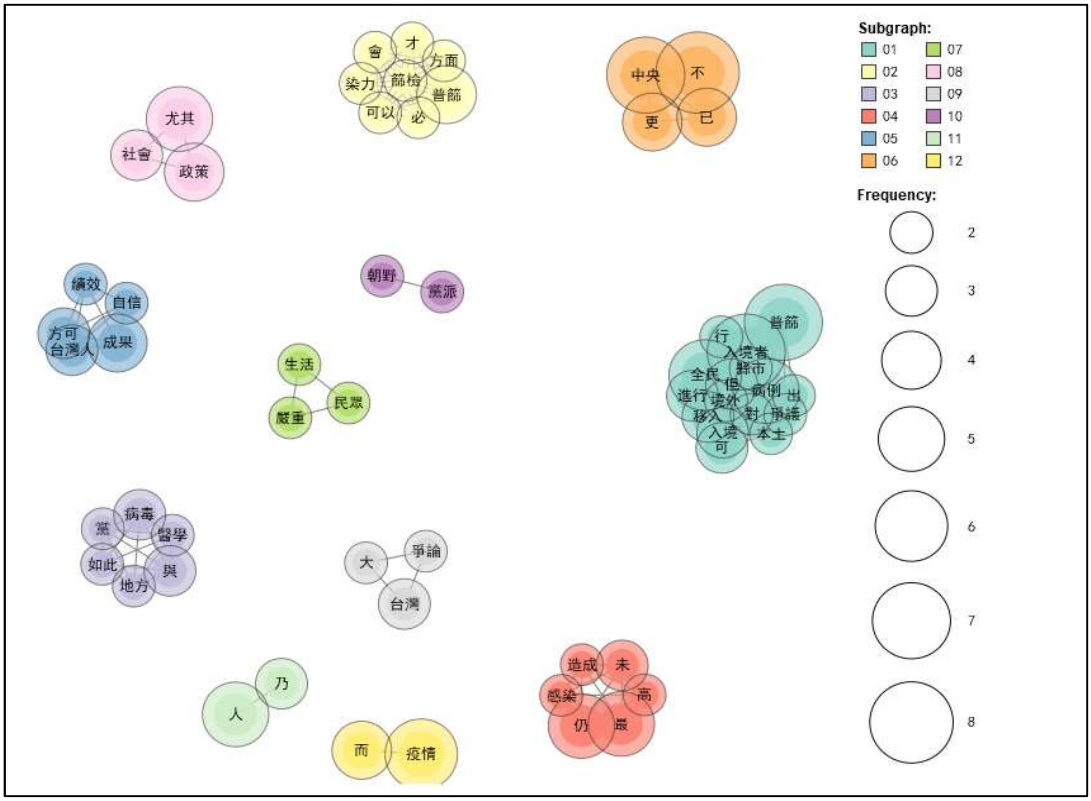


図 7　『自由時報』社説の共起ネットワーク分析

出現語彙の近さを測定する共起ネットワーク分析の結果でも両者の結果は同じで、『自由時報』では「普篩」関係の語彙（上中)、地方の問題に関する語彙(右中、左下)、「社会」「政策」「台湾」に関する語彙（左上、左中)、感染の危険に関する語彙（右下）など、主要な内容に関する語彙群が関係の深いグループごとに８つ抽出された。

『読売新聞』の場合も同様で、「コロナ」「情報」「共有」関係の語彙（上)、「医療」「患者」関係の語彙（右)、自治体の問題に

関する語彙（右下）、「システム」、「導入」に関する語彙（左下）、「保健所」「発生」に関する語彙（左上）などやはり主要内容のまとまりが８つ共起ネットワークとして抽出された。

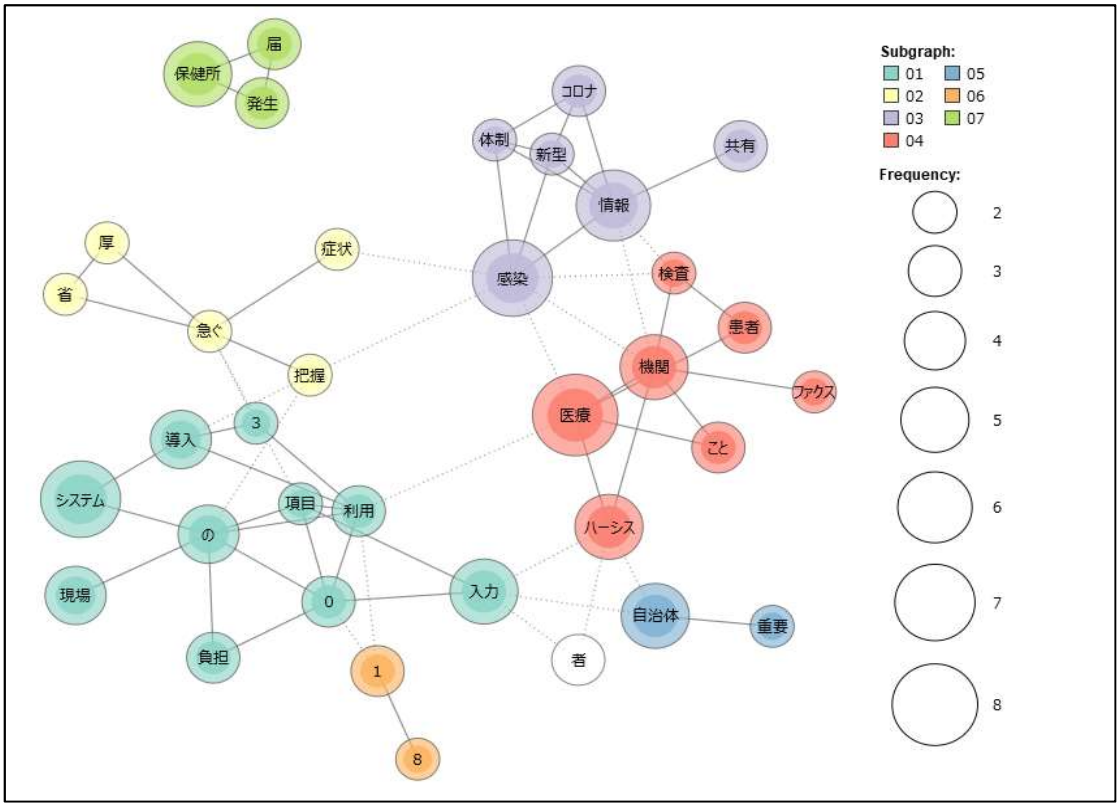


図８　『読売新聞』社説の共起ネットワーク分析

以上から、テキストマイニングを方法として、同じジャンルの文章構成を持つ新聞社説を分析した結果、中国語でも日本語でも結果はまったく同一であり、特定の文章ジャンルでの言語表現は言語の差異を超えていると考えることができる。こうした個々の言語の差異を超えて機能している言語表現のレベルが存在していることをこれは示しており、従来の言語間の差異のみに関心を向ける言語類型論とは別な研究課題を設定して、考察していく必要が言語比較において存在していると考えられる。文章あるいは談話という語、文を超えた言語単位とは何か、まだ未解明の部分が多く、今後の解明の手がかりになると言えよう。

5. おわりに

従来の言語比較は言語間の差異を前提にした比較対照であったが、実際に言語が使用される文章レベルでみると、テキストマイ

ニングの結果から文章構成には言語を超えて共通しているより上位の表現構造があると考えられる。新しい言語の次元について、今後、考察していくことで、言語とは何であるか今までのパラダイムを超えた領域の課題を見出すことができると思われる。本論は、考察の端緒を示したに過ぎないが、この問題は今後、開拓可能な新しい言語研究の領域であると考えられる。

注記

本論文は、2020 年度政治大学日本語文学科国際シンポジウム「東アジア漢字文化の伝承と創造」でのパネルディスカッションの内容に加筆訂正をおこなったものである。また、科技部研究案 MOST 109-2410-H-032 -061 -MY3 の成果の一部である。インターネット資料は 2021 年 1 月 30 日閲覧で確認している。

資料

『自由時報』2020 年 8 月 24 日「社論》平心靜氣論防疫」
https://talk.ltn.com.tw/article/paper/1395098

『読売新聞』2020 年 10 月 8 日社説「コロナ情報共有　現場が使いやすいシステムを」
https://www.yomiuri.co.jp/editorial/20201007-OYT1T50253/

テキスト

芥川龍之介「羅生門」「大川の水」『青空文庫』
https://www.aozora.gr.jp/index_pages/person879.html#sakuhin_list_1

参考文献

落合由治、「日本語テキストマイニング技術の応用可能性―台湾における日本語教育の次のステージを求めて―」、『比較文化研究』140、2020、1-12。

落合由治、『日本語の文章構成に関する基礎的研究―テクスト論と結合して』、致良出版社、2007。

落合由治，曾秋桂，王嘉臨，葉夌、「日本語テキストマイニング技術の文学語学教育分野への応用可能性の検討」、NLP2020 F2:

テーマセッション:言語教育と言語処理の接点、言語処理学会、2020a。

Kaplan, Robert, B., *The anatomy of rhetoric : prolegomena to a functional theory of rhetoric : essays for teachers,* Center for Curriculum Development, Philadelphia, 1972.

Greenberg, Joseph H. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, Joseph H. (Ed.)、 *Universals of Language,* MIT Press, Cambridge,1963,73–113.

郡司隆男、「日本語はどんな言語か?:類型論的観点からの日本語」、『Theoretical and applied linguistics at Kobe Shoin:トークス』14、2011、1-14。

郡司隆男・坂本勉、『言語学の方法』、岩波書店、1999。

Koenig, H. B., *Untersuchungen,* New edition with an introductory article by Konrad Koerner, John Benjamins published, Amsterdam,1983.

Schleicher, August. *Die Sprachen Europas in systematischer Uebersicht, Linguistische*, 1850, by Internet Archive https://archive.org/details/diespracheneurop00schl/mode/2up.

Gleason, H. A. Jr, Contrastive Analysis in Discourse Structure, Georgetown University Monograph Series in *Languages and Linguistics*, 19,1968, 39-63.

永尾章曹、『国語表現法研究』、三弥井書店、1975。

永尾章曹編著、「第四章日本語の文法について」、『日本語学』、和泉書院、1992。

西原鈴子、「日英対照修辞法」、『日本語教育』72、1990、25-41。

Hartmann, Reinhard, Contrastive Textology: Comparative Discourse Analysis in Applied Linguistics (Studies in Descriptive Linguistics Vol. 5), Groos, Heidelberg,1980.

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第二版』、ナカニシヤ出版、2020。

Hinds, John, Inductive, deductive, and quasi-inductive: Expository

writing in Japanese, Korean, Chinese, and Thai. In Coherence in *Writing: Research and Pedagogical Perspectives* by Ulla Conner & Ann M. Johns (Ed), Alexandria, Virginia: Teachers of English to Speakers of Other Languages, 1990, 89-109.

泉子・K・メイナード、『談話分析の可能性』くろしお出版、1977。

本名信行、「日本語の文体と英語の文体」、『講座日本語と日本語教育』5、明治書院、363-385。

李在鎬編、『文章を科学する』、ひつじ書房、2017。