

# テキストのジャンルとメタファー表現のコレスポネンス分析 —「関係」のメタファーを例に—

明星大学 情報学部 情報学科  
大石 亨

## 1 はじめに

近年、大規模なコーパスが整備されるようになったことにより、ジャンルやテキストタイプによる言語現象の違いや類似性の分析に注目が集まってきている。英語では、主に経済記事を対象としたメタファーのコーパス分析が行われてきた。Charteris-Black (2000) は、国際政治と経済を中心に扱うイギリスの週刊誌 *Economist* と、Bank of English の一般雑誌サブコーパスを比較することにより、ビジネス英語に特有のメタファーに基づく表現を発見している。Skorczynska and Deignan (2006) は、大衆向けの一般経済雑誌と研究者を対象とする経済研究専門誌から、同じトピックに関する記事をそれぞれ 40 万語程度になるように抽出して、その中で使われているメタファーを比較している。この結果、それぞれの雑誌で使用されるメタファー表現やその機能は大きく異なっており、想定される読者やテキストの書かれる目的がメタファー選択に影響を与えると論じている。

日本語に関しては、大規模コーパスを用いたジャンル別のメタファー研究はほとんど行われておらず、特定語句の使用率によるテキストジャンルの判別や、ジャンル別の使用語彙に関する調査がようやく始まった段階である。村田 (1999, 2000) は、接続語句と助詞相当句の使用率を指標とし、多変量解析の一手法である正準判別分析によってジャンルを判別している。村田 (2000) のジャンルとは、「経済学教科書」「物理学論文」「工学論文」「文学作品」「新聞社説」の五つである。小磯他 (2009) は『現代日本語書き言葉コーパス(BCCWJ)』と『日本語話し言葉コーパス(CSJ)』を対象に、ジャンルごとの語種や品詞の比率を算出し、線形判別分析によってジャンルを判別するモデルを構築している。また、内田・藤井 (2009) は、BCCWJ の書籍サブコーパスと教科書サブコーパスを対象に、ジャンルの特徴語の抽出とクラスター分析を行っている。小磯らのジャンルは、「行政白書」「新聞」「小説」「WEB データ (Yahoo!知恵袋)」「国会会議録」という BCCWJ の 5 種類のサブコーパスに、「学会講演」「模擬講演」という CSJ の 2 種類のサブコーパスを加えた七つであるのに対し、内田らのジャンルは BCCWJ の「書籍」サブコーパスに付された「総記」「哲学」「歴史」など、日本十進分類法 (NDC) による第 1 次区分 (類) 10 個と、「教科書」サブコーパスの「国語」「数学」「理科」などの 8 科目のことであり、小磯らに比べて一段階細かいレベルの分類に対応している。このように、研究者によって「ジャンル」という用語の定義が異なっているうえに、研究分野による術語の違いもあり、混乱にさらに拍車をかけている。

社会言語学では、人間が用いる言語に現れるさまざまな違いは「バラエティ・バリエーション」あるいは「変異・変種」と呼ばれる。言語変種はその要因によって区別され、主に語られる内容や話題に関連する「ジャンル(genre)」のほかに、特定の集団や職業に関連する「レジスター(register)」や、形式性(formality) に関連する「スタイル(style)」という用語が用いられることが多い (Holmes 2008)。一方、日本語学では、目的や場面など、言語使用に関わる言語以外の属性や条件の類型は「位相」と呼ばれ、それに対応する文章の類型は「文体」「文章体」などと呼ばれている (半沢 2009)。分野による術語の違いに加えて、個々の研究者による術語使用の違いがあるうえに、特定の言語表現には複数の要因が混在しており、複数の観点による分類が常に可能であるため、これらの用語の規定の曖昧さは繰り返し指摘されてきている。

本稿では、これらの術語を厳密に規定することはあえてせず、「テキストジャンル」という用語を「すでに与えられたもの」として、階層的に組織されたコーパス分類を代表する名称として用いる。すなわち、第 1 段階の分類としては、小磯らと同様に BCCWJ のサブコーパスを、第 2 段階の分類としては、内田らと同様に BCCWJ の「書籍」サブコーパスデータに付与された NDC (日本十進分類法) の「類」を表すものとして、2 段階の分析を行う。具体的なデータとその区分については、次節で詳しく述べる。

また、本稿で取り上げるのは、「関係」という語を含む日本語のメタファー表現である。関係概念は、非常に身近でありふれたものでありながら、抽象的なものであり、多様なメタファー表現が用いられている。われわれは、関係の網の目のなかで生活していると言っても過言ではない。なかでも、特に多くの用例が存在する<構築・破壊>を表す語彙と、<結合・切断>を表す語彙を分析対象とし、テキストのジャンルごとに頻度のクロス集計表を作成した。この集計表に、多変量解析手法の一つであるコレスポネンス分析を適用することにより、同一のメタファーを実現するさまざまな表現が、テキストジャンルごとに使い分けられている様子を視覚的に把握することが可能になる。関係のメタファーについては第3節で、コレスポネンス分析については第4節で、それぞれ解説する。第5節は分析結果と考察である。5.1節でBCCWJのサブコーパスを対象とした第1段階の分析結果を、5.2節では書籍サブコーパスのNDC分類に基づく第2段階の分析結果について述べる。第6節はまとめである。

## 2 分析対象コーパス

分析データとして使用したコーパスは、大学共同利用機関法人 人間文化研究機構 国立国語研究所によって開発された『現代日本語書き言葉均衡コーパス(BCCWJ)』モニター公開データ(2009年度版)である(山崎 2006, 山崎他 2006, 国語研 2009)。このコーパスは、書き言葉の「生産」の側面を捉えるための生産実態サブコーパス(1300万語)、書き言葉の「流通」の側面を捉えるための流通実態サブコーパス(1500万語)、および、特定目的の日本語を収集した非母集団サブコーパスとして、ベストセラー(230万語)、行政白書(480万語)、Yahoo!知恵袋(520万語)、国会会議録(490万語)から構成されている。このうち、生産実態サブコーパスは、2001年から2005年の間に国内で発行された書籍であり、国立国会図書館の蔵書目録を電子化した「J-BISC」をもとに決定された母集団から、NDC および発行年度ごとに層別に抽出されたものである。流通実態サブコーパスは、書き言葉の流通実態を図書館の所蔵状況で捉えるものであり、東京都立中央図書館の作成した「ISBN総合目録」から、1986年から2005年までの20年間に発行された書籍から抽出されたものである。ベストセラーは、『出版年鑑』(出版ニュース社)および『出版指標年鑑』(全国出版協会出版科学研究所)のどちらかに、1976年から2005年までの30年間において、各年のベストセラーとして上位20位までに挙げられた書籍から抽出されたサンプルである。これら3種類のサブコーパスは、いずれも出版された書籍からサンプリングされたものであり、「書籍」サブコーパスとして一つにまとめて利用した。「行政白書」は、1976年から2005年までの30年間に発行された政府系刊行物「白書」1,006冊からランダムにサンプリングされた1,500サンプルである。「Yahoo!知恵袋」は、ヤフー株式会社より提供されたWeb上の掲示板Yahoo!知恵袋のデータ(2004年10月から2005年1月にかけて投稿された3,120,839の質問とそれに対する複数の回答)から45,725サンプルが抽出されたものである。また、「国会会議録」は、1976年から2005年までの30年間における国会会議録のうち、国立国会図書館より提供された第77回国会から第163回国会までに開かれた32,986会議の会議録から抽出された159会議である(山崎 2011)。この4種類のサブコーパスを第1段階のテキストジャンルとする。総語数4520万語(短単位)の内訳は、書籍:約3,030万語(10,423サンプル)、白書:約480万語(1,500サンプル)、Yahoo!知恵袋:約520万語(45,725サンプル)、国会会議録:約490万語(159サンプル)となる。

それぞれのサンプルには、以下の3種類の属性情報が収録されている。

- 書誌情報(サンプルを抽出した原本のタイトル、著者、出版年、ISBN、NDCコードなど)
- サンプル情報(サンプルID、原本におけるサンプルの抽出位置、著作権許諾状況など)
- 著者情報(著者ID、著者名、性別、生年代)

書籍サブコーパスのサンプルに付された書誌情報のうち、日本十進分類法(NDC)による「0 総記」「1 哲学」「2 歴史」「3 社会科学」「4 自然科学」「5 技術・工学」「6 産業」「7 芸術・美術」「8 言語」「9 文学」という10個の「類」を第2段階のジャンルとした。それぞれのジャンルの母集団の構成比は表1のとおりである(丸山・柏野 2011)。BCCWJの書籍はNDCと発行年をもとにランダムサンプリングされているため、任意に抽出した部分も母集団のNDCの構成比をそのまま表していることが期待される。表1の第2列は、出版サブコーパスの母集団となる、2001年から2005年の間に国内で出版された書籍・雑誌・新聞の総文字数を推計したデータに基づくジャ

ンル構成比である。ここでは、書籍データのみを対象としているので、国立国会図書館に所蔵されている書籍のうち、2001年から2005年に発行されたものの推計総文字数485.4億文字に対する各ジャンルの推計文字数の割合が記載されている。第3列は、流通サブコーパスの構成比であり、出版サブコーパスの書籍の母集団である約485.4億文字に最も近い母集団を得ることができる、東京都内13自治体の図書館が共通に所蔵する書籍の集合を対象とした推計総文字数の構成比である。モニター公開データでは、2009年時点で著作権処理の終わったものだけが収録対象とされているので、最終的な構成比は表1とは異なるが、おおまかな傾向は把握できるであろう。

表1 書籍データの母集団の構成比（丸山・柏野 2011より抽出・計算）

層	出版 SC の推定構成比	図書館 SC の推定構成比
0. 総記	3. 37%	2. 10%
1. 哲学	5. 35%	4. 90%
2. 歴史	8. 86%	10. 47%
3. 社会科学	25. 56%	18. 69%
4. 自然科学	10. 44%	6. 33%
5. 技術・工学	9. 51%	6. 58%
6. 産業	4. 52%	3. 53%
7. 芸術・美術	6. 71%	8. 47%
8. 言語	1. 83%	2. 00%
9. 文学	19. 24%	32. 34%
n. 記録なし	4. 59%	4. 61%
合計	100%	100%

### 3 関係のメタファー

本研究では、「関係」という語を含むメタファー表現を分析対象としている。関係のメタファーについては、これまでも内省やアンケート調査、Web検索等による言語表現の収集と分類がなされてきた。水野他（2006）は、内省、アンケート調査、各種出版物の用例調査に基づいて、人と人との関係の変化や状態を表すメタファー表現を採取し、青空文庫とGoogle検索によって用例を追加している。水野らは、人間関係のメタファーの根源領域を、「距離」「手」「経路」「磁石」「糸」など、14種類に分類したうえで、イメージスキーマとプライマリー・メタファーによる分析を行っている。それによると、人間関係のメタファーは、経路型、構築物型、環境型の三つに分類でき、経路型はリンク、構築物型は存在／個体、環境型は容器のイメージスキーマとそれぞれ関わりがあるが、すべてのメタファーをイメージスキーマで説明することはできず、プライマリー・メタファーを補完的に利用することにより、より複合的で経験基盤的な動機づけを理解することができると述べている。

鍋島（2005, 2011）は、関係のメタファーを題材にして、認知メタファー理論に概念レベルと知覚レベルの区分を導入することを提案し、この区分を利用した用例の分析を提示している。鍋島（2011）が取り上げているのは、「線（糸、紐、鎖）」「建物」「遠近」「力（密な、濃い、強い）」という4つのメタファー群であるが、このうち、「線」と「建物」を概念レベル、「遠近」と「力」は《親密さは近さである》《感情は力である》という知覚レベルのメタファーとして2分している。ここで、「概念レベルは、具体物に関するフレーム的知識（概念領域）や、反対語、同意語等、さまざまな意味の体系の中に置かれたレベルであり、ここでは主に構造的関係が作用する。一方、知覚レベルとは、五感、快・不快、行動など、身体性に基づいたレベルである」と述べている（鍋島2011, p.314-315）。これにより、基盤性の問題、プライマリー・メタファーの問題、アナロジーの問題、「関係」と共起する語彙の非対称性の問題など、過去の理論的問題が統一的に解決されると主張している。フレーム的知識の多くは視覚的なものであるし、「快・不快」や「行動」を「知覚レベル」と称することにも違和感を禁じ得ないが、身体と密接に結び付いたレベルと、意識的な概

念推論の働くレベルを一旦分離したうえで、両レベルの協業と多重制約充足過程によってメタファーの推論機能と喚起力を捉えようとする試みは、斬新なものである。今後、脳科学の進展によってもたらされるであろう知見との整合性を保ちながら、より詳細な機構が解明されることが望まれる。

さて、本稿が目指すのは、水野らが「糸」「建物」と呼び、鍋島が「線」「建物」と呼ぶ根源領域に関するものである。両者とも、メタファーの従来の規定方法に従って、「《関係は線(糸)である》《関係は建物である》」という標識を付けるための具体物を取り上げているが、これらのメタファーが問題にしているのは、糸や建物という対象ではなく、それらに働きかけた結果実現される機能や、対象に働きかける様態、目的、帰結等、人間の主観的なかわり方であると、われわれは考える。すなわち、「糸」と共起する動詞が関係のメタファーに使われるのは、2つのものを結びつけ、つなげるという機能的な側面を重視するためであり、「建物」と共起する動詞が使われるのは、積み重ねて作り上げたものが一気に崩れるという生成と消滅、および安定性の側面を前面に出すためである。糸や建物が想起されるのは、これらの行為や現象を表す動詞を用いることの付随現象に過ぎない。メタファー表現にとって重要であるのは、これらの動詞によって表される行為の目的や、その行為のもたらす結果に対して人間が感じる思いであって、関係概念が糸や建物に直接対応づけられることではない。そこで、以下では、前者を<結合・切断>語彙、後者を<構築・破壊>語彙と呼び、領域写像というよりはむしろ、これらの語彙の指示する行為が実現する状況を重視していることを明確にする。具体的には、全コーパス中に10回以上「関係」という語と共起し、メタファー表現として出現した、「結ぶ」「断つ(絶つ)」「断ち切る」「断絶する」「切れる」という<結合・切断>語彙5語と、「構築する」「築く」「樹立する」「修復する」「崩壊する」「崩れる」「壊れる」「破壊する」「壊す」「ぎくしゃくする<sup>1)</sup>」という<構築・破壊>語彙10語を分析対象とした。本研究では、これら15語のメタファー表現としての出現状況、すなわち「関係」と共起した回数をジャンルごとに集計してクロス表を作成し、コレスポネンス分析を適用した。それによって見えてきたものは、人間の思いを表現するメタファー表現が、社会的状況によっていかに制約されているかという事実である。次節では、コレスポネンス分析の原理を説明する。

#### 4 コレスポネンス分析

大量のデータを分類・整理・縮約することでデータの全体像を掴んだり、事物の間に潜む相互関係や、変数間の相互関係等を顕在化させたりするための統計手法は、多変量解析と呼ばれるが、コレスポネンス分析もその一つである。コレスポネンス分析は、クロス集計表の行や列に含まれる情報を、次元と呼ばれる少数の成分に圧縮し、それらの関係を散布図上に布置することで、視覚的なデータの俯瞰を可能にする手法である(石川他(編)2010, 内田2006, 高橋2005)。この際、データの性質の違いが問題にならず、カテゴリーデータ(質的データ)の分析に使うことができるという特徴を持つために、アンケートデータの分析などに多用されている。

本稿では、サブコーパスやNDC分類を個体、前節で述べた語彙項目の生起頻度を変数とし、個体同士や変数同士の関係とともに、個体と変数間の相互関係を視覚化して分析することを目的とする。ただし、コレスポネンス分析では、行と列の区別は便宜的なものであるので、変数すなわち語彙項目を第1アイテム、個体すなわちテキストジャンルのことを第2アイテムと呼ぶ。

ここでは、次節で解説することになる実際のデータを用いて、コレスポネンス分析の原理を解説する。表2は、第1列に記載したサブコーパスにおける<結合・切断>語彙のメタファー使用の回数、すなわち「関係を結ぶ」などの生起頻度をクロス集計したものである。

表2 サブコーパスと<結合・切断>語彙のクロス集計表

コーパス/語彙	結ぶ	断(絶)つ	断ち切る	断絶	切れる	計
Yahoo!知恵袋	1	4	1	0	0	6
国会会議録	4	0	0	8	5	17
白書	3	0	1	5	0	9
書籍	111	16	11	18	5	161
計	119	20	13	31	10	193

コレスポネンス分析が行うことは、この表の行と列それぞれの並び順を入れ替えて、表3のように、対角線上にできるだけ大きい数字を集中させることである。こうすることで、表に潜んでいたパターンが把握しやすくなり、いわば、第1アイテムと第2アイテムが強く関係し合う視点が作り出されるのである。表3では、国会会議録と白書が第2行と第3行に移動し、表2では離れていた書籍とYahoo!知恵袋が第4行と第5行に移動している。列では、「切れる」が第2列に移動し、「断つ(絶つ)」と「断ち切る」が第6列と第5列に移動している。同じような語の出てくるジャンルと、同じようなジャンルに出現する語を同時に集めようとしているわけである。

表3 行と列の並び換え(1)

コーパス/語彙	切れる	断絶	結ぶ	断ち切る	断(絶)つ	計
国会会議録	5	8	4	0	0	17
白書	0	5	3	1	0	9
書籍	5	18	111	11	16	161
Yahoo!知恵袋	0	0	1	1	4	6
計	10	31	119	13	20	193

しかし、1回の並び替えだけでは(1次元の尺度だけでは)すべてのパターンを表しきれない。そのため、2回目、3回目…(2次元、3次元…)と並び替えを行って、次元を拡張する。表4は、第2次元の尺度で行った並び換えの結果である。表3とは異なる並び順が得られている。

表4 行と列の並び換え(2)

コーパス/語彙	断(絶)つ	切れる	断ち切る	断絶	結ぶ	計
Yahoo!知恵袋	4	0	1	0	1	6
国会会議録	0	5	0	8	4	17
白書	0	0	1	5	3	9
書籍	16	5	11	18	111	161
計	20	10	13	31	119	193

実際には、並べ替えは両アイテムのカテゴリーにそれぞれ重み(スコア)を与えることによって行われる。表5は第1アイテムである<結合・切断>語彙に与えられたスコアであり、表6は第2アイテムであるサブコーパスに与えられたスコアである。これらの表の次元1のスコアをもとに並び替えられたのが表3であり、次元2のスコアによる並び換えの結果が表4である。

表5 第1アイテムのスコア

	次元1	次元2	次元3
結ぶ	-0.333	-0.6428	-0.257
断つ(絶つ)	-1.3148	2.4994	-0.0931
断ち切る	-0.6137	0.4452	1.1962
断絶	1.5657	0.291	1.5338
切れる	2.5365	1.1699	-3.0568

表6 第2アイテムのスコア

	次元1	次元2	次元3
Yahoo!知恵袋	-2.0785	5.1618	0.4501
国会会議録	2.8225	1.0422	-1.1402
白書	1.388	-0.01	4.3033
書籍	-0.2982	-0.3019	-0.1368

行と列の並び換えと重み(スコア)の付与は、データの行と列の相関を最大化することにほかならず、数学的には「固有値問題」を解くことに相当する。固有値(eigenvalue)は、最大 $k$ 個(第1アイテムと第2アイテムのカテゴリー数のうち小さいほうの数、上の例では4個)求められるが、最大の固有値1を除いた $k-1$ 個の固有値に対応する固有ベクトル(eigenvector)が各アイテムに与えるスコアを定める。

その後、この次元ごとのスコアを組み合わせせて散布図を作成し、分析結果を視覚化する。通例、第1次元を横軸に、第2次元を縦軸にした2次元空間に各アイテムのカテゴリーがプロットされる。

## 5 分析結果と考察

本節では、関係のメタファーに用いられた<構築・破壊>語彙と<結合・切断>語彙を、第2節で解説したテキストジャンルごとに集計したクロス集計表に、前節で解説したコレスポネンス分析を適用した結果および考察を述べる。5.1節はBCCWJのサブコーパスを対象とした第1段階の分析結果、5.2節は書籍サブコーパスのNDC分類に基づく第2段階の分析結果である。

### 5.1 サブコーパス別の分析

表7は、<構築・破壊>語彙の出現頻度をサブコーパスごとに集計したクロス集計表である。第2節で述べたように、書籍サブコーパスの語数が多いために、分析対象語彙のメタファー表現としての出現数も書籍サブコーパスが全体の65.7% (270/411)を占めている。表7から、「関係を築く」という言い回しが最も多く使われていることがわかる (162/411=39.4%)。次に多いのは「関係を構築する」(97/411=23.6%)と「関係を樹立する」(41/411=10.0%)であるが、この2つは、特に行政白書での使用比率が高い。行政白書は、サブコーパスの語数の割合に対して、メタファー表現の出現比率が高い(10.6%:17.2%)なのであるが、そのうちの約66% (= (26+21)/71)が「構築」と「樹立」の2語によって占められている。

表7 <構築・破壊>語彙のサブコーパス別使用状況

コーパス/語彙	構築	築く	修復	崩壊	破壊	壊れる	崩れる	樹立	ぎくしゃく	壊す	計
Yahoo!知恵袋	0	18	1	1	0	5	3	0	2	7	37
国会会議録	9	11	2	0	0	1	2	7	1	0	33
白書	26	18	6	0	0	0	0	21	0	0	71
書籍	62	115	12	13	10	7	11	13	18	9	270
計	97	162	21	14	10	13	16	41	21	16	411

一方、表8は、<結合・切断>語彙のサブコーパス別のクロス集計表である。ここでも、書籍サブコーパスが、全体の83.4% (161/193)を占めている。この割合は、語数の割合(67.0%)に比しても大きな値である。また、「関係を結ぶ」という表現が全体の約61.7% (=119/193)を占めているが、これは結合を表す表現が、「結ぶ」1語しかないことが影響しているかもしれない。「関係を結合する」「関係をつなぐ」などとはめったに言わないのである<sup>2</sup>。これに対して、切断を表す語彙は4つの語彙に分散して用いられているが、そのなかで最も多いのは「関係が断絶する」(31/193=16.1%)で、国会会議録で相対的によく使われている(8/31=25.8%)。

表8 <結合・切断>語彙のサブコーパス別使用状況

コーパス/語彙	結ぶ	断(絶)つ	断ち切る	断絶	切れる	計
Yahoo!知恵袋	1	4	1	0	0	6
国会会議録	4	0	0	8	5	17
白書	3	0	1	5	0	9
書籍	111	16	11	18	5	161
計	119	20	13	31	10	193

これらのクロス集計表にコレスポネンス分析を適用した結果を表9と図1, 2に示す。分析には、「多変量解析システム Seagull-Stat」(<http://www.jomon.ne.jp/~hayakari/>)を用いた。このシステムは、早狩進氏によって開発され、Excel上で動作するマクロ・プログラムであり、石川他(編)(2010)の付属CD-ROMに収録されているものである。表9に示した固有値が、それぞれの次元が語彙とサブコーパスの連関をどれくらい説明できるかを表す数値である。寄与率(contribution rate)は固有値を百分率に加工した値で、累積寄与率(cumulative contribution rate)はその次元までの寄与率の総和である。寄与率が大きいほど、分析対象のクロス集計表が有していた情報が集

約されていることを表しており、次元 1 → 次元 2 → 次元 3 … の順に寄与率の値は必ず小さくなる。表 9 の左側に示した「構築・破壊」語彙では次元 2 までに 97.17% の連関が、右側に示した「結合・切断」語彙では次元 2 までに 88.85% の連関が、それぞれ説明されていることを表している。累積寄与率の値がどの程度であれば分析がうまくいったという統計的な基準は存在しないが、第 2 次元までで 80% 程度の値が出ていれば、おおむね良好な結果であると推定される。

表 9 固有値と寄与率 (サブコーパス別分析)

〈構築・破壊〉語彙	次元 1	次元 2	次元 3	〈結合・切断〉語彙	次元 1	次元 2	次元 3
固有値	0.2192	0.068	0.0084	固有値	0.2476	0.1001	0.0436
寄与率(%)	74.17	23	2.83	寄与率(%)	63.27	25.58	11.15
累積寄与率(%)	74.17	97.17	100	累積寄与率(%)	63.27	88.85	100

図 1 は、〈構築・破壊〉語彙に対する散布図であり、図 2 は〈結合・切断〉語彙に対する散布図である。図の下方に「第 1 群」と示された青い○が対象語彙を表し、「第 2 群」の赤い 2 重丸◎が 4 つのサブコーパスに対応しており、グラフには両方が重ねて表示 (biplot) されている。コレスポネンダ分析の布置図では、マスの比率 (平均の比率) に近いカテゴリーほど原点(0,0)に近くなるという特徴があり、一般に、件数の多いものほど原点に近づく傾向がある。書籍サブコーパスや「築く」、「結ぶ」が原点付近にプロットされているのはこのためである。

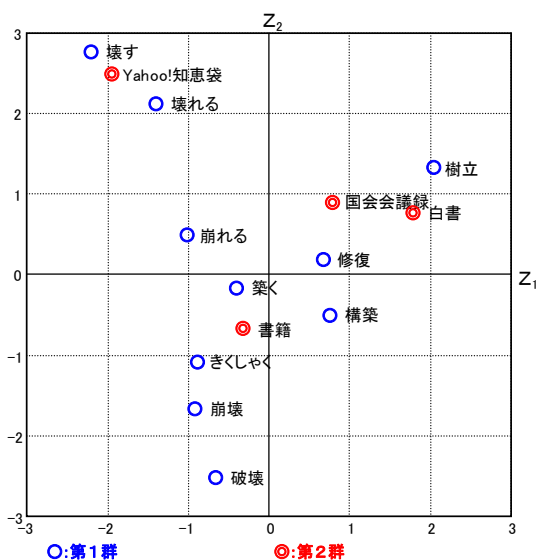


図 1 サブコーパス別の散布図 (〈構築・破壊〉語彙)

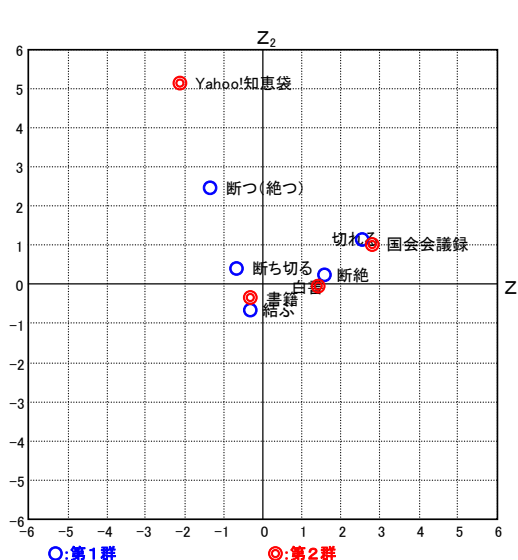


図 2 サブコーパス別の散布図 (〈結合・切断〉語彙)

図 1 では、横軸の正の部分 (右側) に行政白書と国会会議録が、負の部分 (左側) に書籍と Yahoo! 知恵袋がそれぞれプロットされており、第 1 次元は「公私」の区別に対応するものと考えられる。第 1 次元の得点が正の値を持つのは、「樹立」「構築」「修復」という構築または維持を表す漢語語彙であり、これらは外交関係など、公的な関係に用いられることの多い語である。これに対し、負の大きな値を持つものは「壊す」「壊れる」「崩れる」などの破壊を表す和語動詞である。ここから、公的な文書は漢文を基本とした擬似漢文で書き記し、私的な文章は日常の言葉である和語を仮名書きするという、千年におよぶ日本語の書記システムの伝統が、現在もなお脈々と息づいていることが読み取れる (金水他 2008)。さらに、公的な関係は構築し、維持することが問題とされることが多いのに対し、私的な関係は、破壊に言及されることが多いこともわかる。私的な関係は、公的な関係のように意識して作り上げるものではなく、自然に形作られる一方で、ささいなきっかけで壊れやすいものでもあるということを反映していると思われる。一方、第 2 次元 (縦軸) では、Yahoo! 知恵袋が大きく上方 (正方向) にはずれている。第 2 次元の正方向の特徴は「壊す」「壊れる」に

よって、負方向の特徴は「破壊」「崩壊」「ぎくしゃく」によって、特徴付けられている。第2次元は、私的言語のうち、インターネット上の言葉と書籍では使用語彙が異なることを表している。これは、Yahoo!掲示板が、質問応答型の掲示板であり、人間関係に悩みを持つ個人が、壊れそうな関係についての相談を書き込むことが多いためであると考えられる。これに対し、「破壊」「崩壊」という漢語は、1個の例外を除いて、書籍でのみ用いられている。

図2でも、第1次元（横軸）の正負によって、国会会議録、行政白書と、書籍、Yahoo!知恵袋が区分されている。第1次元の正方向の特徴は「切れる」「断絶する」によって、負方向の特徴は「断（絶）つ」「断ち切る」によって、それぞれ特徴付けられている。ここでも、第1軸は公私の別を表すと考えられるが、和語の「切れる」が公的な次元に付置されているのは、国会会議録中の法務委員会における親族関係への言及が含まれているからである。親族関係は本来私的なものであるが、民法という法律によって取り扱われることにより、公的な場に登場することとなり、国会会議録は話し言葉を記録したものであるから、和語である「切れる」が用いられたと考えられる。また、第2次元（縦軸）でYahoo!知恵袋が大きく上方（正方向）にはずれていることも図1と共通している。第2次元を正方向に特徴づけるのは「断（絶）つ」であるが、これは、個人的な男女関係の清算に用いられている。しかし、図2において、Yahoo!知恵袋が外れているのは、全体の件数が少ないためでもある。Yahoo!知恵袋では、関係の有無自体が問題とされることが多く、結びつきに言及されることは比較的少ないからである。

## 5.2 NDC分類別の分析

次に、書籍サブコーパスに付与されているNDCの第1次区分ごとに集計した結果を表10と表11に示す。

表10 <構築・破壊>語彙のNDC分類別使用状況

ジャンル／語彙	構築	築く	修復	崩壊	破壊	壊れる	崩れる	樹立	ぎくしゃく	壊す	計
1 哲学	3	8	1	0	4	2	4	1	2	0	25
2 歴史	2	5	1	1	0	0	0	3	0	1	13
3 社会科学	37	73	5	5	3	1	7	7	9	1	148
4 自然科学	9	9	0	1	2	0	0	0	3	0	24
5 技術・工学	3	7	0	0	0	0	0	0	2	0	12
6 産業	7	2	0	2	0	1	0	0	0	1	13
7 芸術・美術	0	1	0	0	0	0	0	0	0	0	1
8 言語	0	4	0	0	0	0	0	0	0	1	5
9 文学	1	6	5	4	1	3	0	1	2	5	28
0 総記	0	0	0	0	0	0	0	1	0	0	1
分類なし	0	0	0	0	0	0	0	0	0	0	0
計	62	115	12	13	10	7	11	13	18	9	270

表10の<構築・破壊>語彙では、「関係を築く」が全体の42.6%（115/270）を占め、「構築する」が23.0%（62/270）でこれに次ぐ。NDCのジャンル別では、「3 社会科学」が半数以上（148/270=54.8%）を占めている。表1の母集団の推定比率でも、25.56%（生産実態）、18.69%（流通実態）と比較的高率ではあるが、それを大きく上回っている。社会とは、人と人との関係によって成り立つものであるから、これについての言説も、関係概念に言及することが多いものと考えられる。

表11の<結合・切断>語彙では、「関係を結ぶ」が最も多く（111/161=68.9%）、ジャンル別では、やはり「3 社会科学」が多いが3割程度（48/161=29.8%）であり、<構築・破壊>語彙と比較すると、それほど突出しているとはいえない。



表 11 <結合・切断>語彙の NDC 分類別使用状況

ジャンル／語彙	結ぶ	断(絶)つ	断ち切る	断絶	切れる	計
1 哲学	20	3	0	2	1	26
2 歴史	22	2	2	2	1	29
3 社会科学	28	4	4	10	2	48
4 自然科学	4	0	0	2	0	6
5 技術・工学	1	1	1	1	1	5
6 産業	3	0	0	0	0	3
7 芸術・美術	1	1	1	0	0	3
8 言語	4	1	0	0	0	5
9 文学	28	4	3	1	0	36
0 総記	0	0	0	0	0	0
分類なし	0	0	0	0	0	0
計	111	16	11	18	5	161

これらの表に、コレスポンデンス分析を行った結果が表 12 および図 3, 4 である。ただし、対象語彙の出現回数が 1 回以下の「総記」「分類なし」の行は分析から除外している。

表 12 固有値と寄与率 (NDC 分類別分析)

<構築・破壊>語彙	次元 1	次元 2	次元 3	<結合・切断>語彙	次元 1	次元 2	次元 3
固有値	0.2408	0.1171	0.0777	固有値	0.1048	0.0696	0.0242
寄与率(%)	46.31	22.52	14.95	寄与率(%)	50.23	33.38	11.6
累積寄与率(%)	46.31	68.83	83.78	累積寄与率(%)	50.23	83.61	95.21

次元数は、 $(k = \min\{\text{クロス集計表の行数}, \text{クロス集計表の列数}\}) - 1$  個求められるため、<構築・破壊>語彙では 8 次元、<結合・切断>語彙では 4 次元求められているが、表 6 には次元 3 までを表示している。次元 2 までに、<構築・破壊>語彙では 68.83%、<結合・切断>語彙では 83.61% の連関が説明されている。一般に、次元数が増えるほど、寄与率は分散される傾向があり、<構築・破壊>語彙の第 2 次元までの累積寄与率は、他と比べてやや低い値となっている。

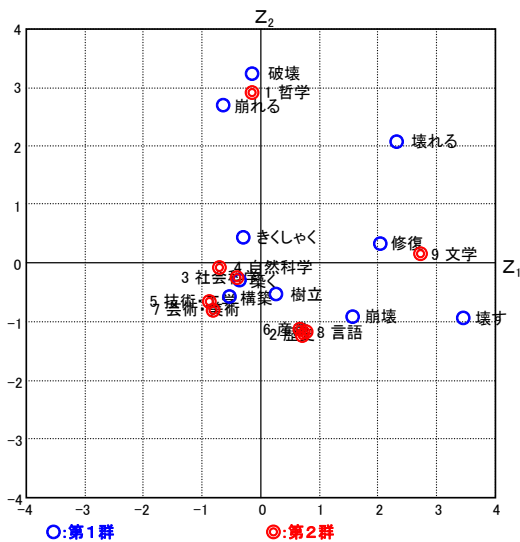


図 3 NDC 分類別の散布図  
(<構築・破壊>語彙)

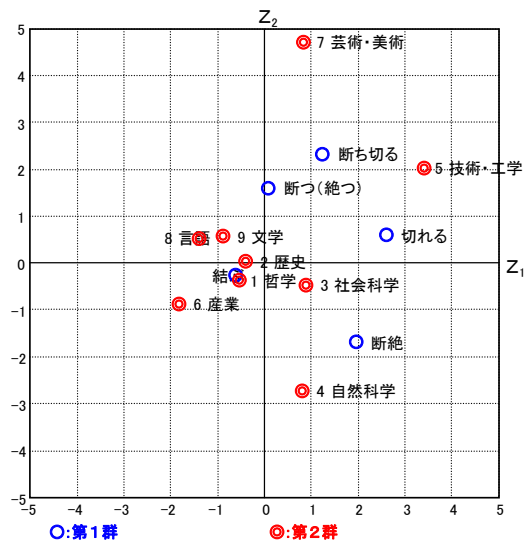


図 4 NDC 分類別の散布図  
(<結合・切断>語彙)

図3では、第1次元（横軸）正方向に文学が位置し、「壊す」「壊れる」という語彙によって特徴付けられている。この2語は、前節で述べたYahoo!知恵袋と共通するものであり、友人関係や家族関係等の個人的な関係に用いられている。「修復」という語も第1次元のスコアが高いが、白書や国会会議録では、外交関係にのみ用いられていたのに対し、文学作品では主に夫婦関係の修復という個人的関係に用いられている。また、第2次元（縦軸）正方向には哲学が位置し、「破壊する」「崩れる」という語彙によって特徴づけられている。これらは社会科学でも用いられており、社会的関係をやや専門的に論述する際に用いられる語である。

一方、図4では、第1次元（横軸）正方向に技術・工学がプロットされ、第2次元（縦軸）正方向には芸術・美術がプロットされている。それぞれの次元のスコアが高い語は、「切れる」と「断ち切る」である。しかし、これらのジャンルに＜結合・切断＞語彙が出現した回数は5回と3回だけであり、「切れる」と「断ち切る」の出現回数は1回ずつである。これらは、統計的に意味のある数値とは言えず、用例を増やして再検証する必要がある。

最後に、「修復」と「断絶」という2語の使用状況について触れておきたい<sup>3</sup>。全コーパス中で、「樹立」という語は、社会科学分野で3例が人間関係に用いられているほかは、38例（92.7%）が外交関係を中心とした組織間の関係に用いられている。これに対し、「壊れる」は国会会議録の1例を除く12例（92.3%）が、「壊す」は文学での1例を除く15例（93.8%）が、友人関係や男女関係といった私的な個人間の関係に用いられている。「樹立」は白書と国会会議録、および書籍の社会科学分野での使用が多いのに対し、「壊す」と「壊れる」はYahoo!知恵袋と書籍の文学での使用が多い。これらの語彙は、特定の種類の関係にのみ使用されることで、そのような関係が言及されることが多いジャンルで多用される傾向があり、前節のサブコーパス別の分析に貢献した語彙である。一方、「修復」と「断絶」は、2国間の外交関係にも、夫婦や親子など、家族間の関係にも用いられる語である。「修復」は21例、「断絶」は31例の使用例があるが、その内訳は、表13のようになっている。

表13 「修復」と「断絶」の使用状況

	「修復」	個人間関係	組織間関係	「断絶」	個人間関係	組織間関係
Yahoo!知恵袋	1	1	0	0	0	0
国会会議録	2	0	2	8	8	0
白書	6	0	6	5	0	5
書籍	12	4	8	18	7	11
計	21	5	16	31	15	16

「修復」が個人的関係に用いられているのは、Yahoo!知恵袋と書籍の文学であり、これは、「壊す」や「壊れる」の場合と一致する。しかし、国会会議録と白書や、書籍の社会科学では、外交関係に用いられており、これは「樹立」と同じパターンである。また、「断絶」は、白書や書籍の社会科学分野では外交関係に用いられているが、国会会議録では参議院法務委員会で、法的な親族関係に用いられている。このように、言及される関係の種類はサブコーパスやジャンルによって比較的はっきり分かれているが、語彙が両方の種類の関係に用いられているために、ジャンルの特徴を表すためには貢献しなかったのである。これらの漢語語彙は、公的な分野から徐々に私的な分野へと使用の裾野を広げている過程にあると考えられる。このように、関係のメタファーを表す言語表現は、言語項目を話し相手や場面、テキストジャンル等に応じてカテゴリカルに切り換える単純なコード・スイッチング (code-switching) としては捉えられない複雑な様相を呈している。

## 6 おわりに

本稿では、＜関係＞に関する日本語のメタファー表現のなかで、特に多くの用例が存在する＜構築・破壊＞を表す語彙と、＜結合・切断＞に関わる語彙を対象とし、テキストジャンル別の使用状況を調査し、コレスポネンス分析によって視覚化した。その結果、関係のメタファーを実現する語彙には、「関係を築く」や「関係を結ぶ」のように、レジスターやジャンルに関わりなく広く用

いられるものと、「関係を樹立する」や「関係が壊れる」のように、ジャンルごとに使い分けられるものがあることが明らかになった。

＜構築・破壊＞語彙に関する分析では、国会会議録、行政白書という公的な場では、「樹立」「構築」「修復」といった関係構築を表す漢語サ変動詞が特徴的に用いられていること、一方、書籍、Yahoo!知恵袋という私的なジャンルは、「壊す」「壊れる」「崩れる」という関係破壊を表す和語によって特徴づけられるということが判明した。これは、千年に及ぶ日本語表記の伝統が生きているということに加えて、それぞれのジャンルにおいて言及される関係の種類（個人間の関係か、国家をはじめとする組織間の関係か）や、関係構築から崩壊に至るどの過程に注目するかが異なっていることに原因が求められる。

＜結合・切断＞語彙でも同様に、公私の区別がされているが、これは「断つ（絶つ）」「断ち切る」と「切れる」が対照的な用いられ方をしていることによる。国会という公的な場でも、親族関係のような私的な関係は「切れる」という和語によって表現されるからである。

さらに、「断絶」や「修復」は、行政白書や国会外交委員会および社会科学関係書籍では外交関係をはじめとする組織間関係に、文学作品や国会法務委員会では個人的な人間関係や親族関係に用いられるという明確な傾向があった。これらの語彙は、公的な用法と私的な用法を両立させている。以上のように、言語が使用される社会的状況によって、述べられる関係の種類は拘束されているが、すべての語彙が関係の種類ごとに使い分けられているわけではないということも明らかになった。

周知のとおり、古くは和語と和文体によった日本語は、より抽象的論理的表現を可能とした外来の漢語と漢文体を獲得し、それとの融合を図ることで、多様な分析的表現を可能としてきた（安部 2009）。重要なことは、漢語が大量に輸入されたのは、7世紀から8世紀にかけての仏教と律令制度の伝来によるということである。仏教は国家イデオロギーであり、律令は刑法と行政のシステムである。この重要なシステムを、われわれの祖先は中国大陸から漢字とともに借用したのである（井上 2011）。したがって、政治や行政の分野で漢語が用いられるのは、むしろ当然なのである。これらの漢語は、当初は僧や役人など特定の職種の間だけで使っていたであろうが、徐々に一般社会にも浸透していったと考えられる。「有頂天」「四苦八苦」のように、日常語化した仏教用語は数多いが、「断絶」「修復」なども同じような日常語化の経過をたどったものと推定される。

従来、認知メタファー理論では、理想認知モデルや認知レベルの写像が重視され、メタファー表現の使い分けも、発話者の認知機能として内面化されたものとして考えられる傾向があった<sup>4</sup>。しかし、特定の語彙が特定の分野で使われる背景には、上に述べたような歴史的背景が大きくかかわっている。最終的には語彙使用の慣習が個人の心的システムに内面化されるとしても、具体的な使い分けの説明をしようとするならば、社会的なシステムや歴史的な経緯に言及せざるを得ない。社会というものは、そこで生きている人間を、あるしかたで行動せざるを得ないように仕向けるものである。言語使用を含む、社会で起きているさまざまな現象を理解しようとするときに、個々の人間の内的なシステムだけを考えると、社会全体で起きていることの本質が見落とされる場合がある。山岸（2010, 2011）は、社会で起こっているさまざまな出来事を、一人ひとりの心が生み出していると考える直観的な理解を、「頭でっかち」という言葉になぞらえて「心でっかち」な考え方と呼んで批判している。「心でっかち」とは、いろいろな社会現象を理解するために、とりあえずすべてを心の問題として置き換えることから出発して、なぜそんな心の問題が生まれたのかを考えるやり方である。

文化は、われわれの行動によって生み出され支えられているものであるが、われわれ一人ひとりは、周囲の人々の行動によって作り出され維持されている社会的環境のなかで行動している。その行動がさらに社会的環境自体を作り出すという循環的な構造のなかで、われわれは生きている。本稿が発見したテキストジャンルによるメタファー表現の使用分布は、ヒトの心的システムだけで説明できるものではなく、心的な思考システムと社会的なシステムが構造的にカップリングした「複合システム」としてメタファー産出機構を捉えなおさなければならないことを示唆している。

## 注

1. 「ぎくしゃくする」という語は建築用語ではないという御指摘を、神戸大学名誉教授 西光義弘先生からいただいた。語源的には、物と物がきしむ時の擬音語と考えられ、動きのぎこちなさを表すこと

が多いと思われるが、複数の部品が組み合わされた関節部分を持つ構築物を前提としていることから、ここに含めている。また、「構築する」「樹立する」という語は、物理的な領域ではほとんど用いられることはなく、メタファー専用の語彙とってよいものである。

2. この制約は、「関係」という語と共起するという条件を付した場合にのみあてはまる。「つながり」「結びつき」「結束」などが、関係概念を表すことは当然である。

3. 日本認知言語学会第12回全国大会での発表時は、ジャンルによる語彙の使い分けを前提としていたために、関係の種類とジャンルの関係についての考察が不十分であった。当日、この点に関するご指摘をいただいた、神戸大学教授 松本曜先生に深く感謝申し上げる。

4. 日本認知言語学会第12回全国大会での発表時に、関西大学教授 鍋島弘治朗先生がこのような考え方を表明された。発話者は、異なる種類の関係の身体的な感覚の違いを表すために、異なる表現を選ぶというものである。このような考え方は、誤りではないが「心でっかち」なものである(本文参照)。

## 参考文献

- 安部清哉 (2009) 「第3章 意味から見た語彙史」, 安部清哉他著『シリーズ日本語史2 語彙史』, 東京: 岩波書店, 73-104.
- 石川慎一郎・前田忠彦・山崎誠 (編) (2010) 『言語研究のための統計入門』, 東京: くろしお出版.
- 井上ひさし (2011) 『日本語教室』, 東京: 新潮社.
- 内田治 (2006) 『すぐわかるSPSSによるアンケートのコーレスポネンダ分析』, 東京: 東京図書.
- 内田諭・藤井聖子 (2009) 『『日本語コーパス』における語彙のジャンル別特徴: クラスタ分析とフレームの観点から』, 『言語処理学会第15回年次大会発表論文集』, 442-445.
- 金水敏・乾善彦・渋谷勝己 (2008) 『シリーズ日本語史4 日本語史のインタフェース』, 東京: 岩波書店.
- 小磯花絵・小木曾智信・小椋秀樹・宮内佐夜香 (2009) 「コーパスに基づく多様なジャンルの文体比較: 短単位情報に注目して」, 『言語処理学会第15回年次大会発表論文集』, 594-597.
- 国立国語研究所の言語コーパス整備計画 KOTONOHA モニター公開データの内容, Online at [http://www.ninjal.ac.jp/kotonoha/ex\\_8.html](http://www.ninjal.ac.jp/kotonoha/ex_8.html). 大学共同利用機関法人人間文化研究機構国立国語研究所.
- 高橋信 (2005) 『Excelで学ぶコーレスポネンダ分析』, 東京: オーム社.
- 鍋島弘治朗 (2005) 「認知メタファー理論に対する一提案: 関係のメタファーを例に」, 『Proceedings of the 30th Anniversary Meeting』, 関西言語学会, 282-291.
- 鍋島弘治朗 (2011) 『日本語のメタファー』, 東京: くろしお出版.
- 半沢幹一 (2009) 「第5章 文体・位相から見た語彙史」, 安部清哉他著『シリーズ日本語史2 語彙史』, 東京: 岩波書店, 127-166.
- 丸山岳彦・柏野和佳子 (2011) 『『現代日本語書き言葉均衡コーパス』におけるサンプリングの設計と実施』, 『『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集』, 21-26.
- 水野真紀子・内田諭・アニタナジ・大堀壽夫 (2006) 「人間関係のメタファーにおけるスキーマ類型」, 『日本認知言語学会論文集』, 第7巻, 120-130.
- 村田年 (1999) 「接続語句・助詞相当句による文章の所属ジャンルの判別—多変量解析法を用いて—」, 『言語処理学会第5回年次大会予稿集』, 213-216.
- 村田年 (2000) 「多変量解析による文章の所属ジャンルの判別—論理展開を支える接続語句・助詞相当句を指標として—」, 『統計数理』, 48 (2), 311-326.
- 山岸俊男 (2010) 『心でっかちな日本人—集団主義文化という幻想』, 東京: 筑摩書房.
- 山岸俊男 (2011) 『「しがらみ」を科学する—高校生からの社会心理学入門』, 東京: 筑摩書房.
- 山崎誠 (2006) 「代表性を有する現代日本語書き言葉コーパスの設計」, 『言語コーパスの構築と活用』 (第13回国立国語研究所国際シンポジウム報告書), 63-70.
- 山崎誠 (2011) 『『現代日本語書き言葉均衡コーパス』の構築と活用』, 『『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集』, 11-20.
- 山崎誠・前川喜久雄・田中牧郎・小椋秀樹・柏野和佳子・小磯花絵・間淵洋子・丸山岳彦・山口昌也・秋元祐哉・稲益佐知子・吉田谷幸宏 (2006) 「代表性を有する現代日本語書き言葉コーパスの設計」, 『言語処理学会第12回大会発表論文集』, 440-443.
- Charteris-Black, J. (2000) Metaphor and vocabulary teaching in ESP economics. *English for Specific Purposes*, 19, 149-165.
- Holmes, J. (2008) *An Introduction to Sociolinguistics*, 3rd Edition. Longman: London.
- Skorczynska, H. and Deignan, A. (2006) Readership and purpose in the choice of economics metaphor. *Metaphor and Symbol*, 21(2), 87-104.

<abstract>

## Correspondence Analysis of Relationship between Text Genres and Linguistic Metaphors: The Case of 'Relation' Metaphor

**Akira OISHI**

Meisei University

This paper investigates the relationship between text genres and linguistic metaphors concerning 'relation'. Although the notion of 'relation' is very popular and commonplace, it is also abstract and is expressed by various metaphorical expressions. Among others, we focus on the verbs expressing 'construction or destruction' and 'linking or cutting', both of which occur in the corpus many times. We counted up the frequencies of the verbs in each genre of text and compiled them into cross-tabulation tables. Then, we applied the correspondence analysis on the tables to visualize the correlation between text genres and linguistic metaphors.

The analysis revealed the fact that there exist two types of the verbs: one that are widely used irrespective of the text genres; for example, "*kizuku* (build)" and "*musubu* (connect)", and the other, are those which are used only in a few genres of text; "*juritū* (establish)" for the Diet records and white papers, "*kowasu* (destroy)" for electronic BBS on the Internet and novels, and so on.

In the public sectors such as the Diet records and white papers, the verbs of Chinese origin expressing construction are major, while the private sectors such as electronic BBS on the Internet and novels are characterized by the native Japanese verbs expressing destruction. This is a reflection of the Japanese writing system lasting over one thousand years.

The social situation restricts the relations to be referred to, and consequently the verbs metaphorically expressing the relations in a specific genre are constrained. The findings we obtained suggest the need for reconsidering the metaphor production as a complex system which structurally coupling the human mind system and the social system.