上越教育大学研究紀要　第40巻第２号　令和３年３月
Bull. Joetsu Univ. Educ., Vol. 40(2), Mar. 2021

日本語母語話者／学習者の文章に特徴的な語彙ネットワーク構造

鯨　井　綾　希*

（令和２年８月31日受付；令和２年11月30日受理）

要　　旨

本研究では，母語話者の文章と学習者の文章が対比され，そこから分析される日本語の文章上の特徴を考察した。文章中に現れる語彙を同一文内の共起に基づく語彙ネットワークによって整理し，その特徴を探索的に分析した。分析の結果から以下のことが明らかになった。母語話者の文章は，課題ごとに独立した語彙を用いる傾向にある。また，母語話者は中心的な語を強く関連付けることに重点を置いた文章形成を構想している。さらに，母語話者は，局所的な結束性よりも大きな文章全体の中での位置づけを考慮して文章を作成している可能性がある。学習者の文章は，複数の課題ごとに共通する少数の語彙的まとまりを用いる傾向にある。また，学習者は，ある語から別な語への単発的なネットワークの形成を優先的に行い，特定かつ少数の語と別な語を結び付けることによって内容的連結を形成している。さらに，学習者は，母語話者に比べて局所的な結束性の構築に重点を置いた文章を作成している可能性がある。

KEY WORDS
Quantitative lexicology 計量語彙論，Collocation コロケーション，Text mining テキストマイニング，Written Corpus of Yokohama National University YNU書き言葉コーパス，Graph description グラフ説明

1　はじめに

現代では，国際的な交流の活発化に伴い，母国語や自らの出自に依拠する母語の習得のみならず，他言語の学習も様々な地域で行われるようになってきている。それによって，日本語という言語を取り巻く環境でも，日本語を母語とする者の文章のみならず，日本語を外国語として学んだ者（日本語学習者）による文章や，外国語を翻訳することで作られた日本語の文章などを目にする機会が増えている。また，たとえばある日本語の文章が「外国語的な文章である」といった印象を母語話者に抱かせることがあるように，日本語母語話者の文章とそれ以外の文章は，多くの場合，印象のレベルにおいて何らかの違いを想起させる。しかしながら，たとえ母語話者であっても，日本語と異なる言語をその背景として持つ日本語の文章と，母語話者によって書かれた日本語の文章との構造的な違いを明瞭に説明することは困難である。これはつまり，母語話者自身が自らの言語を客観的・体系的な姿として十分に捉えきれていないことを意味する。

このことから，日本語母語話者と学習者に見られる言語表現の差異を明らかにすることは，学習者による日本語理解という面だけでなく，母語話者による日本語理解という面でも重要であると言える。また，母語話者と学習者のそれぞれの日本語を対照することで，母語話者が書いた日本語の分析だけでは見出せない日本語の言語的特徴が浮かび上がる可能性もある。実際，近年では，学習者の日本語を日本語という言語の体系に取り入れることで，日本語そのものの特徴をより多面的に見出していくことが様々に試みられている（阿部・庵・佐藤（編）2015，中俣2016，鯨井2017，森（編）2018，野田・迫田（編）2019など）。

本稿では，日本語の母語話者と学習者の双方が同一条件のもとで作成した文章を電子化した『YNU書き言葉コーパス』という資料を利用して，母語話者の文章を学習者のそれと対比させ，日本語の文章の書き方の多様性と特徴について考察する。また，文章を構成している一要素である語彙を利用して，文章という複雑な構造を持つ単位の特徴を定量的に分析していき，具体的な研究手法を示すことで，文章という大きな単位に有効な分析方法を提示することも目標とする。

*人文・社会教育学系

2 先行研究と本稿の分析視座

本稿では，日本語母語話者および学習者の文章上の特徴を調査するにあたり，計量分析に主眼を置く。本節では，文章を対象とした定量的な方法に基づく先行研究を取り上げ，その方法論的な有効性と問題点を指摘することで，本稿の分析視座を明示する。

文章を対象とした計量分析は，近年「テキストマイニング」という専門用語によって括られる枠組みの中でしばしば見受けられる。石田・金（編著）（2012）では，「コーパスとテキストマイニング」というフレーズのもとで，言語学や心理学のみならず，金融や政治に至るまでの幅広い分野での電子化テキストを用いた計量的な研究事例が掲載されている。また，岸江・田畑（編）（2014）でもテキストマイニングという語が使われており，それをキーワードとしつつ，そこから広がる言語研究の可能性が示されている。いずれの研究でも，近年分析資料として定着を見た電子化テキストを利用し，その利点である大きなデータを用いた計量研究が行われている。さらには，語や語同士のコロケーションの頻度だけでなく，それらを利用した多変量解析やネットワーク分析なども行われており，語や語彙を利用した文章の多角的な計量分析法が提示されている。

ただ，テキストマイニングという用語のもとに行われる言語分析は，総じて問題意識が限定的であり，ある方法を用いた時点で，その分析対象や分析結果の方向性も概ね決まってしまっているという問題がある。たとえば，多変量解析は文章ジャンルや執筆者の分類を行うために用いられ，語彙を用いたネットワークの分析は特定の文章の内容整理に用いられる，という具合である。現状のテキストマイニングによって行われている諸研究は，そこで用いられる分析方法とそれによって解明できることの結びつきを必要以上に強めてしまい，言語研究に関わる方法論的な多様性を十分に示すことができていない。

ところで，石井（2007，2019）では，記述統計学の一種である「探索的データ解析」という概念を言語研究の中に取り入れることが試みられている。また，計量分析を通した言語学的な仮説の発見という側面にも力が入れられている。石井（2007，2019）が示しているように，言語研究における定量的データの分布調査やその可視化に際しては，それによる発見的・仮説的側面との関わりを重視する必要がある。テキストマイニングの手法はもとより定量的であるが，先述したように，少なくとも言語分析を目的とした研究に限って言えば，その方法と問題意識は必ずしも言語そのものの発見的・仮説的側面に重点を置いたものになっていない。言語研究における方法論的な広がりを示すためには，石井（2007，2019）が考えるような探索的な分析視座を積極的に導入していくことが求められる。

また，従来の方法論的手続きに縛られない多様な応用例を示すことは，将来的にテキストマイニングという方法論を言語研究上の手法として定着させることにつながると考えられる。現時点でも，そういった事例が全く見られないというわけではない。たとえば，河瀬・市村・小木曽（2014）における『虎明本狂言集』の会話文の計量分析は，登場人物の身分・役割（商人・百姓・漁師など）ごとの語彙の運用上の差異を，語彙ネットワークの違いを通して探索的に明らかにしており，近年の発見的・仮説的計量分析の一例として注目に値する。小林（2019）は，単語と単語の共起関係を可視化する一例として，Web上に現れる「ブラックサンダー」という商品名を中心とする語彙ネットワークを作成し，「ブラックサンダー」が「バレンタイン」「欲しい」「あげる」といった語と結びつきやすかったことを報告するとともに，その言語学的発展の可能性について触れている。本稿でも同様に，文章中に現れる語彙の現れ方を定量的に捉えるとともに，その特徴を可視化し，探索的に分析していくことで，質的な研究では見出しにくい日本語の特徴を発見していくことを試みる。

3 分析資料と分析方法

本稿では，金澤（編）（2014）の付属CD-ROMに収録されている『YNU書き言葉コーパス』を資料として用いた。『YNU書き言葉コーパス』は横浜国立大学（Yokohama National University）でのプロジェクトに基づいて作成されたコーパスであり，金澤（編）（2014）によれば，「大学生の日常における「書く」という言語活動に注目し，日本人大学生（30名）と同大学に所属する留学生（韓国語母語話者30名，中国語母語話者30名）を対象に，12の課題による書き言葉の資料，計1080編（母語別各グループ360編ずつ）を集めたものである」とされる（金澤（編）2014：3）。

『YNU書き言葉コーパス』は，日本語母語話者と学習者の日本語の使用に関わる分析を容易にするために，統一的な基準に基づいて資料の収集が行われている。特に，母語話者・学習者が同一の課題に対して文章を作っている点が，本稿において重要である。その設定により，特定課題に対する母語話者・学習者の日本語表現法の違いを直接に対照できるためである。本稿では上記の利点を考慮し，『YNU書き言葉コーパス』を分析資料として選んだ。

また，『YNU書き言葉コーパス』では学習者の日本語能力を「上位群」「中位群」「下位群」の三つに分けている。しかし，「中位群」「下位群」は，文法上・語法上の誤りが目立つため，本稿で分析する文章という単位での比較は，適当ではない。したがって，本稿では母語話者と対照する学習者の資料として，文法や語法上での誤りがあまり見られなくなる「上位群」の20名を選択した。

『YNU書き言葉コーパス』で母語話者・学習者に課せられた課題は全部で12種類ある。各課題は，「手紙，PCメール，ケータイメール，投書，レポートなど，さまざまなスタイルのものとなるように考慮」されている（金澤（編）2014：8）。以下の表1に課題の一覧を，表2にその課題に関する母語話者・学習者（「上位群」）の語彙的概要を示す。

表1：『YNU書き言葉コーパス』で課される課題（タスク）

タスク1	面識のない先生に図書を借りる
タスク2	友人に図書を借りる
タスク3	デジカメの販売台数に関するグラフを説明する
タスク4	学長に奨学金増額の必要性を訴える
タスク5	入院中の後輩に励ましの手紙を書く
タスク6	市民病院の閉鎖について投書する
タスク7	ゼミの先生に観光スポット・名物を紹介する
タスク8	先輩に起こった出来事を友人に教える
タスク9	広報誌で国の料理を紹介する
タスク10	先生に早期英語教育についての意見を述べる
タスク11	友人に早期英語教育についての意見を述べる
タスク12	小学生新聞で七夕の物語を紹介する

（金澤（編）（2014）：53）

表2：母語話者・学習者の課題ごとの延べ語数と語彙多様度（Guiraud値）

	母語話者(30名)		学習者(20名)	
	延べ語数	Guiraud値	延べ語数	Guiraud値
タスク1	2788	9.13	2194	10.12
タスク2	1416	8.77	1686	11.01
タスク3	2252	10.41	1646	9.05
タスク4	4225	14.42	3963	16.23
タスク5	8151	17.67	6021	17.54
タスク6	5356	15.62	4058	17.20
タスク7	4269	18.93	3828	17.88
タスク8	2025	8.96	1649	10.86
タスク9	5727	18.10	5405	18.83
タスク10	4042	15.02	3448	14.85
タスク11	3424	15.11	2807	14.42
タスク12	11701	12.63	8006	13.60
合計・平均	55376	13.73	44711	14.30

表2の「Guiraud値」は，語彙の多様性を示す指標であるType-Token Ratioをもとに，文章の長さの影響をできるだけ小さくした補正値で，以下の式によって表される。

$$\text{Guiraud値} = \frac{V}{\sqrt{N}}$$

Vは異なり語数，Nは延べ語数を表す。Guiraud値が大きければより多くの語で文章を構成していると言え，小さければより少ない語で文章を構成していると言える。

表2を見ると，少なくとも上位学習者に限ってみれば，母語話者と同等かそれ以上の語彙多様性を示していることが分かる。したがって，本稿の分析に際しては，母語話者と学習者が持つそもそもの語彙量の差はひとまず考慮しないことにする。

なお，『YNU書き言葉コーパス』には，書かれた文章をできる限りそのまま文字化した「オリジナルデータ」と，漢字の誤りや送り仮名をはじめとした表記上の問題を適宜修正し，一行一文の形に加工した「補正データ」の二種類が存在するが，本稿は表記上の差異や改行位置の差異を問題としないため，表記上の均一化が図られている「補正データ」を利用した。

『YNU書き言葉コーパス』を本発表の資料として使うにあたっては，中・長単位解析器Comainuを用いてテキス

トデータを長単位と呼ばれる長い単位の語に分割した。形態論情報付きの分割結果は，目視で確認して誤解析部分の修正を行った。

分析方法としては，語彙内の各語の文章中における関連付けの仕方を見るために，文章中の同一文内で用いられた語同士をつなぎ合わせ，それによって生まれる語彙ネットワークを分析することとした。上記の分析を通して，日本語母語話者・学習者の文章がどのような情報構造下で構築されているのかを明らかにしていく。

4　同一文内の共起関係から見た母語話者・学習者語彙ネットワークの特徴

4. 1　語彙ネットワークの分析に関する概要

語の共起に基づく語彙ネットワークは，テキストマイニングの手法の一つであるネットワーク分析を応用したものである。ここでのネットワークとは，ある語とある語が何らかの範囲で共起して用いられたとき，両者に関連があると考え，それらを線分で結んだ結果出来上がる関係図を指す。たとえば，(1) のような文章があり，一文内で共起した各語に関係性があるとみなしたとき，語彙ネットワークは図１のような形で作られることになる。

(1) 太郎は花屋に行った。太郎はバラを買った。太郎は花子にバラをあげた。

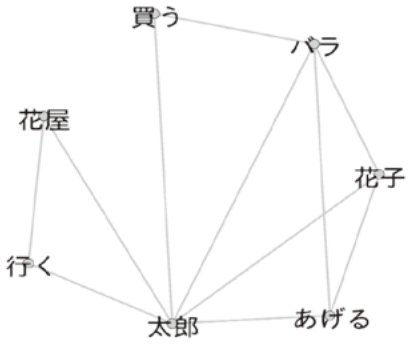


図１：用例（1）に基づく語の共起ネットワーク図

図１に示した語彙ネットワークでは，(1) で登場した各語が，それぞれの文内での共起に合わせて線分で結ばれている。具体的には，「太郎」「花屋」「行く」という結びつきと，「太郎」「バラ」」「買う」という結びつきと，「太郎」「花子」「バラ」「あげる」という結びつきの三つが，語彙ネットワークに反映されていると言える。このうち，「太郎」と「バラ」の二つの語は複数の文に亘って登場しているため，図１において，個々の語を線分でつなぐ役目を担っていることが分かる。特に「太郎」は，(1) の文脈において，全ての語をつなげる重要な語であると言うことができる。このように，語彙ネットワークの構築は，文章中に現れる各語が相互にどのように関連付けられるのかを可視化することを意味し，文章と語彙との相互関係の分析に有効な手法である。

本稿で構築した語彙ネットワークは，視覚的な分析のしやすさを考慮し，そもそもの語の頻度が母語話者で10以上，学習者で８以上の語を対象とした。母語話者と学習者の収集頻度の差は，それぞれの延べ語数の差に基づき，便宜的に設定した。

また，語彙ネットワーク構築に際してはプログラミング言語Pythonによる自作プログラムで共起頻度の重み付き行列を作り，統計解析用言語ＲでGML形式という表記形式に変換した後，グラフの可視化ソフトウェアGephi（ver.0.9.1）のOpenOrd（＋Noverlap）と呼ばれる描画法によってネットワークを描画した。

4. 2　『YNU書き言葉コーパス』全体における語彙ネットワーク

はじめに，対象資料全体における母語話者と学習者の語彙ネットワークを以下の図２に示す。構築されたネットワークは非常に複雑であり，この段階では個々の語の関連性を見るには至らない。ただ，大局的に見ても両ネットワークからいくつかのことが分かる。

まず，ネットワークの中心になっている円形のまとまりから離れたかたまりが両者ともに見られる。続いて，それらの中心からの離れ方を見ると，母語話者（左）側のネットワークの方が分離した枝の数が多い。母語話者側は大きく三つから五つ程度，中心から離れた枝が見られる。一方の学習者（右）側は，大きな枝は二つか，せいぜい三つで

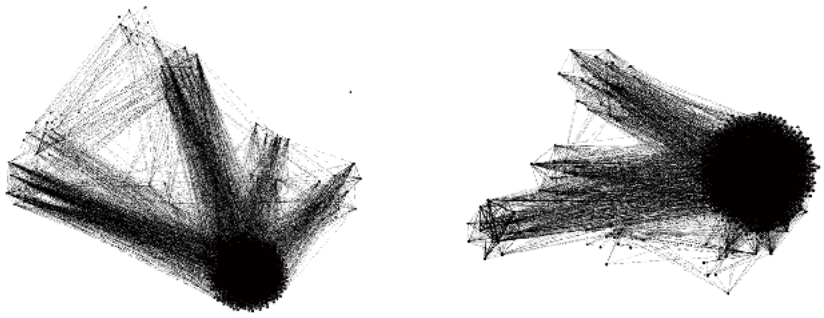


図２：日本語母語話者（左）と学習者（右）の語彙ネットワーク

ある。また，母語話者側の枝は個々の枝の重なり具合から見て分離が比較的明瞭であるが，学習者側の枝は，左下のものと左上のものの間に三つ目の枝が混ざりこんでおり，その境界線が母語話者側に比べ曖昧である。さらに，個々の分離した枝の先を見ると，母語話者側は枝の先でのつながりが薄くではあるが認められるのに対し，学習者側は分離した枝の先同士でのつながりが見られない点も指摘できる。

中心の円形のかたまりから分離された枝状のかたまりは，その他の文章とは異なる独立した語彙が使われているときに形成されるものである。本稿で対象とした課題は表１にも示したように12種類あるが，上記の図２より，母語話者は学習者に比べて課題ごとに独立した語彙を用いる傾向にあるものの，同時にそれらは完全に独立しているわけではなく，相互に弱い関係性を持たせつつ，課題ごとの文脈を構築していると考えられる。一方で，学習者は母語話者に比べて複数の課題ごとに共通した少数の語彙的まとまりを用いる傾向にあり，それらは課題相互にはあまり関与せず，それらの語彙的まとまりごとに文脈を構築していると考えられる。

上記の語彙ネットワークに対しては，全体の構造的特徴をさらに計量的に分析していくことも可能であるが，その点は今後の課題としたい。本稿では分析の容易さの問題もあり，次節では比較的ネットワークの規模が小さい母語話者・学習者の枝の先の部分を取り上げ，そこに見られる語彙ネットワーク構造の分析を行う。

4. 3　「グラフ説明」の文章群を事例とした語彙ネットワークの分析

図２で示した語彙ネットワークの末端を観察すると，母語話者・学習者ともにタスク３の「グラフ説明」とタスク９の「料理」が枝の先に位置づけられていた。このうち「料理」については，郷土料理の名前の登場や料理上の手順に伴う語の使用差などがあり，母語話者と学習者の直接的な比較が難しい。そのため，本稿では共通した内容になることが確実な「グラフ説明」における語彙ネットワークを取り上げ，分析を行うこととした。

上述したように，ここで取り上げるグラフ説明とは，『YNU書き言葉コーパス』中の「タスク３」に当たる。タスク３は，2004年から2010年にかけてのＡ社のデジタルカメラの販売台数の推移を説明する課題であり，その内容は以下のような形で提示される。

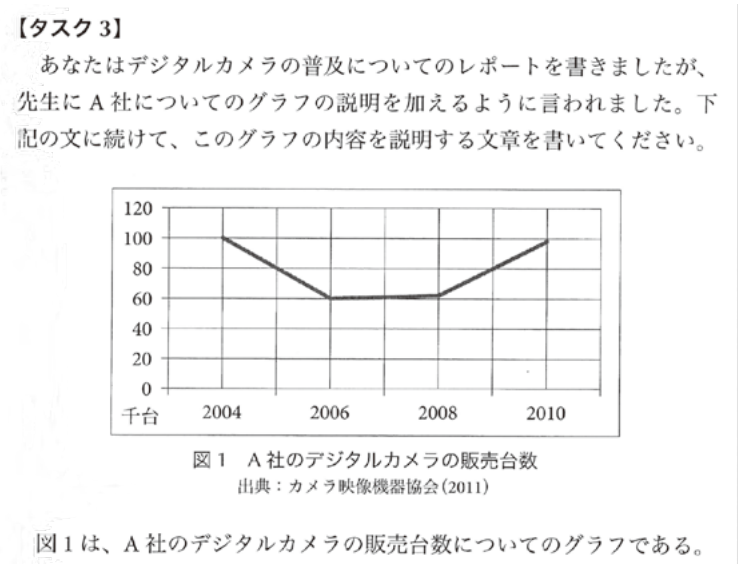
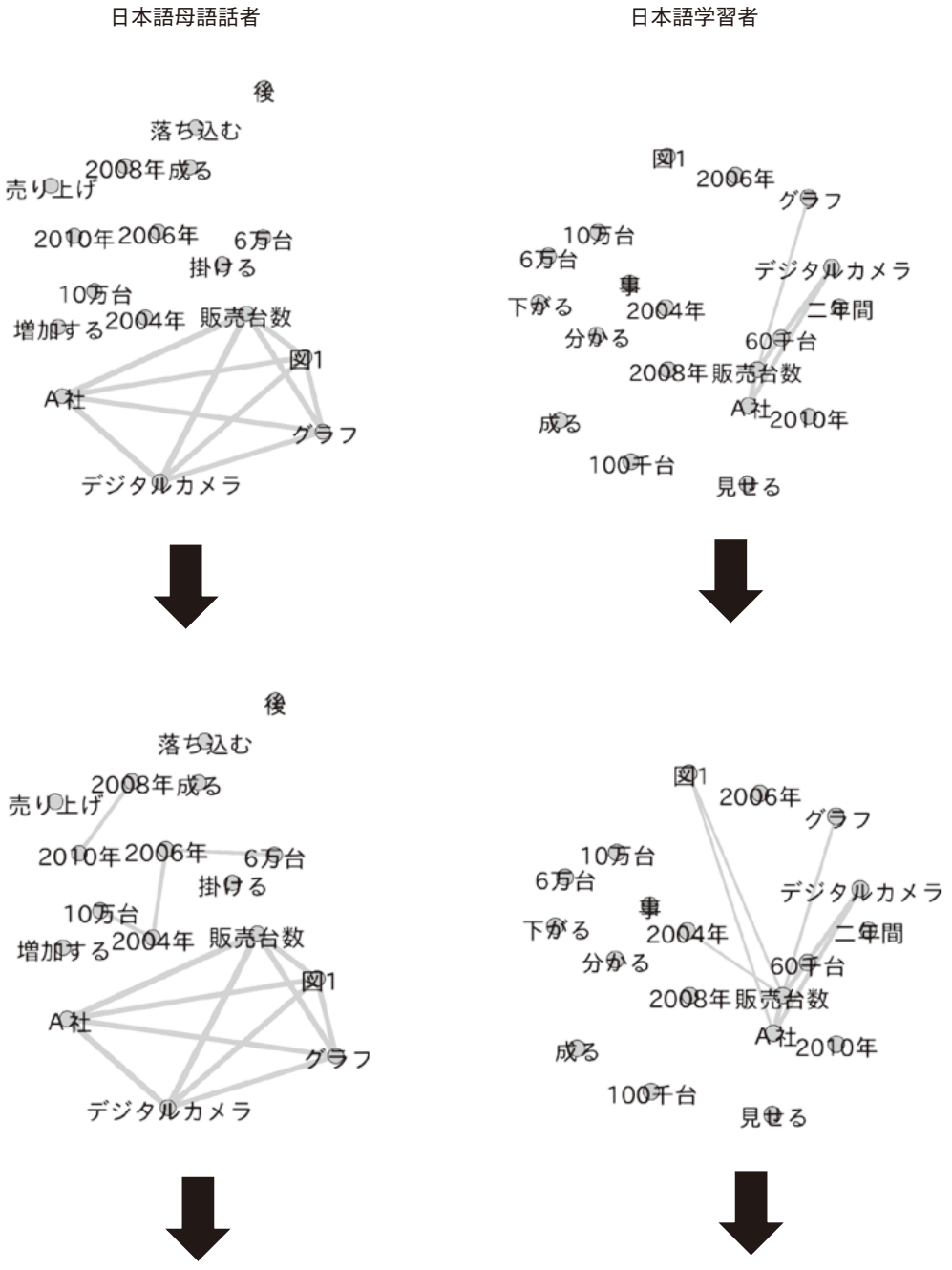


図３：金澤編（2014：87）からの引用

上記の課題に対応する形で書かれたグラフ説明の文章を取り上げ，その中で用いられた語彙をもとにして，語同士のつながりの強度が強い順（共起頻度の大きい順）に，段階的にネットワークを形成していった。すると，母語話者の文章と学習者の文章における語彙ネットワークの形成過程に違いが見られた。図４に，その変化過程を示す。なお，図中では，つながりの強度の強いものほど，太い線分で表示される。

図４では，左側に日本語母語話者の語彙ネットワークの遷移を表示し，右側に日本語学習者の語彙ネットワークの遷移を表示した。それぞれに，特徴的な変化を見せた際のネットワークを示し，上から下へ遷移の様子を示した。



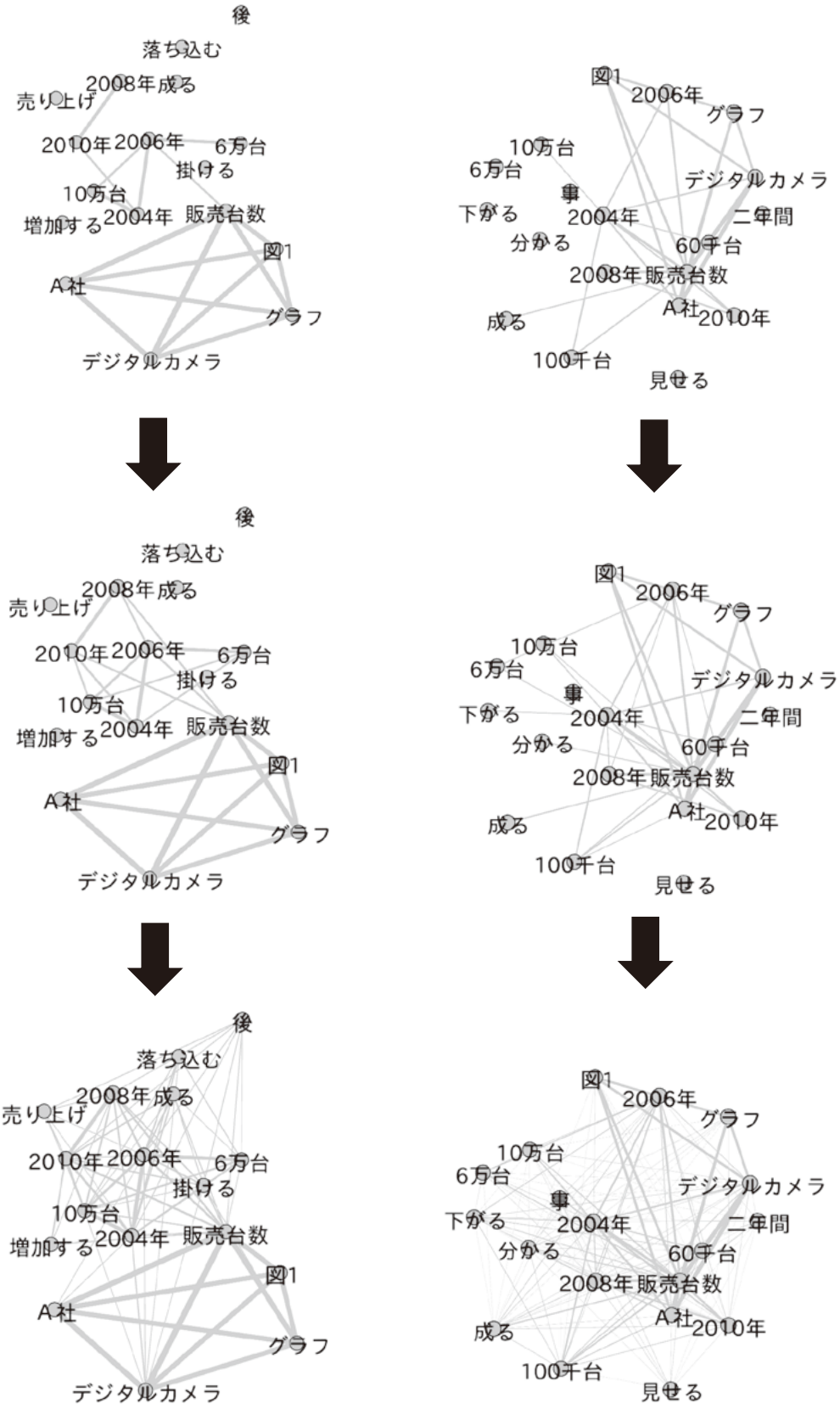


図 4 ：日本語母語話者と学習者の語彙における共起強度のネットワーク遷移

まず，日本語母語話者（左側）に注目して図４を見てみる。すると，グラフ説明を行うという文章において，日本語母語話者ははじめに「A社」「デジタルカメラ」「グラフ」「図１」「販売台数」という名詞群を相互に太い線分で，すなわち強い共起関係に基づくネットワークで結んでいる様子が見て取れる（図４左側の最初のネットワーク）。すなわち，日本語母語話者は，これら文章の主題に関わる語群をいくつかの文で独立に表示することをせず，一つの文の中で共起させて使うことを重視しているということを指摘できる。さらに，母語話者の語彙ネットワークにおいては，そのようにして中心的な語群を結びつけた後，それらとは別な箇所で新たなネットワーク関係を形成している（図４左側の２番目のネットワーク）。別な箇所で作られたネットワークは相互の関連性を強化しつつ，最初のネットワークとの関連付けを「販売台数」と関連付けることで，話題全体を結びつけている（図４左側の３番目・４番目のネットワーク）。最終的には，「販売台数」と「デジタルカメラ」「A社」を中心に様々な語群を結びつけ，語彙ネットワークの全体像を形成していく（図４左側の５番目のネットワーク）。

以上のことから，少なくとも「グラフの推移を説明する」という文脈では，日本語母語話者は，中心的な語をできるだけ同一文内で表示していると言える。すなわち，日本語母語話者は中心的な語を強く関連付けることに重点を置いた文章形成を構想していると考えられる。さらに，中心的な語同士の連結とは別な連結が別個に形成されていることから，中心的な語のネットワークが形成された後は，必ずしもそれらの語と直接的な連結を行わないまま，枝葉の話題についての詳細な説明を行っていると言える。それら独立した文内共起の関係は，最終的に「デジタルカメラ」や「販売台数」などの中心的な語と，それに関わる詳細な説明の語とが結びつき，それよって全体像を構築していく。

一方，学習者のネットワークにおいては，母語話者と異なり，中心的な語群の相互ネットワークが最初に形作られるわけでは必ずしもない。学習者の場合は，ある語から別な語への単発的なネットワークの形成が優先的に行われている（図４右側の１番目・２番目のネットワーク）。その傾向は全体を通して続くが，ある程度複雑性が増してくると，ネットワークは，特定の語，たとえば「販売台数」という語を基点として放射状の広がりを見せるようになる（３番目・４番目のネットワーク）。最終的に出来上がるネットワークを見ると，母語話者のものに比べて各語の結びつきが多く，細い線分によって複雑に結びつける形状となっていることが分かる（５番目のネットワーク）。

これらのことから，学習者は中心的な語の相互の関連付けを必ずしも優先していないことが分かる。むしろ，その時々で，特定かつ少数の語を基点に据えて，その語と別な語を随時結び付けることによって内容的連結を形成することに力点が置かれていると言える。

こうした差異は，日本語母語話者の文章が局所的な結束性よりも大きな，文章全体の中での位置づけが考慮された構造をなしているという可能性と，学習者の文章が母語話者とは異なり局所的な結束性の構築に重点が置かれた構造をなしている可能性という二つの可能性を示唆する。

ただし，図４は，個々の文章に見られる語の使用の時系列的な推移ではなく，特定の課題によって書かれた文章群における，文内共起の頻度の多寡に基づく変化を示したものである。そのため，ここで検討されていることは，あくまでもある文章群の情報構造を静的に整理したときの，個々の語の結びつきの強弱の推移，すなわち，より強い結びつきがどのようなもので，その強度の変化に応じてどのように語彙ネットワークが拡張していくかという観点である。これに対して文章の展開に応じたネットワーク拡張の分析もありえる。そうした動的観点からの分析は，日本語母語話者の日本語のみに注目した分析であるものの，鯨井（2020）において試みられている。

ここまでの，語彙ネットワークの構築とその部分的な形成過程を観察することで得た知見を以下に図示する。

表３：文章上で見られる語彙ネットワークの特徴

語彙ネットワークの特徴	母語話者	学習者
分岐形状	明瞭	曖昧
分岐数	多い（３～５）	少ない（２～３）
分岐先同士の関係	薄くだが有り	無し
形成過程（グラフ説明時）	中心語の相互連結	基軸語からの派生

表３のうち，分岐形状はジャンルの書き分け方につながり，分岐数はその書き分けの数につながる。分岐同士の関係は分岐形状と同様であり，解釈については今後さらなる検討が必要であるが，今回は中心的な書き方（円形の部分）から離れたもの同士にも語彙的なつながりが緩く存在しているかどうかという観点として理解したい。また，本稿におけるネットワークの形成過程は，グラフを説明するときに書き手がどのような情報上のつながりを重視して書いているかという点を説明するものである。

5　おわりに

本稿では，日本語の母語話者と学習者の双方が同一条件のもとで作成した文章を電子化した『YNU書き言葉コーパス』という資料を利用して，母語話者の文章を学習者のそれと対比させ，そこから分析される日本語の文章上の特徴について考察した。その際には，文章中に現れる語彙を同一文内の共起に基づく語彙ネットワークにより整理し，その特徴を探索的に分析していくことで，質的な研究では見出しにくい日本語の特徴を発見し，併せて文章という大きな単位に有効な分析方法を提示することを目標として設定した。分析の結果から明らかになった日本語母語話者と学習者のネットワーク構造の特徴と，それに基づいて本稿で述べた両者の日本語の文章上の特徴に関する解釈を整理し，以下に示す。

(2) 日本語母語話者と学習者に見られる共起関係に基づく語彙ネットワークの特徴と文章上の差異

・日本語母語話者

1．語彙ネットワークの分岐形状が明瞭であり，分岐数が多く，分岐先同士の関係が認められる。このことから，母語話者は課題ごとに独立した語彙を用いる傾向にあるが，同時にそれらは完全に独立しているわけではなく，相互に弱い関係性を持たせつつ，課題ごとの文脈を構築していると考えられる。

2．グラフ説明という特定の文脈内で語彙ネットワークの形成過程を観察すると，中心的な語同士の相互連結に重点を置いていることが分かる。このことから，母語話者は中心的な語を強く関連付けることに重点を置いた文章形成を構想していると考えられる。また，枝葉の話題に関する語彙は少数の中心的な語と結びつき，特定の話題と枝葉の話題が関連付くことによって，全体的な文脈を形成していると考えられる。

3．母語話者は，局所的な結束性よりも大きな文章全体の中での位置づけを考慮して文章を作成している可能性がある。

・日本語学習者

1．語彙ネットワークの分岐形状が曖昧であり，分岐数が少なく，分岐先同士の関係が認められない。このことから，学習者は複数の課題ごとに共通した少数の語彙的まとまりを用いる傾向にあり，それらは複数の課題相互にはあまり関与せず，それらの語彙的まとまりごとに文脈を構築していると考えられる。

2．グラフ説明という特定の文脈内で語彙ネットワークの形成過程を観察すると，学習者は，ある語から別な語への単発的なネットワークの形成を優先的に行うため，そのネットワークは特定の語を基点として放射状の広がりを見せることが分かる。母語話者のものに比べて各語の結びつきが多く，細い線分によって複雑に結びつける形状となっている。これらのことから，学習者は特定かつ少数の語と別な語を結び付けることによって文章の内容的連結を形成していると考えられる。

3．学習者は，母語話者に比べて局所的な結束性の構築に重点を置いた文章を作成している可能性がある。

以上に述べた本稿の分析結果は，文章中に現れる語彙の様相を定量化し，その値をもとに分析を行った結果，見えてきたものでもある。特にネットワーク分析を通した言語研究は，本稿の分析結果からも分かるように，文章のような大きな言語単位において有効性を発揮しうるものであると言える。もちろん，本稿からさらに分析を加えるべき部分は多々ある。たとえば，図2や図4で示したネットワークは，本稿ではその外観を質的・感覚的に捉えて分析したが，ネットワーク分析で用いられる種々の定量化によって量的にその特徴を捉え直し，より客観的かつ詳細なネットワーク構造の分析を行うことが可能である。また，図4のようなネットワーク形成の原因は，(2) で示したように，文章構築における結束性（cohesion）と首尾一貫性（coherence）の差異に求められる可能性があり，その点のより精確な検証が必要である。それら残された問題の考察は全て今後の課題としたい。

参考文献

① 阿部二郎・庵功雄・佐藤琢三（編）(2015)『文法・談話研究と日本語教育の接点』くろしお出版.

② 石井正彦（2007)「日本語研究における探索的データ解析の有用性」田野村忠温・服部匡・杉本武・石井正彦『特定領域研究「日本語コーパス」平成18年度研究成果報告書（JC-L-06-10） コーパスを用いた日本語研究の精密化と新しい研究領域・手法の開発Ⅰ』, pp.157-165.

③ 石井正彦（2019)『探索的コーパス言語学－データ主導の日本語研究・試論－』大阪大学出版会.

④ 石田基広・金明哲（編著）(2012)『コーパスとテキストマイニング』共立出版.

⑤ 金澤裕之（編）（2014）『日本語教育のためのタスク別書き言葉コーパス』ひつじ書房.
⑥ 河瀬彰宏・市村太郎・小木曽智信（2014）「『虎明本狂言集』における会話文の計量分析」『言語処理学会　第20回年次大会　発表論文集』，pp.662–665.
⑦ 岸江信介・田畑智司（編）（2014）『テキストマイニングによる言語研究』ひつじ書房.
⑧ 鯨井綾希（2017）「日本語母語話者／学習者の文章における語彙的差異－『YNU書き言葉コーパス』を用いて－」『国語学研究』56，pp.156–170.
⑨ 鯨井綾希（2020）「文章の内容展開に伴う語彙的結束性の形成過程－中学校教科書の「モアイは語る」を例に－」『国語学研究』59，pp.199–212.
⑩ 小林雄一郎（2019）『ことばのデータサイエンス』朝倉書店.
⑪ 野田尚史・迫田久美子（編）（2019）『学習者コーパスと日本語教育研究』くろしお出版.
⑫ 中俣尚己（2016）「学習者と母語話者の使用語彙の違い－『日中Skype会話コーパス』を用いて－」『日本語／日本語教育研究 7』ココ出版，pp.21–34.
⑬ 森篤嗣（編）（2018）『コーパスで学ぶ日本語学　日本語教育への応用』朝倉書店.

分析資料と分析ツール

『YNU書き言葉コーパス』（金澤（編）（2014）の付属CD-ROMに所収）.
「Comainu」：http://comainu.org　　2020年 8 月26日アクセス確認
「Gephi – The Open Graph Viz Platform」：https://gephi.org　　2020年 8 月26日アクセス確認
「Welcome to Python.org」：https://www.python.org　　2020年 8 月26日アクセス確認
「The R Project for Statistical Computing」：https://www.r-project.org　　2020年 8 月26日アクセス確認

付記

本研究は，2016年 7 月に行われた第396回国語学研究会での口頭発表をもとに，現在の研究動向を踏まえ加筆・修正したものである。また，本研究は「公益信託田島毓堂語彙研究基金」研究助成「日本語母語話者／学習者の語彙運用上の差異の解明－コーパスと計量分析を活用して－」（研究代表者：鯨井綾希，2015年度－2016年度）およびJSPS科研費JP19K13176「文章展開メカニズムの解明に向けた語彙拡張プロセスに関する研究」（研究代表者：鯨井綾希，2019年度－2021年度）による研究成果の一部である。

Lexical Network Structures Characteristic of Japanese Native Speakers' and Learners' Sentences

Ayaki KUJIRAI*

ABSTRACT

In this study, the sentences of native Japanese speakers were contrasted with those of learners, and the characteristics of Japanese sentences analyzed in the contrast were discussed. The lexical features of the sentences were exploratively analyzed by organizing the lexical networks based on co-occurrence within the same sentence. The results of the analysis revealed the following.

Native speakers' sentences tended to use independent vocabulary for each task. In addition, native speakers conceive of sentence formation with an emphasis on the strong association of central words. Furthermore, native speakers may consider their place in the larger whole sentence rather than local cohesion in their sentence formation.

Learners' sentences tended to use a small number of common lexical cohesion for each of the multiple tasks. Learners also prioritize the formation of singular networks from one word to another and form content linkages by linking a specific and few words to another. In addition, learners may be more focused on building local cohesion in their sentences than native speakers.

* Humanities and Social Studies Education