

コーパスにみる話しことばと書きことばの連続性
—BCCWJ と CSJ におけるチョット/スコシ、ヨ/ネの出現頻度の比較を通じて—

秋本 瞳

キーワード： 日本語、話しことば、書きことば、出現頻度、コーパス

要旨

音声言語と話しことば、文字言語と書きことばとのずれを示すために、現代日本語書きことば均衡コーパス(BCCWJ)、及び日本語話し言葉コーパス(CSJ)を用い、チョット/スコシ、ヨ/ネの出現頻度から各ジャンルの特徴を考察した。その結果、チョット・スコシ、ヨ・ネの出現比率と、チョット・スコシのクラスター分析の結果から、BCCWJ と CSJ の各ジャンルは、文字言語と音声言語といった基準によって分類できることがわかった。一方で、チョットとスコシ、ヨとネをそれぞれ総合した形、あるいは対立させた形でジャンルごとにみると、話しことば的、あるいは書きことば的な特徴の有無による分類等が考えられることが明らかになった。以上のことから、本稿では、音声言語/文字言語という枠組みではない、各ジャンルのテキストの文体的位置づけが行われ得ることを示唆した。

1. はじめに

日本語の文体やスタイルと呼ばれるものは、一对一の対話や、一对多の講演、さらに新聞や雑誌、ネット上の書き込みなど、どのような場で用いられるかによって、様々なものがあるが、それらの文体やスタイルについて、書きことば/話しことばという観点と、音声言語/文字言語という観点は区別されるべきである。

例えば、ネット上の個人の HP と報告レポートでは、同じ書記言語であっても、前者は話しことば的な性質を持ち、後者は書きことば的な性質をもつことが予想される。他にも、ブログと一对一の対話では、いずれも話しことばの表現が使われるとしても、音声言語と文字言語という差異がある。さらに、その違いから、同じ文体とは言えない側面が見いだせる可能性がある。

本稿では、このように様々な場で用いられる日本語がお互いにどのような関係にあり、どのように区分できる可能性があるのか、様々なジャンルのテキストを用いて考察する。コーパスを用いた文体研究については、いくつか散見されるが、その中で文体が異なるコーパスを比較したものとしては小磯他(2009)が挙げられる。小磯他(2009)では、現

代日本語書きことば均衡コーパス、及び日本語話し言葉コーパスで用いられている語と語種を比較している。その結果、漢語率において、国会会議録や学会講演は、小説より「新聞や白書に近い傾向を示す」、「白書、新聞、学会講演よりも小説や模擬講演の方が、副詞率、形容詞率が高い傾向が見られ」といったことが示され、文字言語、音声言語の枠を越えて近い特徴を持つジャンルがあることが示された。小磯他(2009)は、さらに、漢語率や接続詞率、副詞率等、各ジャンルの使用傾向の考察に続き、線形判別分析を行っているが、漢語率や接続詞率など、「各ジャンルの判別に寄与」し、各ジャンルの類似度を計る上で役立つと考えられるものがあることを明らかにした。しかし、「話し言葉と書き言葉の比較を単純に接続詞率あるいは副詞率で見ることは妥当ではない」(小磯他 2009:596)と述べているように、品詞単位で文体を検討する場合、実際には同じ品詞の中にも様々な語がその中に含まれていることから、検討の余地があることを示唆している。

今回は、具体的な語の現れ方から、コーパスの様々なジャンルの文体を見ていくことを試みる。いくつかの語を選び、実際の用例数や用例そのものを検討しながら、各ジャンルの関係を観察し、各ジャンルの特性を明らかにする一助としたい。

2. BCCWJ・CSJ

本稿で使用するコーパス¹は、国立国語研究所コーパス開発センターによって構築された現代日本語書きことば均衡コーパス(以下 BCCWJ)と、国立国語研究所及び通信総合研究所によって開発された日本語話し言葉コーパス(以下 CSJ)である。いずれもコアデータ、非コアデータを合わせたデータで用例を収集した。BCCWJ は 13 のレジスター² (特定目的・ブログ、特定目的・国会会議録³、特定目的・知恵袋、特定目的・ベストセラー、出版・雑誌、図書館・書籍、出版・書籍、出版・新聞、特定目的・韻文、特定目的・

¹ 用例抽出にあたって、BCCWJ では「中納言」(通常版)(Version 2.1.1)を用いた。CSJ の用例については、第四版のデータを対象に、用例抽出には「ひまわり」を用いた。用例を「中納言」や「ひまわり」で一度収集した後、それぞれチョット、スコシ、ヨ、ネに該当しない例を除いた。

² 本稿では、ブログや国会会議録など特に BCCWJ の各々のジャンルについて言及する際には、レジスターと呼び、学会講演や対話など特に CSJ の各々のジャンルについて言及する際にはタイプと呼ぶが、これらはそれぞれのコーパス内での呼称に依るものである。

³ BCCWJ の国会会議録は、音声言語の記録というととらえられ方をされる可能性がある。しかし、BCCWJ を音声言語とすることには問題がある。国会会議録は CSJ のようなものと異なり、発話そのものを記録したものではなく、記録者が、読み手に発話内容がわかるように記録したものであるためである。したがって、発話そのものの記録であれば、記録されてあるはずのフィラーなども今回収集した用例の中では見られない。さらに、もし、国会会議録が CSJ と同様の音声言語であれば、今回明らかになった CSJ の各タイプに共通する性質を国会会議録も持っていると考えられるが、実際の結果はそうではない。

教科書、特定目的・広報誌、特定目的・白書、特定目的・法律)がある⁴。今回用いた CSJ のデータ⁵は、大きくわけて四つのタイプ(学会講演、模擬講演、その他、対話)がある。「その他」は、講演会や講義における独話音声である。

3. チョット・スコシ、ヨ・ネと文体

文章のスタイルによって使い分けられると考えられるチョット・⁶スコシ⁷、およびヨ・ネ⁸を取り上げ、それらが様々なジャンルの日本語のコーパスでどの程度用いられているのかみる。ここで、チョット・スコシ、ヨ・ネと、文体との関連について簡単にふれる。

3.1. チョット・スコシ

チョット・スコシは、『日本国語大辞典 第二版』によると、「ちよっと」は、副詞としては、「①時間、物事の量や程度がわずかであるさまを表わす語。大げさでなく行なうという気持ちをこめていう。ちっと。すこし」「②(打消しのことばを伴って)事態や判断が簡単には成立しがたいことを表わす語」「③差し障りのあること、具合の悪いことを曖昧にぼかして表わす語」という意味があり、感動詞としては「(①の『ちよっと…』の陳述部分が略された用法から)軽く相手に呼びかけるときのことばとして用いる」という意味があるとされている。「すこし」は一つの用法が挙げられ、「数量、程度が小であるさま。わずかに」という用法が記述されている。このように、「ちよっと」はいくつかの用法があり、副詞や感動詞として用いられるが、「すこし」は用法も一つに限られ、感動詞としての用法は取り上げられていない。

『大辞泉第二版』においても、「ちよっと」の複数の用法を持つことや、「ちよっと」が感動詞としても用いられる点、「すこし」については、一つの用法のみ記述されている点が、『日本国語大辞典第二版』と共通している。さらに、『大辞泉第二版』においては、「すこし」と「ちよっと」の用法を比較し、『ちよっと』は、ややくだけた言い方

⁴ 以下、BCCWJ の各レジスターについて言及する際は、「ブログ」や「新聞」の前の「特定目的」や「出版」を省略し、「ブログ」や「新聞」とする。ただし、「書籍」に限っては、「図書館・書籍」、「出版・書籍」の二つがあるため、省略しない。尚、「図書館」や「出版」を省略した「書籍」は、「図書館・書籍」と「出版・書籍」を合わせたものを指す。

⁵ 今回は「朗読/再朗読」のデータは用いていない。「朗読/再朗読」は、被験者が用意された文章を読み上げたデータであり、発話としての自発性が低いため除外した。詳細は、『日本語話し言葉コーパス』の概観 Version. 2.0 参照。

<URL: http://pj.ninjal.ac.jp/corpus_center/csj/manu-f/overview.pdf>

⁶ チョットとスコシ、ヨとネの間に挿入されている「/」「・」は、チョットとスコシ、もしくはヨとネを対立させていること、総合していることをそれぞれ示している。

⁷ チョット・スコシは異表記も含めて検索している(「ちよっと」「ちょっと」「一寸」「チョット」、「少し」「スコシ」など)。

⁸ ヨ・ネの用例についても、チョット・スコシと同様、「ヨ」「よー」「ねえ」など異表記のものが含まれている。

とされている。これらのことから、「ちょっと」は用法が多岐にわたり、格式張らない表現であると考えられる。

さらに、国立国語研究所(1955)によれば、副詞チョット、スコシは、話しことば⁹である談話語¹⁰と、書きことばと位置づけられる新聞および雑誌『主婦之友』とでは使用度数順位が異なる。国立国語研究所(1955)では、それぞれ使用度数が上位20位以内の副詞が挙げられているが、談話語では、「ちょっと」が8位、「すこし」が12位であるのに対し、新聞ではいずれも20位以内に含まれず、雑誌では、「すこし」4位、「ちょっと」16位となっている。この結果から、チョット・スコシは話しことばで用いられること、チョットは書きことばよりも話しことばで用いられる頻度が高い傾向が読み取れる。スコシは、書きことばの雑誌で、話しことばの談話語よりも使用度数順位が高いが、ニュースでは20位以内に含まれておらず、話しことばや書きことばといった枠組みでは説明しきれない部分も見られる。

チョットとスコシの差異については、ソピアック(2015)においてもBCCWJの用例を用いた考察があり¹¹、『ちょっと』は話し手と聞き手が存在する文脈に多く使われる」と述べている。さらに、チョットの用法が多岐にわたる点に関しても、「副詞として、『少量・小程度』『かなり』の意味で使われ、機能については、派生的な意味や配慮的なコミュニケーション機能」(ソピアック 2015:10)を有していると述べている。

この記述から、チョットは、スコシにはない機能があり、それが用いられる文脈にも特徴があることがうかがえる。

⁹ 「話しことば」「書きことば」の定義については様々なものがあるが、ここでは、「談話語」が話しことば、雑誌、ニュースが書きことばと位置づけられていることに関し、その裏付けとして、定延(2005)で挙げられている話しことばと書きことばの定義を挙げたい。定延(2005)では、Tannen(1980)、Chafe(1982)、定延(2003)を基に典型的な話しことばと書きことばの定義をまとめている。これによると、典型的な話しことばは、「(i)音声メディアで、(ii)公共性の低い個人の体験を、(iii)『この同じ場に一緒にいることを楽しむ』という共在重視のストラテジーに基づいて」「(iv)事前の準備なしに、とぎれとぎれに、冗長な表現で発せられるもの」である。一方で、典型的な書きことばは、「(i)文字メディアで、(ii)公共性の高い一般的知識を、(iii)「相手の知らない知識を伝えよう」という情報のやりとり重視のストラテジーに基づいて、(iv)事前に準備され、周到に練り上げられた高度に圧縮された表現で発せられるもの」(定延 2005:113)とされている。この定義を照らし合わせると、談話語は話しことば、雑誌とニュースは書きことば的な傾向が強いものと判断される。

¹⁰ 国立国語研究所(1955)による「談話語」は日常会話であり、話しことばとして調査されている。

¹¹ ソピアック(2015)では、BCCWJの用例を対象に「ちょっと」「すこし」の用法の違いについて言及しているが、用例検索にあたって年代を限定しているため、本稿とは抽出した用例数が異なる結果になっている。

3.2. ヨ・ネ¹²

終助詞ヨ・ネについては、既に多くの研究がなされているが、ここでは、本稿の内容に関わると考えられるものを中心に挙げたい。

まず、ヨ・ネは終助詞の一種であるが、終助詞は、「話し手の種々の心情を相手に伝えるとともに、完結した文にする働きをする」(山口・秋本 2001)とあるように、それが用いられることにより、何らかの話し手の心情が表されるとしている。

ヨ・ネについて、文体の面では、中村(1951)が、話しことばでは「終助詞を好んで用いる」一方で、書きことば¹³は「終助詞をあまり用いない」と述べている。その後、実際の談話を調査した国立国語研究所(1955)では、「談話語では文末助詞¹⁴のあるものが73%を占めているのに対してニュースでは全く文末助詞は現れなかった」(国立国語研究所 1955:117)と述べ、この73%の中で「よ」(14.8%)「ね」(25.0%)が多く見られるという結果を示している¹⁵。談話語は話しことばとみられ、ニュースは音声媒体としているが、基本的に原稿を読み上げるものであることから、書きことば的であるとする、話しことばでは終助詞が用いられることが多いが、書きことばでは終助詞はあまり用いられないという傾向が読み取れる。さらに、レポートや論文では、「書き手自身の存在を強く感じさせる語・表現や、読み手に話しかけるような語・表現は、余り使わない」(定延 2005:115)ため、終助詞を用いないとしているものもあり、ヨ・ネを用いるか否かは、ジャンルによってかなり異なると考えられる。

ヨ/ネの差異については、文体の違いと直接結びつけることは困難だが、個々のジャンルの文章の特徴を把握するために、ここで簡単に取り上げる。益岡・田窪(1992)によると、『ね』は相手も当該の知識を持っていると想定される場合に用いられる。自分の知識と相手の知識が一致していると想定し、これを相手に確認するときは同意要求になり、自分の知識が不確かなときは確認になる、『よ』は、基本的には、相手が知らないことに注意を向けさせる働きをする。したがって、場合によって、単なる知らせ、注意、警告、等の様々な意味を表す」としている。このように、ヨ、ネの使用には自分の

¹² 助詞ヨ・ネに関しては、間投助詞の用法と終助詞の用法を分ける立場と分けない立場があるが、今回は、ヨ・ネともに、終助詞の用例としても間投助詞としても考えられる用例を終助詞として併せて用例収集等行った。いずれの用法も話しことばでよく用いられると考えられる。

¹³ 中村(1951)は、話しことばと書きことばの差異を「話しことばと文字ことばとの構造上の相違」としてこのほかにもいくつか挙げている。尚、中村(1951)の「文字ことば」はここでいう書きことばに相当するものと考えられる。国語学会(1999)においても、中村(1951)の挙げたこれらの違いを話しことばと書きことばの違いとしている。

¹⁴ 「文末助詞」は終助詞として理解してよいと考えられる。

¹⁵ 間投助詞についても、談話の文の中で8.3%の割合を占めているが、ニュースでは見られなかったとしている(国立国語研究所 1955:130-132)ため、間投助詞としてもヨ・ネは話しことば的な場では見られるが、書きことば的な場では見られないということが予想される。

知識や相手の知識の状態が大きく関与する。加えて、ネには「間投用法」があることについても言及されている。

ヨ・ネは、それを用いることによって伝達場面や聞き手の特徴がどのようなものであるのか示し、またそれを用いることによってある発話効果が生じることがわかる。さらにヨとネはそれぞれ、自分の知識や相手の知識の状態が大きく関与し、使い分けがなされている。

4. チョット・スコシの出現頻度と分析結果

まず、チョット/スコシの出現頻度と割合、クラスター分析の結果をみる。

4.1. BCCWJにおけるチョット・スコシのレジスター別出現頻度

BCCWJにおけるチョット、スコシの出現頻度をレジスター別にまとめた結果が表1である。図1は、各レジスターの語数の中でのチョット・スコシの割合をグラフ化したものである。

表1 BCCWJにおけるチョット・スコシの用例数と割合

レジスター	チョット	チョットの割合	スコシ	スコシの割合	語数 ¹⁶	チョット・スコシの割合
ブログ	7,391	0.07 ¹⁷ %	5,853	0.06%	10,194,143	0.13%
知恵袋	4,159	0.04%	5,029	0.05%	10,256,877	0.09%
国会会議録	2,790	0.05%	1,341	0.03%	5,102,469	0.08%
ベストセラー	1,247	0.03%	1,633	0.04%	3,742,261	0.08%
図書館・書籍	6,922	0.02%	10,569	0.03%	30,377,863	0.06%
雑誌	1,199	0.03%	1,330	0.03%	4,444,492	0.06%
出版・書籍	4,777	0.02%	8,715	0.03%	28,552,283	0.05%
韻文	11	0.00%	84	0.04%	225,273	0.04%
教科書	37	0.00%	184	0.02%	928,447	0.02%
新聞	80	0.01%	174	0.01%	1,370,233	0.02%
広報誌	145	0.00%	277	0.01%	3,755,161	0.01%
白書	2	0.00%	52	0.00%	4,882,812	0.00%
法律	0	0.00%	0	0.00%	1,079,146	0.00%
計	28,760	0.03%	35,241	0.03%	104,911,460	0.06%

¹⁶ 短単位での語数で、コア・非コア含む。

¹⁷ 以下、表で挙げる割合は全て小数第二位までを表示する。

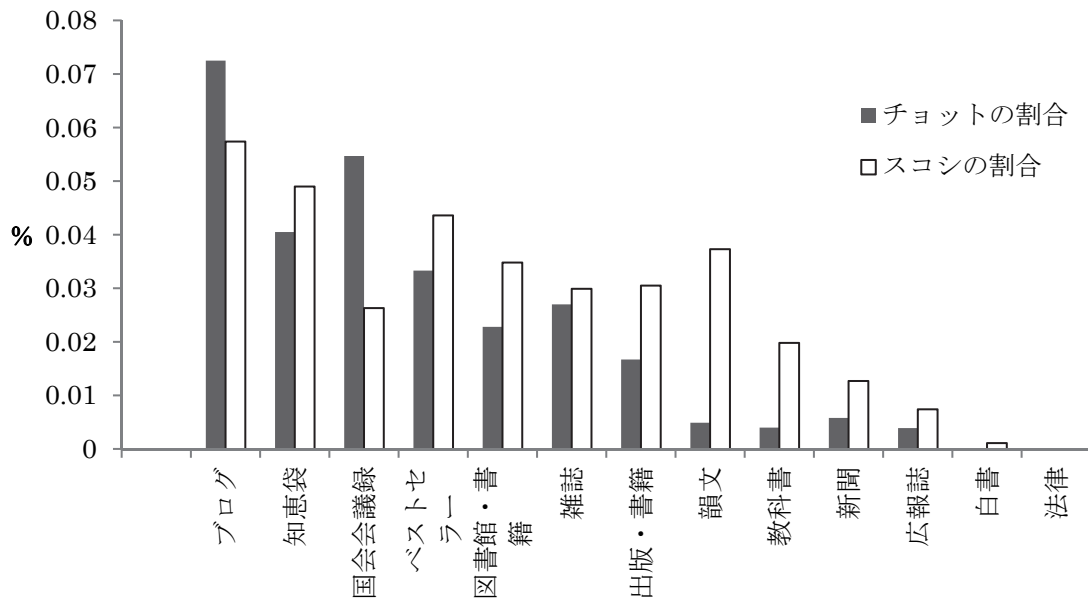


図1 BCCWJ におけるチョット・スコシの割合

BCCWJ 全体における、チョット・スコシの割合は、0.06%、また内訳としては、スコシ(0.03%)、チョット(0.03%)であり、スコシ/チョットの割合は同程度であることがわかる。レジスター別にみると、チョット・スコシの割合は、ブログ(0.13%)、知恵袋(0.09%)、国会会議録(0.08%)、ベストセラー(0.08%)、図書館・書籍(0.06%)、雑誌(0.06%)、出版・書籍(0.05%)、韻文(0.04%)、教科書(0.02%)、新聞(0.02%)、広報誌(0.01%)、白書(0.00%)、法律(0.00%)の順に高く、レジスターごとにチョット/スコシの割合にばらつきが見られた。

この中で、スコシ、チョットがいずれも最も高い割合で用いられているのがブログであり、チョットの方が割合は高い。続いて国会会議録では、ブログに次いでチョットの割合が高いが、ブログよりもスコシの割合が低い。ブログと国会会議録は、チョットがスコシよりも高い割合である点で共通している。

次に知恵袋やベストセラー、雑誌、書籍、新聞、韻文、教科書、広報誌だが、スコシがチョットよりも割合が高いという特徴がある。知恵袋に関しては、この8つの中で最もチョット・スコシの用例が多い。ベストセラーや書籍に関しては、会話文が含まれることがチョット・スコシの出現頻度に影響していると考えられる。雑誌、新聞、教科書、広報誌は、様々なインタビューやコラム、時事解説などの要素で成り立っているため、ある程度チョットやスコシが用いられるが、知恵袋やベストセラーに比べ、チョット・スコシの割合が低いと推測される。

最後に、白書や法律では、白書、法律の各レジスター別語数におけるチョット・スコシの割合が、他と比較して少ないか、もしくは該当する用例がない。

以上のことから、BCCWJ の各レジスターは、チョット/スコシの割合に着目すると、
 1. ブログ、国会会議録のようにチョットの割合が高いタイプ、2. 知恵袋、書籍、新聞、

韻文、教科書のようにある程度チョット、スコシが用いられるがスコシの方が多く用いられるタイプ、3. 白書、法律のようにチョット、スコシの用例の割合が非常に低い、もしくは全く用いられないタイプに分類可能である。

4.2. CSJにおけるチョット・スコシのタイプ別出現頻度

CSJにおけるチョット・スコシの用例数と割合は以下の通りである。表2にチョット、スコシの用例数とその割合をまとめた。チョット、スコシの割合をグラフ化したものが図2である。

表2 CSJにおけるチョット・スコシの用例数と割合

タイプ	チョット	チョットの割合	スコシ	スコシの割合	短単位数	チョット・スコシの割合
対話	328	0.22%	37	0.02%	149,826	0.24%
模擬講演	7,421	0.21%	1,263	0.04%	3,605,729	0.24%
その他	487	0.17%	127	0.04%	282,728	0.22%
学会講演	4,531	0.14%	1,200	0.04%	3,279,364	0.17%
計	12,767	0.17%	2,627	0.04%	7,317,647	0.21%

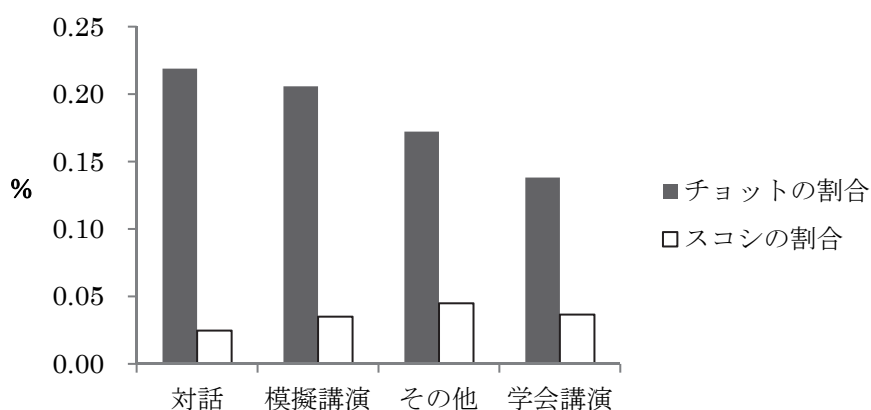


図2 CSJにおけるチョット・スコシの割合

CSJにおけるチョット・スコシの割合は、0.21%、そのうちチョットは0.17%、スコシは0.04%である。タイプ別にみると、対話(0.24%)と模擬講演(0.24%)が他の二つのタイプよりも高く、その他(0.22%)、学会講演(0.17%)と続く。いずれのタイプもチョットの割合は、0.1-0.2%である。特に、対話、模擬講演でチョットが多く用いられている。スコシの割合は、チョットに比べて低く、全体に占める割合も0.02-0.04%である。比率はタイプによって異なるが、どのタイプもチョットの割合がスコシに比べて高く、その点においてはBCCWJのブログや国会会議録と同様である。

4.3. BCCWJ と CSJ におけるチョット/スコシのクラスター分析結果

今回収集した BCCWJ・CSJ の用例については、さらに統計学的な見地からデータを検証するため、クラスター分析（Ward 法、平方ユークリッド距離）を行った。図 3 は、チョット・スコシの用例を合わせたものを分析した結果である¹⁸。法律は、用例数 0 例のため、分析の対象から除外している。

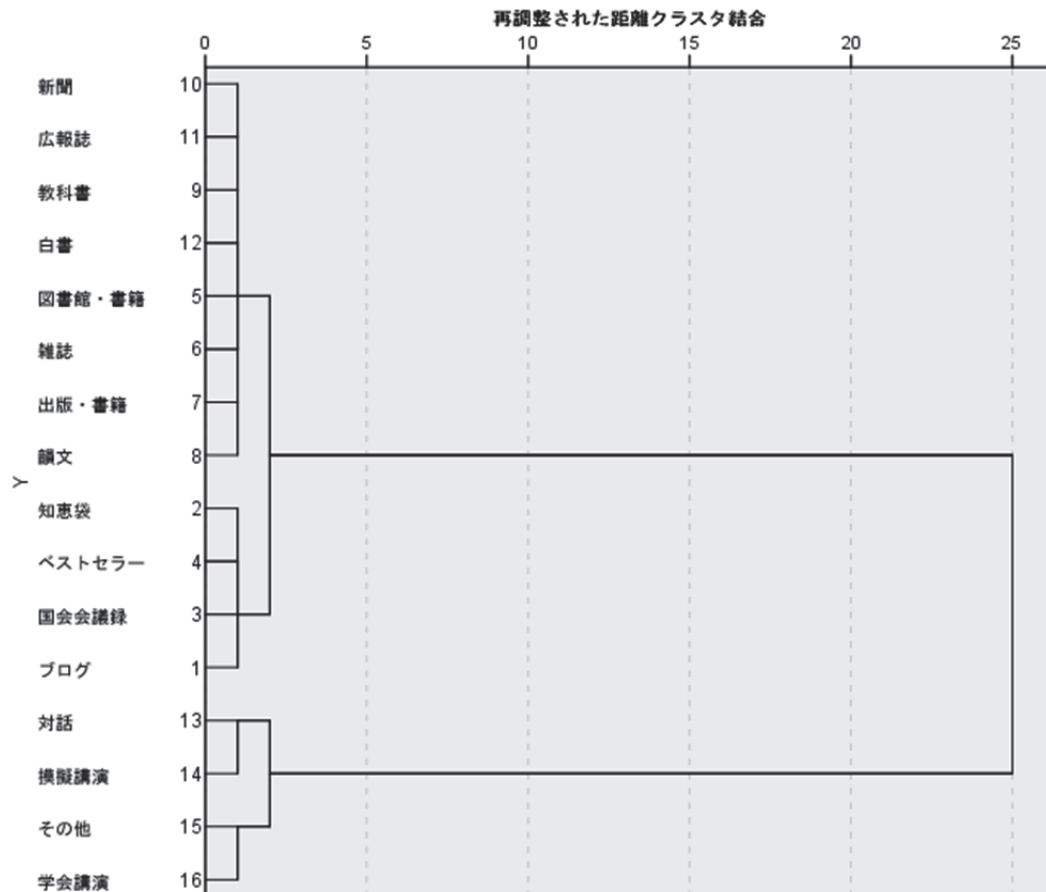


図 3 チョット・スコシの割合を基にしたデンドログラム

チョット・スコシについては、大きく二つのクラスターに分かれる(図 3)。新聞や知恵袋などが含まれるクラスターと対話などが含まれるクラスターである。

この二つのクラスターをさらに詳しくみると、BCCWJ のジャンルのクラスターの中では、早い段階で融合する二つのクラスターに分かれている。新聞や韻文のクラスターと、知恵袋やブログのクラスターである。新聞や韻文が含まれるクラスターは、知恵袋やブログのクラスターと比べ、チョット・スコシの割合が低い。

CSJ のジャンルでまとまったクラスターは、対話と模擬講演、その他と学会講演でそ

¹⁸ デンドログラム(図 3、図 7)については、チョット/スコシ、ヨ/ネいずれもヨ・ネ、チョット・スコシをそれぞれ合わせたものと同様の結果である。

それぞれクラスターを形成し、それらが早い段階で CSJ の一つのクラスターを形成している。対話と模擬講演は、今回扱ったジャンルの中で最もチョット・スコシの割合が高いジャンルであり、チョット・スコシの出現頻度は、その他や学会講演よりも対話と模擬講演の方が高い。

以上、クラスター分析の結果から、BCCWJ と CSJ の各ジャンルで大きく二つのクラスターに分類され、さらに、BCCWJ・CSJ で分かれた各クラスターをみると、チョット・スコシの割合により、さらに分類されることが読み取れた。

4.4. チョット/スコシの出現頻度、クラスター分析結果についての考察

BCCWJ・CSJ におけるチョット・スコシの用例から、いくつかのことが明らかになった。

出現頻度の面からチョット・スコシの割合を見ると、BCCWJ 全体では 0.06%(チョット(0.03%)/スコシ(0.03%))、CSJ 全体では 0.21%(チョットは(0.17%)/スコシは(0.04%))であり、CSJ の方がチョット・スコシの割合が高く、また、その中でチョットの割合が特に高いことが明らかになった。チョット・スコシは、3.1 で述べた通り、いずれも話しことばにおいて出現頻度が高い語であると考えられるが、そのような語が BCCWJ よりも CSJ において多く用いられることがわかる。さらに、CSJ 全体として BCCWJ より、より話しことばで多く用いられる語の出現頻度が高いと考えられる。チョットは国立国語研究所(1955)において、談話語に多くみられたが、CSJ においても多くみられるという結果であった。

ジャンル別にチョット・スコシの割合をみると、CSJ の対話(0.24%)と模擬講演(0.24%)、その他(0.22%)、学会講演(0.17%)、BCCWJ がそれに続き、ブログ(0.13%)、知恵袋(0.09%)、国会会議録(0.08%)、ベストセラー(0.08%)、図書館・書籍(0.06%)、雑誌(0.06%)、出版・書籍(0.05%)、韻文(0.04%)、教科書(0.02%)、新聞(0.02%)、広報誌(0.01%)、白書(0.00%)、法律(0.00%)の順に高い。ジャンル別ではこの順に、チョット・スコシのような話しことばにおいて出現頻度が高い副詞の出現頻度が高いことが明らかになった。

チョット/スコシの割合の差異に着目すると、各ジャンルは、1.チョットの割合が高いタイプ(ブログ、国会会議録、CSJ の全てのタイプ)、2.ある程度チョット、スコシが用いられるがスコシの方が多く用いられるタイプ(知恵袋、ベストセラー、書籍、雑誌、新聞、韻文、教科書、広報誌)、3.チョット、スコシの用例の割合が非常に低い、もしくは全く用いられないタイプ(白書、法律)に分けられる。ソピアック(2015)が『『ちょっと』は話し手と聞き手が存在する文脈に多く使われる』と述べていることを踏まえると、上記1はこの中で最も「話し手と聞き手が存在する」、もしくはそれに近い状態での会話の記録や文章であることが考えられる。

クラスター分析の結果(図 3)からは、各ジャンルが、新聞や知恵袋などが含まれるクラスターと、対話などが含まれるクラスターの二つのクラスターに大きく分類されるこ

とがわかった。これは BCCWJ と CSJ の別と一致しており、結果的に、文字言語と音声言語で大きく分かれている。このことから、チョット・スコシは、文字言語と音声言語の区別をする指標として用いられる可能性がある。

また、BCCWJ と CSJ で分かれたクラスターは、それぞれ二つのクラスターが融合して形成されたものである。チョット・スコシの割合によって分けられたクラスターが BCCWJ と CSJ のそれぞれのクラスターを早い段階で形成する形になっている。

BCCWJ のジャンルが全て含まれたクラスターでは、知恵袋、ベストセラー、国会会議録、ブログが分類されているクラスターとそれ以外のレジスターが分類されているクラスターが早い段階で融合し、BCCWJ のレジスターでまとまったクラスターを形成している。知恵袋、ベストセラー、国会会議録、ブログは実際に発話されたか、会話のスタイルで書かれた文章が多いことが考えられる。国会会議録は、実際の発話を記録したものであり、知恵袋やブログはネット上でのやりとりではあるが、会話のスタイルで書かれたものが多く見受けられる。ベストセラーも文学のジャンルの比率が高い¹⁹ことから会話のスタイルの文章が多く見受けられる可能性がある。今回の結果のみで結論づけることは難しいが、これらのレジスター内の会話のスタイルの文章の占める割合が高いものとそうでないもので分けられていることが考えられる。

CSJ のジャンルでまとまったクラスターをみると、対話と模擬講演のクラスター、その他と学会講演のクラスターが早い段階で融合している。今回扱ったジャンルの中で最もチョット・スコシの割合が高い対話と模擬講演²⁰は、公的な場での発話ではないため、場の改まり度が低いという特徴を持っていると考えられる。一方で、その他や学会講演は、実際の講演や講義の場での発話を記録したものであり、対話と模擬講演よりも場の改まり度が高い可能性がある。

5. ヨ・ネの出現頻度と分析結果

次に、ヨ・ネの出現頻度²¹と割合、クラスター分析の結果をみる。

¹⁹ 書籍やベストセラーの各テキストは、「総記」「哲学」「歴史」「社会科学」「自然科学」「技術・工学」「産業」「芸術・美術」「言語」「文学」の 10 のジャンルに細分化され、それに当てはまらないものについては「分類なし」という情報が付与されている。チョット・スコシの用例は、書籍、ベストセラーで、いずれも「文学」が最も高い割合を示している。ベストセラーのチョット・スコシの用例の中で、「文学」に位置づけられるのは、各ジャンルの全語数中 0.05%、同様に図書館・書籍では 0.03%、出版・書籍では 0.02%である。ベストセラーは書籍と比較すると、「文学」の割合が高い。

²⁰ 『日本語話し言葉コーパス』の概観（注 4）によれば、模擬講演は、「できるだけ年齢と性別のバランスをとった一般話者による、日常的話題についての講演」であり、「あらかじめ指定された 3 種の一般的テーマ」についてのスピーチである。「発話スタイルは概して学会講演よりもくだけたものとなっている」と述べられている。

²¹ 今回は「ヨネ」の形式については扱っていない。用例収集の際、それぞれ、ヨにネが後続する例、ネにヨが後続する例を除いた。

5.1. BCCWJにおける助詞ヨ・ネの出現頻度

BCCWJにおけるヨ・ネの頻度と割合を表3にまとめた。また、ヨ・ネそれぞれの各レジスターあたりの割合について図4にまとめた。

表3 BCCWJにおけるヨ・ネの頻度と割合

レジスター	ヨ	ヨの割合	ネ	ネの割合	語数 ²²	ヨ・ネの割合
知恵袋	32,702	0.32%	30,993	0.30%	10,256,877	0.62%
ブログ	22,193	0.22%	38,863	0.38%	10,194,143	0.60%
韻文	489	0.22%	86	0.04%	225,273	0.26%
ベストセラー	7,719	0.21%	6,370	0.17%	3,742,261	0.38%
図書館・書籍	43,528	0.14%	37,885	0.12%	30,377,863	0.27%
雑誌	5,088	0.11%	5,560	0.13%	4,444,492	0.24%
出版・書籍	30,041	0.11%	24,519	0.09%	28,552,283	0.19%
国会会議録	3,814	0.07%	6,559	0.13%	5,102,469	0.20%
教科書	542	0.06%	412	0.04%	928,447	0.10%
新聞	320	0.02%	364	0.03%	1,370,233	0.05%
広報誌	420	0.01%	609	0.02%	3,755,161	0.03%
白書	2	0.00%	3	0.00%	4,882,812	0.00%
法律	0	0.00%	0	0.00%	1,079,146	0.00%
計	146,858	0.14%	152,223	0.15%	104,911,460	0.29%

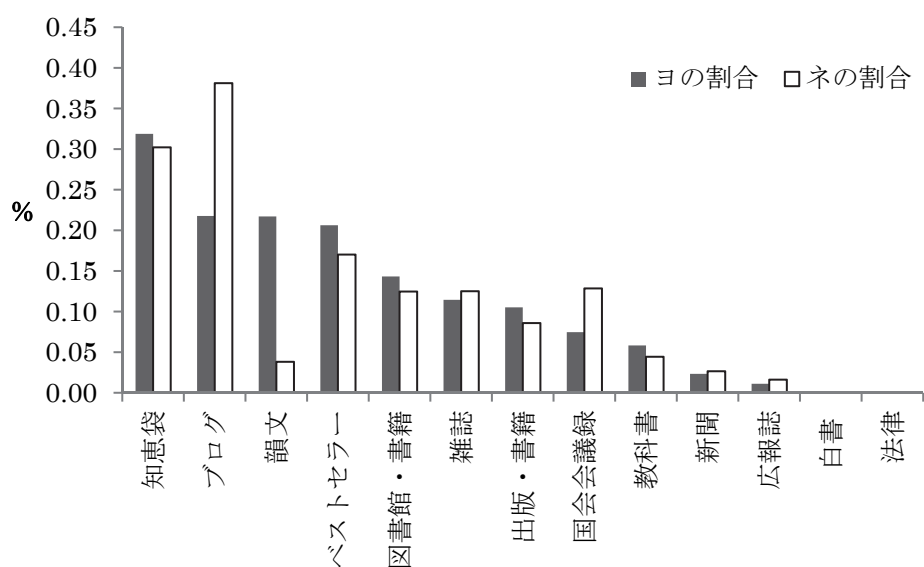


図4 BCCWJにおけるヨ・ネの割合

²² 短単位での語数で、コア・非コア含む。

BCCWJ では、全てのレジスターを合わせてヨ・ネが 0.29%を占め、ヨとネの割合としては、ヨ(0.14%)、ネ(0.15%)で頻度に大きな差異はない。しかし、レジスターごとにみると、ヨとネで割合が異なる。

この中で、知恵袋(0.62%)、ブログ(0.60%)が、ヨ・ネの割合が最も高い。ブログは、ヨよりもネの割合が高く、知恵袋はネよりもヨの割合が高い。

韻文は、ヨ・ネの割合が 0.26%で、知恵袋、ブログに続き、ヨの割合が高い(ヨ(0.22%)、ネ(0.04%))。これは、韻文のヨ・ネの用例が、詩や短歌、俳句の用例で占められており、ヨについては、それらの一種の技法としてヨが多用されているためである²³。韻文では、ネの割合は低い。

ベストセラー(0.38%)、図書館・書籍(0.27%)、雑誌(0.24%)、出版・書籍(0.19%)、国会会議録(0.20%)、教科書(0.10%)は、知恵袋やブログよりもヨ・ネの割合が低いが、新聞や広報誌、白書よりは、ヨ・ネが高い割合で用いられている。ベストセラー、書籍、教科書は、ヨよりもネの方が高い割合であり、雑誌、国会会議録は、ネよりもヨの方が割合が高い。

その他、新聞(0.05%)、広報誌(0.03%)、白書(0.00%)は、割合が知恵袋やブログと比較して低い。法律では用例が 0 例であった。

5.2. CSJ における助詞ヨ・ネの出現頻度

CSJ で用いられている、助詞ヨ・ネの頻度と割合を表 4 にまとめた。ヨ・ネそれぞれの割合については、タイプごとに図 5²⁴、図 6 にまとめた。

表 4 CSJ におけるヨ・ネの頻度と割合

タイプ	ヨ	ヨの割合	ネ	ネの割合	短単位数	ヨ・ネの割合
対話	444	0.30%	3,603	2.40%	149,826	2.70%
その他	135	0.05%	3,908	1.38%	282,728	1.43%
模擬講演	4,203	0.12%	30,181	0.84%	3,605,729	0.95%
学会講演	553	0.02%	12,850	0.39%	3,279,364	0.41%
計	5,335	0.07%	50,542	0.69%	7,317,647	0.76%

²³ ヨについては「一握の 米の白さよ 虫の声」<栗生純夫(1981)『増補現代俳句大系』角川書店>のように、俳句や短歌の一種の技法として用いられている用例が 489 例中 319 例見られる。

²⁴ ヨとネの割合に差があるため、別々のグラフに示した。

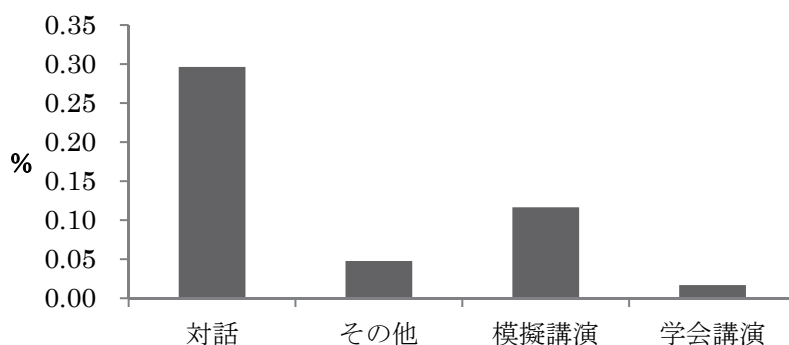


図5 CSJにおけるヨの割合

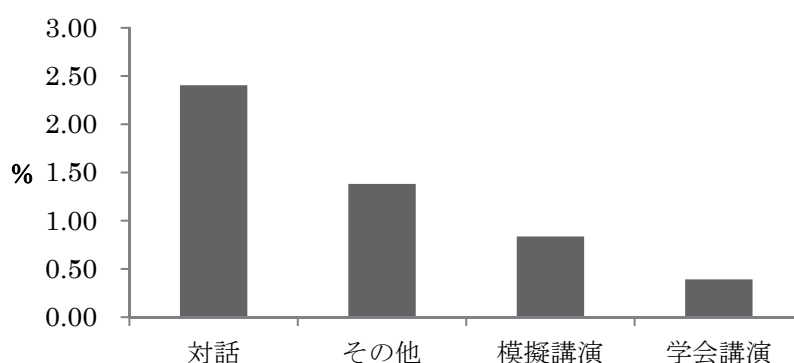


図6 CSJにおけるネの割合

CSJ 全体のヨ・ネの割合は 0.76%である。ヨの割合は 0.07%、ネの割合は 0.69%であり、ネの割合が高いのが特徴である。

タイプごとにみると、対話(2.70%)、その他(1.43%)、模擬講演(0.95%)、学会講演(0.41%)の順にヨ・ネの割合が高く、いずれのタイプもヨよりもネの割合が高い。この中でヨ・ネいずれも最も高い割合を示している対話は、ヨ(0.30%)、ネ(2.40%)である。対話に続き、ヨ・ネの割合が高いタイプであるその他は、他の CSJ のタイプと比べ、ヨ(0.05%)、ネ(1.38%)と、相対的にヨの割合が低い。模擬講演((ヨ(0.12%)ネ(0.84%))と学会講演(ヨ(0.02%)ネ(0.39%))では、模擬講演の方がヨ、ネいずれも割合が高い。

5.3. BCCWJ と CSJ におけるヨ/ネのクラスター分析結果

図7は、ヨ・ネの用例を合わせたものを分析した結果である。

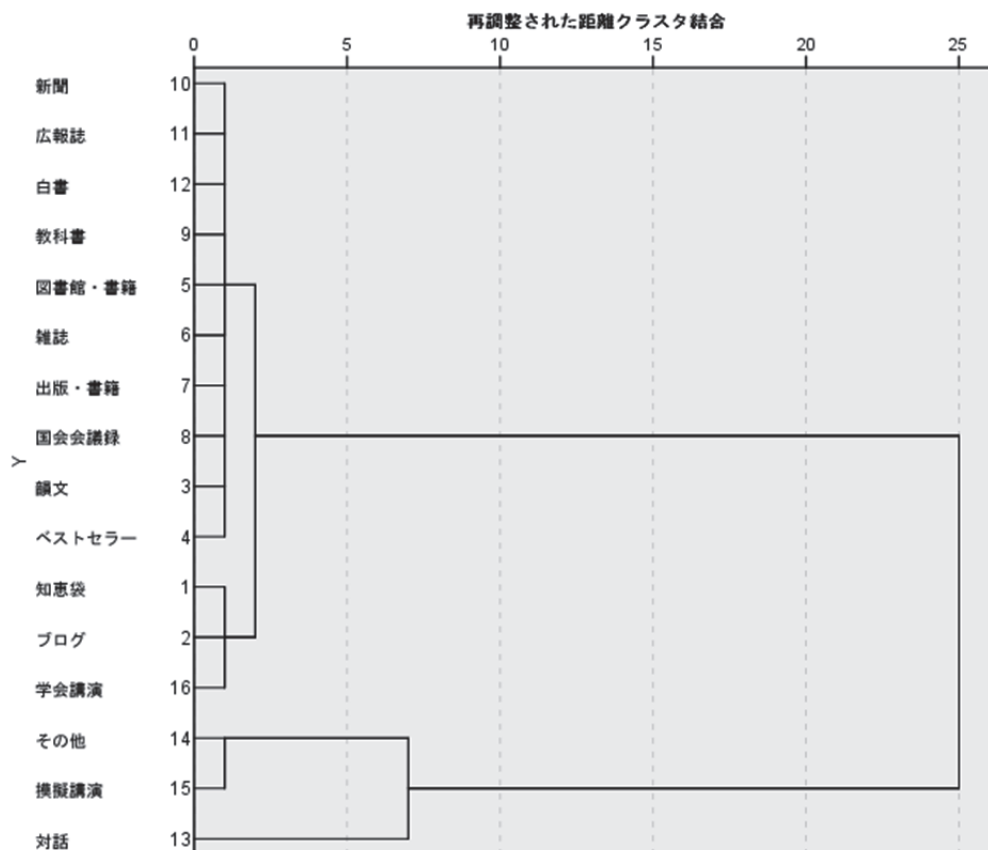


図7 ヨ・ネの割合を基にしたデンドログラム

ヨ・ネの結果をみると、CSJの学会講演以外のその他、模擬講演、対話が一つのクラスターを形成し、CSJの学会講演とBCCWJのジャンルすべてが一つの大きなクラスターを形成している。

CSJのその他、模擬講演、対話のクラスターは、学会講演、BCCWJの各ジャンルなどが含まれるクラスターと比べ、ヨ・ネが多く用いられるものであり、ヨ・ネが多く用いられるジャンルでまとまっているようである。このクラスターは、その他、模擬講演が分類されたクラスターに対話も加わった形で成立している。対話はヨ・ネの割合が最も高いものであるため、その点で区別されていると考えられる。

CSJの学会講演とBCCWJのジャンルすべてが含まれるクラスターは、新聞や広報誌、ベストセラーなどを含むクラスターと、知恵袋、ブログ、学会講演が含まれるクラスターに分かれている。これらは、早い段階で融合する。知恵袋、ブログ、学会講演は、その他、模擬講演、対話に続きヨ・ネの割合が高いものであり、新聞や広報誌、ベストセラーのクラスターと分けられている。

5.4. ヨ/ネの出現頻度、クラスター分析結果についての考察

5.1、5.2より、BCCWJ、CSJにおけるヨ・ネの出現頻度の結果から、ヨ・ネの割合は、

BCCWJ(0.29%)よりもCSJ(0.76%)の割合が高いことが明らかになった。ジャンル別にみると、対話(2.70%)、その他(1.43%)、模擬講演(0.95%)、知恵袋(0.62%)、ブログ(0.60%)、学会講演(0.41%)、ベストセラー(0.38%)、図書館・書籍(0.27%)、韻文(0.26%)、雑誌(0.24%)、国会会議録(0.20%)、出版・書籍(0.19%)、教科書(0.10%)、新聞(0.05%)、広報誌(0.03%)、白書(0.00%)、法律(0.00%)の順にヨ・ネの割合が高い。中村(1950)、国立国語研究所(1955)では、話しことばで終助詞が用いられる傾向があるとしていたが、この結果を見る限りは、CSJ 全体、BCCWJ のジャンルの中では知恵袋やブログも比較的話しことば的な性質を持つことが予想される。

CSJ の学会講演が、CSJ の他のタイプと比べ割合が低いことについては、学会講演の内容が影響し、学会講演の場は、専門的な内容を説明する場だと考えられるため、3.2で挙げた、「話し手の種々の心情を相手に伝えとともに、完結した文にする働きをする」(山口・秋本 2001)、終助詞の出現頻度が低いと考えることができる²⁵。

ベストセラーや書籍については、会話のスタイルの文章が含まれていることが、ヨ・ネの使用に関わっている可能性がある。特にベストセラーは、さらに細分化されたジャンル分けでは、書籍よりも文学の割合が高い²⁶。そのため、そこに含まれると考えられる会話のスタイルの文章が多くなり、書籍よりもヨ・ネが高い割合で用いられていると推測される。

雑誌、国会会議録、教科書は、知恵袋やブログよりもヨ・ネの割合が低い、新聞や広報誌、白書よりは、ヨ・ネが高い割合で用いられている。そのうち、雑誌や教科書は、様々なスタイルの文が入り交じっているが、その中に含まれる会話のスタイルの文章でヨ・ネが用いられると考えられる²⁷。国会会議録は、実際の国会でのやりとりを記録したもののだが、公的な場でのやりとりであるという点で、知恵袋やブログと異なっている。

²⁵ 学会講演については男性の発表者が多いため(『日本語話し言葉コーパス』の概観 Version. 2.0 p.4)、性差が影響しているとも考えられる。しかし、それを明らかにするためには具体的に各用例でどのようにネが用いられているのか、といった点についてみる必要がある。今後の課題とする。

²⁶ 上述の通り、書籍やベストセラーの各テキストは、10 のジャンルに細分化されている。これらのジャンルのうち、書籍、ベストセラーでは、いずれも「文学」が最も高い割合を示している。ヨ・ネの用例の中で、「文学」に位置づけられるのは、全語数ベストセラーで中 0.27%、同様に図書館・書籍では 0.17%、出版・書籍では 0.11%である。書籍と比較して、ベストセラーは「文学」の割合が高い。

²⁷ 雑誌、教科書では以下のような用例が見られる。(i)は雑誌、(ii)は教科書の用例である。

(i) でも、普段走っていてブレーキングだけは負けないとか、そういうつもりで走っていたことはないね。

<はた☆なおゆき(2002)『AUTO SPORT』三栄書房>

(ii) 第1次産業は農林水産業、第2次産業は鉱工業、第3次産業はサービス業などのことだよ。

<田邊裕ほか(2005)『新しい社会 地理』東京書籍株式会社>

国会のやりとりは、形式上は一对一のやりとりであっても、与党と野党、あるいは政治家と国民の間のやりとりというとらえ方もできる²⁸。

BCCWJ 新聞、広報誌、白書については割合が非常に低く、法律では用例が 0 例であった。これらのテキストに共通しているのは、報告や説明などがその主たる目的であって、基本的には対話という形をとらないということである。ヨ・ネが多く用いられている知恵袋やブログは、読み手に話しかける形や対話形式になっていると考えられるが、新聞や広報誌、白書はそのような形式をとるものではない。その点が、ヨ・ネの割合に影響していると推測される²⁹。法律は用例が見当たらなかったが、これは、法律が語りかける相手の存在を全く想定しないテキストであることを示している。

ヨ/ネのそれぞれの割合は、BCCWJ では頻度に大きな差はない(ヨ(0.14%)、ネ(0.15%))が、CSJ ではネの方が、割合が高い(ヨ(0.07%)、ネ(0.69%))。この結果については、CSJ の全てのタイプでヨよりもネの割合が高いことが影響していると考えられる。さらに、付け加えると、CSJ においてネの割合が高いことに関しては、CSJ は音声言語であり、間投助詞としてのネが使用されやすいことも考えられる。

ヨ/ネの割合をジャンル別にみると、1.ネの割合が高いタイプ(ブログ、雑誌、国会会議録、新聞、広報誌、CSJ の全てのタイプ)、2 ヨの方が多く用いられるタイプ(知恵袋、韻文、ベストセラー、書籍、教科書)、3.ヨ、ネの用例の割合が非常に低い、もしくは全く用いられないタイプ(白書、法律)の 3 つに分けられる。上記 1 のネの割合が高いジャンルは、「相手も当該の知識を持っていると想定される場合に用いられる」(益岡・田窪 1992)という機能を持つネが多く用いられるジャンルであり、3.2 で挙げた「談話語」と同じ特徴を有している。

尚、上で挙げた図 4-図 6 を見ればわかるように、国会会議録、CSJ のその他、ブログについては、他のジャンルと比べ、相対的にネの割合が高い。このことについては、それぞれのジャンルにおいて話し手や書き手に何が求められるのかが関係しているよう

²⁸ 国会でのやりとりには、事前に資料を準備し、それを基に進めるという流れがある点も、影響していることも考えられる。

²⁹ 広報誌、新聞、白書では、一部インタビュー記事など発話が引用される形でヨ・ネが用いられている。(i)は広報誌、(ii)は新聞、(iii)は白書の例である。

(i) 食べ終わった園児は、「おいしかった。全部食べちゃったよ」と空になったお皿を自慢げに見せてくれました。

<編著者不明(2008)『広報いず』2008 年 10 号静岡県伊豆市>

(ii) 「遠くから太鼓の音が聞こえてきたら、なんだかワクワクしませんか？僕は、血が騒ぐんですよ」。穏やかに、でも、力強く語った。

<神戸新聞社(2003)『神戸新聞』神戸新聞社>

(iii) B さんは本当に楽しそうに農業をしている。いくつになっても明日に夢をみられる仕事はいいですね。いつかは農業が自分の天職であると言えるようがんばります。」としており、周囲からも研修後の A 氏の活躍が期待されている。

<編著者不明(2001)『食料・農業・農村白書』(財)農林統計協会>

に考えられる。国会会議録、CSJ のその他³⁰については、益岡・田窪(1992)の「同意要求」や「確認」をしながら、聞き手にわかりやすく且つ丁寧に話を進める必要性があるために、ネが相対的に多い結果になっていると考えることができる。ブログについては、その性質上、ヨの出現頻度が低いことが可能性として考えられる。益岡・田窪(1992)によれば、『よ』は、基本的には、相手が知らないことに注意を向けさせる働きをする。したがって、場合によって、単なる知らせ、注意、警告、等の様々な意味を表す」とされているが、ブログは読み手の「知らないことに注意を向けさせる」ことではなく、むしろネット上で様々な人間とコミュニケーションをとることが目的としてあるように考えられる³¹。そのため、ヨ・ネの割合では同程度の知恵袋とヨ/ネの割合を比較すると、ヨの割合が低い結果になっていると推測される。

知恵袋³²は、ブログと同様ネット上で会話のスタイルで書かれた文章と考えられるが、ブログとは異なり、知恵袋が Q&A 方式のサイトであり、誰かが質問に回答する際に「相手が知らないことに注意を向けさせる働きをする」(益岡・田窪 1992)ヨを用いる機会がブログよりも多いため、相対的にヨの割合がブログよりも高くなっていると予想される。

ヨ・ネのクラスター分析結果は、5.3 に挙げた図 7 のように、CSJ のその他、模擬講演、対話が分類されるクラスターと、CSJ の学会講演と BCCWJ のジャンルすべてが含まれるクラスターの二つの大きなクラスターに分かれることを示している。CSJ は、その他、模擬講演、対話と、ヨ・ネの割合が CSJ のの中では最も低い学会講演が異なるクラスターに分類された形になったが、上で述べた通り、学会講演は、学会という場の特性を考えると、終助詞の出現頻度そのものが低い可能性がある。そのため、ヨ・ネのクラスター分析結果は、チョット・スコシのクラスター分析結果のように、BCCWJ と CSJ で分類されていない原因と推測することも可能である。

CSJ のその他、模擬講演、対話のヨ・ネの割合が高いクラスターは、最初にその他と模擬講演のクラスターがまず融合し、そのあとでヨ・ネの出現頻度がこの中で最も高い対話と結びつく形となっている。対話は、一対一で話すスタイルであり、模擬講演やその他は、一人の話し手が多くの聴衆に向けて話すスタイルである点が異なる。

CSJ の学会講演と BCCWJ のジャンルすべてが含まれるクラスターは、新聞や広報誌、ベストセラーなどを含むクラスターと、知恵袋、ブログ、学会講演が含まれるクラスター

³⁰ CSJ のその他は、学会講演や模擬講演と異なり、「話者と聴き手の関係が専門家と一般聴衆の関係である」音声文字化されたものである。(『日本語話し言葉コーパス』の概観 Version. 2.0 p.4)

³¹ ブログでは、「今回のツアーでよかったのは、赤レンガ倉庫。趣があっていいねえ。」<Yahoo!(2008) Yahoo!ブログ>のように、ネの同意要求と解釈できる用例が多くみられる。

³² 知恵袋は、「ユニホームやタオルなどは、どこに行ったら買えるんですか?」「ネットでも買えますよ。↓みたいなオフィシャルショップで買えます。」<Yahoo!(2005)Yahoo!知恵袋>といった Q&A によって成立しており、回答部分でヨが用いられることが多いようである。

一が早い段階で融合してできたものである。このクラスターは、知恵袋やブログ、学会講演よりもヨ・ネの割合が低い、文字言語、且つブログのようにネット上でのやりとりを記したものの以外のジャンルでまとまっているようである。国会会議録もこれらのジャンルと同じクラスターに分類されているが、実際に話された内容の記録であっても CSJ の各ジャンルとは異なる特徴を持っていることを示唆する結果となっている。

知恵袋、ブログ、学会講演は、CSJ のその他、模擬講演、対話に続き、ヨ・ネの割合が高いジャンルだが、音声言語で専門的な内容を聴衆の前で講演したものを記録した学会講演と、文字言語でネット上でのやりとりを記録したものであるブログ、知恵袋は、専門的な内容か否か、ネット上でのやりとりか否かといった差異がある。しかし、この三つのジャンルは、新聞や広報誌、ベストセラーなどが含まれるクラスターに分類されたジャンルが持つ特徴、即ち文字言語で、ブログのようにネット上でのやりとりを記したものの以外のジャンルとは反対の特徴を持つものであるととらえられる。

6. チョット、スコシ、ヨ、ネの出現頻度、クラスター分析結果についての考察

以上、チョット・スコシ、ヨ・ネを取り上げ、それらが BCCWJ、CSJ の各ジャンルでどのように用いられているのかをみた。チョット・スコシ、ヨ・ネは、文体と関わりがあるとみられる語であり、いずれも話しことばで用いられる語であると考えられるが、BCCWJ、CSJ におけるこれらの語の出現頻度やその割合、出現頻度の割合を基にしたクラスター分析の結果から明らかになったことをまとめる。

チョット・スコシ、ヨ・ネの割合をみると、いずれも CSJ の方が BCCWJ よりも出現比率が高く、その点においては、CSJ と BCCWJ の違いが表れる結果となっており、音声言語と文字言語の差異を示す結果と考えられる。

ジャンル別にチョット・スコシ、ヨ・ネで共通する特徴に着目すると、チョット・スコシ、ヨ・ネいずれも、出現頻度の割合が高いジャンルと低いジャンルが一致している。まず、CSJ の対話、模擬講演、その他は割合が高く、教科書、新聞、広報誌、白書、法律は割合が低い。ベストセラー、書籍、雑誌、韻文、国会議事録は、チョット・スコシ、ヨ・ネによって割合が変わるが、割合の高いジャンル(対話、模擬講演、その他)と、割合の低いジャンル(教科書、新聞、広報誌、白書、法律)の間に位置する。それ以外に、ブログと知恵袋も BCCWJ の中ではチョット・スコシ、ヨ・ネの割合が高いという特徴がある。CSJ の学会講演は、CSJ の各タイプの中で最も低い割合を示している。

チョット/スコシ、ヨ/ネの割合については、BCCWJ では、チョット/スコシ、ヨ/ネの割合に大きな差異はないが、CSJ では、スコシよりもチョット、ヨよりもネの割合が高く、CSJ が話しことば的要素をもっていることを示す結果が出た。一方で、用例の割合が非常に低い、もしくは全く用いられない白書、法律を除き、ジャンル別にチョット/スコシ、ヨ/ネを合わせて見た場合、CSJ と BCCWJ の区別とは異なる区分がなされることがわかる。それは以下の三つである。

1. スコシよりもチョットの割合が高く、ヨよりもネの割合が高いタイプ
 <ブログ、国会会議録、CSJの全てのタイプ>
2. チョットよりスコシの割合が高く、ヨよりもネの割合が高いタイプ
 <雑誌、新聞、広報誌>
3. チョットよりスコシの割合が高く、ネよりもヨの割合が高いタイプ
 <知恵袋、ベストセラー、書籍、韻文、教科書>

国立国語研究所(1955)の話しことばである「談話語」の調査結果では、スコシよりもチョットの割合が高く、ヨよりもネの割合が高かったことから、その点において、上記1は「談話語」と同じ特徴を持っている。上記2はヨよりもネの割合が高いという点のみ「談話語」と同じであるが、上記3はその2点においては、「談話語」の特徴と一致していない³³。

クラスター分析の結果については、チョット・スコシ、ヨ・ネそれぞれ、大きく二つのクラスターに分けられたが、チョット・スコシでは、BCCWJとCSJのジャンルで分かれたのに対し、ヨ・ネでは、CSJの対話、その他、模擬講演のクラスターとCSJの学会講演とBCCWJのすべてのジャンルを含むクラスターに分類された。しかし、この結果は、学会講演を例外とすれば、BCCWJとCSJで分類される結果が示されたとも見ることも可能である。もし、そのようにヨ・ネの結果をとらえるならば、クラスター分析の結果では、CSJのジャンルとBCCWJのジャンルの違いが明確に出たと判断される結果となる。しかし、それぞれの大きなクラスターを見ると、場の改まり度や、一対一の対話スタイルか一対多のスタイルか、といったことからジャンルの分類を解釈できる可能性も示された³⁴。

これらのチョット、スコシ、ヨ、ネの出現頻度等の結果から、BCCWJとCSJの各ジャンルは、チョット・スコシのクラスター分析結果のように、文字言語と音声言語という枠組みで分類されるものもあるが、その枠組みではとらえきれない、即ちBCCWJ、CSJの各ジャンルが混じった分類があり得ることが考えられる。

7. 今後の課題

今回は、いくつかの語の出現頻度からジャンルの差異を見出すことを試みたが、ジャンルの分類については、実際には様々な要因が関わっていると考えられる。したがって、ジャンルの分類を行うという点において、本稿はその一面を取り上げたに過ぎない。し

³³ チョット/スコシ、ヨ/ネの割合については、各ジャンルがどの程度話しことば的であるかの一つの尺度として考えられる可能性もあるが、その際は、文体とは異なる面でネよりもヨの割合が高いことから、3の知恵袋や韻文をこの分類から除いた方がよいと考えられる。

³⁴ 場の改まり度や会話のスタイルの文章の割合の間の関連性について、また本稿で挙げた、ジャンルの分類に関わると考えられる他の特徴については更に考察が必要である。

かし、チョット/スコシ、ヨ/ネの出現頻度を、そのテキストが話しことば的であるか、書きことば的であるかの評価の指標とできる可能性はある。

その上で、今後は、本稿で扱えなかった、書きことば特有の表現が各コーパスでどのように用いられているのかについても検討したい。

文末の「です・ます」についても、文体を考える際に重要であると考えられるが、今回は取り上げることができなかった。併せて考察することも今後の課題である。

参考文献

- エアン・ソピアック(2015)『日本語の「ちょっと」とクメール語の ʔəŋə/bəntəc/の対照』平成 26 年度修士論文、麗澤大学大学院言語教育研究科
- 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--短単位情報に着目して--」言語処理学会第 15 回年次大会 発表論文集、pp.594-597
- 国語学会編(1999)『国語学大事典 第十版』東京堂出版
- 国立国語研究所(1955)『談話語の実態』秀英出版
- 定延利之(2003)「体験と知識 コミュニカティブ・ストラテジー」『国文学 解釈と教材の研究』43-12 学燈社、pp.54-64
- (2005)「ケース 18 話しことばと書きことば(文字編)」『ケーススタディ 日本語のバラエティ』上野智子・定信利之、佐藤和之、野田晴美編、おうふう、pp.108-113
- 小学館大辞泉編集部編(2012)『大辞泉 第二版』松村明監修、小学館
- 中村通夫(1951)「談話のことばと放送のことば」『言語生活』11 月号、pp.16-20
- 日本国語大辞典第二版編集委員会・小学館国語辞典編集部(2001)『日本国語大辞典 第二版』小学館
- 益岡隆志・田窪行則(1992)『基礎日本語文法・改訂版』くろしお出版
- 山口明穂・秋本守英(2001)『日本語文法大辞典』明治書院
- Chafe, Wallace L. (1982). Integration and involvement in speaking, writing, and oral literature, In Tannen, Deborah (ed.), *Spoken and Written Language: Exploring Orality and Literacy*, Norwood, NJ: Ablex, 35-53.
- Tannen, Deborah(1980). Spoken/written language and the oral /literate continuum, *Proceedings of the sixth annual meeting of the Berkeley Linguistics Society*, University of California, Berkeley, 207-218.

