



Title	語彙から見た英語の話し言葉と書き言葉の相違 : ハリデーの語彙密度を中心に
Author(s)	梅咲, 敦子
Citation	大阪大学言語文化学. 1992, 1, p. 113-123
Version Type	VoR
URL	https://hdl.handle.net/11094/78205
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

語彙から見た英語の話し言葉と書き言葉の相違

—ハリデーの語彙密度を中心に—*

梅咲（古田）敦子**

The ultimate goal of this study is to elucidate lexical differences between spoken and written English. According to Halliday (1985), lexical density (the proportion of content words to total running words) is much higher in writing than in speech. In order to give some statistical evidence of lexical density, this article examines the English used in oral presentations and published papers. Two paper-oral presentation pairs were used for the analysis. In each pair, one oral presentation and one published paper with identical contents were produced by the same scientist. Frequencies according to word classes and type/token ratio were calculated. Comparison was made with the existing corpus data (Biber, 1989). The result obtained showed higher lexical density and higher type/token ratio in writing. Nouns and adjectives appeared more often in writing. Lexical verbs and pronouns were more often used in speech. A contrastive pair analysis indicated that the use of nouns and adjectives raises lexical density sharply and that the repetition of words reduces the amount of information. Lexical density is an important feature which distinguishes spoken from written English.

1 はじめに

英語の話し言葉 (spoken English) と書き言葉 (written English) の相違について、Halliday (1985, pp. 61-67, 79-80) は、語彙密度 (lexical density) が両者の重要な差であると指摘している。彼は、語彙密度とは、使用語数に対する内容語

*An Analysis of Spoken and Written English based on Halliday's Theory of Lexical Density (Atuko Furuta UMESAKI)

**言語文化研究科博士後期課程

の割合であり、書き言葉は話し言葉よりはるかに語彙密度が高いと主張している。しかし、証明としてあげられているのは、実際に書かれた文とそれに対応するように Halliday が話し言葉に置き換えた文のペア数個である。

そこで、本研究では、実際に同じ人が同じ内容を、話して伝える場合と書いて伝える場合でどの様に語彙の使用が異なるかを調べる。分析方法として、ペアを比較する（いわゆるミクロ的分析）のみならず、かなりの長さのテキストをコンピュータを使用して統計的にも比べる（マクロ的分析）。統計分析は、既存のコーパスのデータとの比較を可能にする。さらに、語彙使用の相違の原因を、内容語と機能語の両方から探る。

2 話し言葉と書き言葉

「話し言葉 (spoken language)」と「書き言葉 (written language)」は、それぞれ、音声により伝達される言葉、文字により伝達される言葉という意味で一般に用いられている。例えば、Biber (1989, p. 66) は London-Lund コーパスのことを「イギリス英語の話し言葉 (British spoken English) のコーパス」、LOB コーパスを「イギリス英語の書き言葉 (British written English) のコーパス」と説明し、実際、前者に入っているテキストはすべて、録音を文字化したものであるのに対し、後者のテキストは全て印刷物からとられている。

同時に、それらの特徴は口語体と文章体といった文体的概念とも深く関係していると言える。話し言葉と書き言葉は、すべてのジャンルのテキストを完全に2つのグループに分けるような特徴を持つのではなく、典型的な話し言葉のテキストと典型的な書き言葉のテキストを両端に持つ連続線上にさまざまなテキストが並ぶような特徴を持つ (Leech ら, 1982, p. 140)。この連続線上に並ぶテキストのうち、典型的な話し言葉に近い特徴を示すテキストを、音声伝達であれ文字伝達であれ、口語体のテキストと呼び、典型的な書き言葉に近い場合を文章体のテキストと呼ぶと考えることができる。

Halliday (1985, p. 61) も、実際に書かれた文を、話す時に用いられる形に換えた例を示す際、「話し言葉 (speech) に多く用いられるであろうが、特に口語的 (specially colloquial) というわけではない」という説明の仕方をしていることから、一般的な意味での話し言葉と書き言葉を連続線的に文体と関連づけているこ

とが解る¹⁾。

3 内容語と機能語

品詞は一般にその成員がほとんど不変で有限個であるものと、新造語が加わる可能性がある無限個のものに二分される。Halliday (1985, p. 61) は、前者に属する語を「文法語 (grammatical words)」または「機能語 (function words)」と呼び、後者に属する語を「語彙 (lexical words)」または「内容語 (content words)」と呼んでいる。Quirk ら (1985, pp. 67-72) は、品詞をクローズドクラス、オープンクラスとその他の3つの小さなカテゴリーに分けている。そして、クローズドクラスには、前置詞、代名詞、決定詞、接続詞、法助動詞、助動詞 *be, have, do* が属し、オープンクラスには、名詞、形容詞、動詞、副詞が入り、その他のカテゴリーとして、数詞、間投詞、及び *not* や不定詞の *to* などの属するクラスの3つをあげている。Halliday の機能語には、Quirk らのクローズドクラスとその他の小さなカテゴリーに属する語が入り、内容語はオープンクラスと同じであると考えられる。ただし、内容語と機能語の境界となる副詞、例えば *always, there, here* 等を、Halliday は機能語に分類している。動詞、助動詞は²⁾、例文から判断する限り、語彙的動詞 (lexical verbs) の *be, have, do* を助動詞とともに機能語と見なされている³⁾。

4 分析するテキスト

分析対象として、自然科学系の国際会議において、同じ科学者が同じ内容で行なった口頭発表と会議後に発行される論文集の論文を2組取り上げる。

2つの口頭発表は、1988年7月カリフォルニアで開かれた、第四回国際非晶

¹伝達方法と文体を明確に区別する用語としては、Halliday & Hasan (1985, pp. 58-59) は、前者に視角チャネル (graphic channel) と音声チャネル (phonic channel)、後者に話し言葉体 (spoken medium) と書き言葉体 (written medium) を用いている。

²Halliday (1985, p. 61) には、'finite verbs' (定動詞) が機能語であると書かれているが、一般的な定動詞の定義 (数、人称、時制を示す動詞) からすれば、不適切であるように思われる。

³Halliday は語 (words) とは別に項目 (items) という用語も使っている。句動詞 (*take off*) や句前置詞 (*according to*) はひとつの語彙項目 (a lexical item) であり、同じ語幹から派生した語、例えば、*differ, differed, different*, もひとつの語彙項目とみなしている。

質物質会議でなされた。それらは、筆者が録音し、各発表の開始から終了までのテープ起こしを行い、イギリス人科学者にチェックしてもらった。発表は2人とも、完全原稿を読むのではなく、準備した内容をメモや OHP に映されたグラフをもとに伝えていた。2つの論文は、会議の約半年後に発行されたプロシーディングズ⁴⁾ から取り、図表と注は分析用テキストからは除いてある。口頭発表と出版された論文の一方の組は、著名なイギリスの物理学者（40 才代前半の男性）によるもので、内容は「中性子と X 線による非晶質構造学」である⁵⁾。

もうひと組は、別の著名なイギリス物理学者（40 才代半ばの男性）による「ケイ酸塩ガラスにおけるカルシウムイオンの局所構造」についての口頭発表と論文である⁶⁾。

なお、本分析に用いるテキストは、シンタックス面では、LOB コーパスおよび London-Lund コーパスの各ジャンルとともにクラスター分析をした結果、口頭発表と論文の間に有意差のあることがすでに判明している（梅咲、1992）。

5 分析項目と方法

5.1 マクロ分析

4つのテキストをコンピュータに入力し⁷⁾以下の分析をする。

1. 総語数に対する総内容語の割合：プログラム KWICE⁸⁾ を使用して語彙頻度リスト及び KWIC コンコーダンス（前後文脈表示付単語リスト）を作成する。次に、それらを利用して内容語と機能語を分類し、割合を計算する。分類の大枠は Halliday に従い、内容語と機能語を構成する品詞の細かい定義は Quirk ら（1985）に従う。短縮形（*isn't* 等）は、Halliday の例に従い、一語として数える。ただし、次の分析のために、句前置詞等は 1 語として

⁴⁾Wagner, C. N. J. & Wright, A. C. (eds.), *The Structure of Non-Crystalline Materials* (North-Holland) 1988.

⁵⁾Wright, A. C., "Neutron and X-ray Amorphography"

⁶⁾Gaskell, P. H., "The Environment of Ca Ions in Silicate Glasses"

⁷⁾分析に使用したコンピュータ機器：富士通 M150F、NEC PC9801、富士通 FMR50、光学自動読み取り機 KURZWEIL K5100。

⁸⁾大阪大学言語文化研究科言語情報科学講座助手、三木邦弘氏作成。

扱う⁹⁾。

語彙的動詞としての *be, have, do* については、内容語の動詞として集計し、その数を付記する。

2. 内容語の使用特性：名詞、形容詞、副詞、動詞の総語数に対する割合を計算する。
3. 機能語の使用特性：機能語を品詞別に総語数に対する割合を計算する。
4. 出現語数に対する異なり語数の割合：ここでいう異なり語数とは、同じ語が何度使用されていても1つと数えた数値である。市販ソフト Micro-OCP¹⁰⁾ を使用して、テキストの最初から400語までに出現する異なり語数を4で割った数値を比較する。派生語や複数形など、形の異なるものは別の異なり語として数えるようプログラムされている。得られた値を、同じ要領で Biber(1989) が数えた LOB コーパスの各ジャンルの数値及び London-Lund(LLC) コーパスの各ジャンルの数値をそれぞれ平均した値と比較する。

5.2 ミクロ分析

マクロ分析で得られた数量的結果について、口頭発表と論文の対応する部分を手作業で抜き出し比較する。

6 結果と考察

6.1 マクロ分析

6.1.1 語彙密度

総語数に対する機能語と内容語の割合は、表1の通りである。

⁹⁾ 本分析中の総語数は、内容語と機能語の合計である。1語として数えられた句は、句前置詞：*according to/due to/* 句接続詞：*so far as/so that/* 句法助動詞：*ought to/be going to/have to/* 句副詞：*et. al./* その他：*in order to*。

¹⁰⁾ Copyright: Oxford University Press 1987.

表1：機能語と内容語の割合

	論文		口頭発表	
	イギリス人1	イギリス人2	イギリス人1	イギリス人2
総語数	6583	1819	4159	2819
総機能語数	3040	882	2246	1637
総内容語数	3543	937	1913	1182
機能語の割合	46.2%	48.4%	54.0%	58.1%
内容語の割合	53.8%	51.6%	46.0%	41.9%

総語数に対する総内容語の割合として語彙密度を見ると、論文のほうが口頭発表よりもふたりの場合とも語彙密度が高いことが判る。この結果は、Halliday の主張に一致する。

6.1.2 内容語

表2は内容語の品詞別使用数と総語数に対する割合を示している。

表2：内容語に関する品詞別使用状況

	論文		口頭発表	
	イギリス人1	イギリス人2	イギリス人1	イギリス人2
名詞	1941/29.5%	547/30.1%	992/23.8%	586/20.8%
形容詞	715/10.9%	148/ 8.1%	285/ 6.9%	128/ 4.5%
副詞	102/ 1.5%	41/ 2.3%	71/ 1.7%	48/ 1.7%
動詞	785/11.9%	201/11.1%	565/13.6%	420/14.9%
内 be, do, have	192/ 2.9%	42/ 2.3%	215/ 5.2%	153/ 5.4%

名詞、形容詞の使用率が、口頭発表よりも論文の方がふたりの場合とも高い。さらに、形容詞については、表3に示されているように、特に限定用法の使用率が論文において高い。即ち、語彙密度を押し上げているのは名詞および名詞を修飾する形容詞であると言える。Halliday (1985, p. 72) は、語彙的「内容」は動詞句と比べ名詞句が伝えると主張しているが、上述の結果はその主張に沿っている。

表3：形容詞の限定用法と叙述用法

	論文		口頭発表	
	イギリス人1	イギリス人2	イギリス人1	イギリス人2
限定用法	605/9.2%	119/6.5%	240/5.8%	97/3.4%
叙述用法	110/1.7%	29/1.6%	45/1.1%	31/1.1%

動詞¹¹⁾については、論文よりも口頭発表のほうが使用率が高い。これは、語彙的動詞としての *be, do, have* の使用が口頭発表に多いことが大きな原因である。

内容語の副詞については、イギリス人1は口頭発表に、イギリス人2は論文に多く用いており、傾向が異なっている。従って、内容語の副詞は、話し言葉と書き言葉を区別する特徴ではないと考えられる。

6.1.3 機能語

表4は機能語の品詞別使用数と総語数に対する割合を示している。

表4：機能語に関する品詞別使用状況

	論文		口頭発表	
	イギリス人1	イギリス人2	イギリス人1	イギリス人2
前置詞	932/14.1%	279/15.3%	469/11.2%	291/10.3%
代名詞	162/ 2.5%	31/ 1.7%	413/ 9.9%	310/11.0%
決定詞	915/13.9%	284/15.6%	544/13.1%	404/14.4%
接続詞	313/ 4.8%	84/ 4.6%	228/ 5.5%	186/ 6.6%
法助動詞	69/ 1.0%	15/ 0.8%	61/ 1.5%	56/ 2.0%
<i>be, have, do</i> 助動詞	136/ 2.1%	48/ 2.6%	45/ 1.1%	38/ 1.3%
副詞	226/ 3.4%	59/ 3.3%	267/ 6.4%	185/ 6.6%
数詞	6/ 0.1%	5/ 0.3%	16/ 0.4%	7/ 0.2%
間投詞	0/ 0.0%	0/ 0.0%	61/ 1.5%	101/ 3.6%
その他	281/ 4.3%	77/ 4.2%	142/ 3.4%	59/ 2.1%

機能語の中で、論文と口頭発表の相違が一番著しい品詞は、代名詞である。代名詞は、論文より口頭発表に多く用いられている。特に、代名詞に含まれる人称

¹¹⁾ 詳細は梅咲(1992)。竹蓋ら(1990)の分類による基本動詞を、論文よりも口頭発表のほうが多く用いていることが解った。

代名詞（所有格を除く）は、論文では、イギリス人2に *she* が3回、*we* が1回使用されているだけで、イギリス人1には全く現れていない。他方、口頭発表では、イギリス人1が190回、イギリス人2が124回使用している。これは、書き言葉では客観性がより求められるのに対し、口頭発表では目の前にいる人との対人関係が重視されるため、聞き手や自分をさす *you* や *I* が多く用いられているためと思われる。

6.1.4 異なり語の割合

論文と口頭発表でどちらが豊富に様々な語を用いているかを調べると、表5のように、ふたりとも論文のほうが口頭発表よりもさまざまな語が用いられていることが判る。

表5：異なり語の割合

書き言葉		話し言葉	
イギリス人1	52.3	イギリス人1	47.5
イギリス人2	57.8	イギリス人2	47.5
LOB 平均	53.3	LLC 平均	47.3
LOB 最大値	56.5	LLC 最大値	49.7
LOB 最小値	47.8	LLC 最小値	44.9
(95%信頼区間： $\mu \pm 0.18$)			

得られた値を、LOB コーパスと London-Lund コーパスの平均値と比べると、口頭発表は、ふたりともほぼ平均値に等しい。論文については、イギリス人2は平均値よりもかなり高い値を示しており、イギリス人1の値は、平均よりやや低い。論文と口頭発表の4つのテキストを異なり語の使用について連続線上に並べると、口頭発表は両方とも平均的な話し言葉であり、イギリス人2の論文は、書き言葉の度合が相当高く、イギリス人1の論文は書き言葉の度合がやや低いことになる。同じように語彙密度について見れば、イギリス人1のほうがイギリス人2よりも書き言葉の度合が高く、異なり語数の割合とは違った結果を得ることになる。

6.2 ミクロ分析

代表例を比べると、例1、2のペアでは、名詞、形容詞が書き言葉に多く用いられ、その為、語彙密度が話し言葉の場合より相当高くなることが解る。また、(1a)(2a)で使われている動詞は全て *be, have, do* であり、これらを機能語として扱えば、口頭発表の語彙密度はさらに低くなる。人称代名詞 *I, we, you* は、1、2及び次の3、4の例から解るように、話し言葉に多く用いられている。

- (1a) [話し言葉] Now, if we are doing a neutron diffraction experiment, first of all, we have in fact a choice. There are two sorts of diffraction experiment we can do. [語数:28, 内容語:12(名詞8, 形容詞0, 動詞4), 密度:42.9%]
- (1b) [書き言葉] Two distinct types of neutron diffraction experiment may be performed on an amorphous solid, viz. total diffraction and elastic diffraction. [語数:20, 内容語:12(名7, 形4, 動1), 密度:60.0%] (イギリス人1)
- (2a) [話し言葉] I can do a third measurement, in which I have the glass with 50% of these two isotopes. [語数:18, 内容語:5(名3, 形0, 動2), 密度27.7%]
- (2b) [書き言葉] Glasses containing almost 50% CaO have been examined by neutron scattering with substitution of Ca. [語数:15, 内容語:8(名6, 形0, 動2), 密度53.3%] (イギリス人2)

3、4のペアは、話し言葉と書き言葉の相違について新たな示唆を与えている。話し言葉では、言い直しや繰り返しが起こる。

- (3a) [話し言葉] Those are the sort of parameters we are interested in. What we actually get from a diffraction experiment, we don't get that, of course, what we get in the diffraction experiment is a series of correlation functions.... [語数:37, 内容語:17, 密度:45.9%]
- (3b) [書き言葉] In general, the parameters which characterise the order in ranges I to III are not those which are obtained directly from experiment, diffraction or otherwise. The primary real space result from a diffraction experiment is a one dimensional correlation function. [語数:40, 内容語:20, 密度:50.0%] (イギリス人1)

- (4a) [話し言葉] The other point to note is that the form factor has got Q dependence. And for this reason these peak functions have different shapes because the Q dependence of the form factor we just divide through by the number of electrons in the atom, the Q... (1)this is just vitreous silica, the Q dependence of the form factors is different for different elements. And so when we calculate the peak function we broaden the experimental data. Here they are for vitreous silica. I'll show you this one and then move it out.... What we find is that they actually have different shapes for the different, various different ..er.. pair correlation functions. [語数:111, 内容語:52, 密度:46.8%]
- (4b) [書き言葉] Since the Q dependence of the form factor $F(Q)$ varies for each element, the resulting component peak functions [EQUATION(8)] have a different shape for each component as illustrated for vitreous silica in Fig 2. [語数:33, 内容語:17, 密度:51.5%] (イギリス人1)

3の例では話し言葉に繰り返しが多いために、ほぼ同じ語数の書き言葉のほうがはるかに多い情報量を伝えている。4の例では、書き言葉が伝えている内容とほぼ同じ内容を伝えるのに、話し言葉では多くの語数を使用している。(4a)では *different* が何度も繰り返されている。また、下線部(1)のような挿入のために、本来言おうとしていたことが中断され、あとで言い直されている。

今回のマクロ分析で行なった語彙密度の計算方法では、この繰り返しや言い直しの多さが反映されていない。そのために、6.1.4で書き言葉の度合を考えた際、語彙密度と異なり語の割合の間に、不一致があったのではないと思われる。

7 結論

語彙密度、異なり語数の割合のどちらも口頭発表より論文のほうが高かった。品詞別には、名詞、形容詞（特に限定用法）が論文に、動詞、代名詞が口頭発表に多く用いられていた。ミクロ分析からも名詞、形容詞が語彙密度を上げるのに大きく貢献することが解った。このように、Hallidayの語彙密度の分析は、話し言葉と書き言葉を対比させて特徴を見いだすには非常に有効である。ただ、テキストが典型的な書き言葉と典型的な話し言葉を両端にもつ連続線のどこに位置するかを決める指数として語彙密度が有効かどうかは今後さらに研究を要する。ま

た、書き言葉と話し言葉におけるこれらの現象の背後にあるメカニズムの解明も今後の課題である。

参考文献

- Biber, D., *Variation across Speech and Writing* (Cambridge University Press, 1988).
- Halliday, M. A. K., *Spoken and Written Language* (Oxford University Press, 1985).
- Halliday, M. A. K. & Hasan, R., *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective* (Deakin University, 1985).
- Leech, G., Deuchar, M. & Hoogenraad, R., *English Grammar for Today* (Macmillan, 1982).
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J., *A Comprehensive Grammar of the English Language* (Longman, 1985).
- 竹蓋 幸生、中條 清美（日本時事英語学会 『時事英語学研究』 24、1990年、pp. 63-77）
- Umesaki F. A., “A Comparison of Scientific Lectures and Papers with regard to Grammar and Lexis” (近代英語協会『近代英語研究』 9、1992年掲載予定)