

フランス語初級学習者の発話における使用語彙の特徴

杉山香織

Kaori SUGIYAMA

0. はじめに

コーパスに基づいた研究は、言語学の分野において益々重要な位置を占めるようになってきた。しかし、その研究の多くは英語学を中心に発展してきており、特に言語教育の分野では英語以外の言語に応用されるようになってからまだ日が浅く、研究の数もまだ少ない。

本論文ではコーパスに基づき、日本人フランス語学習者の話しことばにおける語彙の使用について、母語話者の使用語彙と比較し分析する。母語話者と比較するのは、母語話者の使用語彙が学習者の学習言語の規範であり、言語習得の最終目標であると考えためである。比較を通して、学習者の使用語彙が規範と異なるのか、また異なるとすればどのように異なるのかを分析する。まず、『パリ話しことばコーパス』、『エックス話しことばコーパス』、『学習者話しことばコーパス』の3つのコーパスに含まれる話者の使用語彙を基に、この3グループを再分類化する。次に、グループ間における使用語彙の差を分析する。

1. 先行研究

コーパスの登場は、言語学の分野において大きな影響を与えた。特に第二言語習得の分野では、学習者コーパスの登場によって、学習者言語の研究手法も変化した。学習者コーパスとは、学習者が発したテキストを体系的に集積したものである(Nesselhauf 2004: 127)。コーパス登場以前は、外国語学習者の研究といえば主に実験データ(多肢選択式タスクや文法判断タスク)に基づくものであったのに対し、学習者コーパスによって、実際の産出データ(または、それに近いデータ)を分析できるようになった(Nesselhauf 2004:130-31)。また、コンピュータ上の多くの自然発話テキストに簡単にアクセスできるようになったことで、内省に基づいてきた言語記述の正確さを検証することが可能になり、言語構造や言語使用に関する新しい考えを得ることができるようになった(Tsui 2004 :39-40)。

現在学習者コーパスには、第二外国語教育分野の研究者から大きな関心が寄せられ、第二言語学習者の発話における熟練度を研究する際に、学習者コーパスの担う役割は増えつつある(Tono 2007: 163)。コーパスに基づくアプローチの利点の1つは、数量化を簡単に行うことができることである(McEneaney, Xiao and Tono 2006: 80)。そのため、学習者の産出言語における語彙の頻度情報を使用した語彙研究が中心に行われている。学習者の使用語彙に関する研究として、学習者による語、語句、文法項目、統語構造の過剰使用、過少使用に関する研究が多く行われており、エラーではない中間言語の外国語らしさについての分析結果が発表されている(Granger 1999 :132)。学習者の発話時における過少使用

／過剰使用語彙の研究は、英語学習者を中心に行われており、研究成果は辞書編纂や教材開発などに応用されている (cf. De Cook and Granger 2004, 石川 2008)。しかし、フランス語学習者の話しことばにおけるこれらの研究はあまり行われておらず、日本人フランス語学習者に関する研究は皆無と言ってよい。

2. リサーチエスチョン

本研究では、以下の2点について明らかにしていく：

1. 学習者の発話における使用語彙は母語話者と異なるか。
2. 異なるとすればどのように異なるのか。

母語話者グループと学習者グループ間の使用語彙の差は直感的また経験的に異なるだろうと予測できるが、本研究では従来の質的分析に加えて量的分析も行い、検証する。

3. 調査方法

母語話者話しことばコーパスと学習者話しことばコーパスを用いて、使用語彙を比較する。母語話者コーパスとして、自然会話コーパスである『フランス語(パリ)話しことばコーパス(以下パリコーパス)』と『フランス語(エクス)話しことばコーパス(以下エクスコーパス)』を使用した。¹⁾ 主題は、大学生活について、また好きな音楽についてなど身の回りの生活についてである。『学習者コーパス』は、フランス語学習歴2年目のフランス語専攻の大学生39名による発話を集めたものである。ヨーロッパ共通参照枠で定められたA1レベル(初級者レベル)のスピーキングタスクを教材から選択し、学習者が54タスクを行い、それを録音した。²⁾

タスク例

普段、どのような映画を見ますか？

今日のお昼に何を食べましたか？

レストランのメニューにあなたの知らない料理がありました。どのようにウエーターに質問しますか？

あなたは今、カフェにいます。友達のアピアンに、何を頼むか聞いてください。

まず、AntConc 3.2.1 (Anthony 2007)というコンコーダンス・ソフトウェア⁴⁾を使用し、『パリコーパス』、『エクスコーパス』、『学習者コーパス』の各コーパスの使用語彙リストを作成する。次に、『パリコーパス』の語彙リストの頻度を基準に各コーパスの使用語彙頻度を比較し、クラスタ分析を行う。クラスタ分析によって3グループを再分類化した後、k-means法を使用してグループ間における語彙の使用頻度差を分析する。最後に、学習者による過少使用語について分析を行う。

4. 調査結果

『パリコーパス』、『エックスコーパス』、『学習者コーパス』の3つのコーパス内で、異なり語数と総語数の割合に差があるのかを調べるためにカイ二乗検定を行う。その後、3グループを再分類化するために、クラスタ分析を行う。最後に、k-means法によって、グループ間における各単語の使用割合の差を明らかにする。

4. 1. 各コーパスの概要

3コーパスそれぞれの異なり語数と総語数は、以下の通りであった(表1参照)。

表 1. 各コーパスの異なり語数と総語数

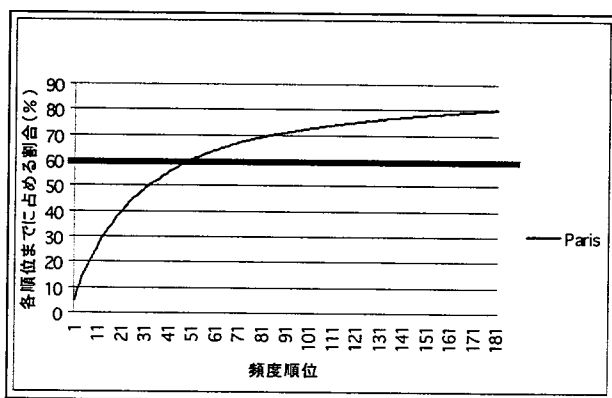
	異なり語数	総語数
パリ	2505	31796
エックス	2825	30956
学習者	2096	28393

この3グループでは異なり語数と総語数の割合に差があるといえる($\chi^2_{\omega}=54.958, p<.001$)。次に、この3グループをさらに下位分類するため、クラスタ分析を行った。

4. 2. クラスタ分析

本研究では『パリコーパス』を基準とし、『パリコーパス』における頻度の順位と総語数に占める各語数の割合とを累積比率で換算した(図1参照)。

図 1. 『パリコーパス』における語彙の累積比率



累積比率の割合の変化が、総語数における割合のおおよそ60%あたりまでに急激に上昇し、以降は緩やかに上昇していることから、全体の使用語彙の60%あたりまでに発話の核となる語が含まれると考える。そこで、『パリコーパス』の使用語彙の60%を占める語彙50語を基準に、3コーパスを比較

した。クラスタ分析の結果、『パリコーパス』と『エックスコーパス』の2つは同グループと再分類でき、母語話者グループと学習者グループの2グループへと再分類化できた。

4. 3. k-means 法

各語の総語数に対する割合を、母語話者グループ間で平均し、学習者の割合と比較したものが以下の結果である(図2参照)。

図2. k-means 法

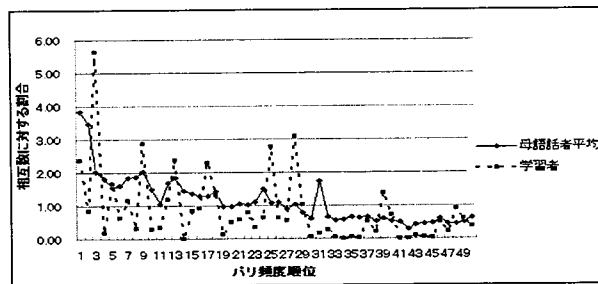


図3より、2グループ間の使用差が大きい語彙がいくつかあることが分かる。その中でも、学習者グループの使用割合が多い語と、母語話者グループの使用割合が多い語の2種類あることが分かった。本分析では、学習者グループと母語話者グループの単語使用割合の差が±1.5%のものをそれぞれ過剰使用語、過少使用語とした(表2参照)。

表2. 学習者による過剰使用語

		学習者	母語話者平均	差
過剰使用語	je	5.64	2.02	3.62
	j	2.76	1.07	1.69
	à	3.08	1.03	2.05
過少使用語	c	0.84	3.46	-2.62
	oui	0.19	1.81	-1.62
	ça	0.30	1.87	-1.57
	eah	0.17	1.73	-1.56

(%)

過剰使用はコーパスの特性に左右されやすいが、過少使用の分析はコーパスの影響を受けにくいという指摘がある(Shirato and Stapleton 2007:399)。今回分析に用いた学習者コーパス構築に用いたタスクは、母語話者コーパスのタスクと完全に同一なものでないため、過少使用のみに焦点を当てることにする。

5. 分析

学習者によって過少使用される4語のうち、指示代名詞である「c」と「ça」の2語が含まれていた(表2参照)。本稿では過少使用語彙の分析としてこれらの指示代名詞の使用について焦点を当て、詳細に見ていくことにする。

5. 1. «c (ce)»の過少使用

«c»は単独で用いられることはなく、「c'est」や«c'était»のようにêtre動詞と共に共起する。être動詞の活用形である«est»は、母語話者によって最も用いられた単語であるため、「c'est」という語連鎖について分析する。「c'est」の出現回数は、母語話者コーパス平均994.5回に対して、学習者コーパスでは216回であった。

次に«c'est」と共に用いられる語について分析するため、3語の連続であるトライグラムを抽出し各コーパス間で比較したところ、学習者コーパスに含まれるこのトライグラムの数と母語話者コーパスに含まれる数の差が大きかった。学習者コーパス内に現れるこのトライグラムのうち、もっとも出現回数の多かった55回を基準に、各グループのコーパスに55回以上出現するトライグラムを抽出したところ以下の様になった。また、「c'est」の総出現回数に対する各トライグラムの出現回数の割合も算出した(表3参照)。

表3. «c'est»を含む単語トライグラム

学習者コーパス上位トライグラム				母語話者コーパス上位トライグラム					
トライグラム	母語話者平均(回)	割合(%)	学習者(回)	割合(%)	トライグラム	母語話者平均(回)	割合(%)	学習者(回)	割合(%)
que c'est	43.5	4.37	55	25.46	c'est pas	82.5	8.30	1	0.46
					c'est un	67.5	6.79	6	2.78
					c'est c	62.5	6.28	6	2.78
					est c'est	60	6.03	6	2.78
					mais c'est	59.5	5.98	2	0.93
					ça c'est	56	5.63	0	0
					oui c'est	50.5	5.08	1	0.46

以上の表より、母語話者グループと学習者グループでは«c'est」の使用頻度が異なるだけでなく、「c'est」を含む表現の方法が異なることが分かる。また、母語話者グループでは«c'est」を含む表現の種類が偏らずバラエティーも多いことが使用の割合から分かる。

母語話者が使用するトライグラムは、学習者によってあまり使用されていない。たとえば、母語話

者によって最も使用されていた「c'est pas」という表現は、学習者によって1回しか用いられていない。この表現は「ne」の脱落と関連する。「c'est pas」を「ne」を用いて言い換えると「ce n'est pas」となるが、この表現は母語話者グループでは平均1回のみしか使用されていない。一方、学習者グループでは10回出現した(表4参照)。

表4. 「c'est pas」と「ce n'est pas」の使用頻度差

	母語話者平均(回)	学習者(回)
c'est pas	82.5	1
ce n'est pas	1	10

母語話者はこの表現を用いる際、ほぼ「ne」を脱落させるといえる。反対に、学習者は「ne」を脱落させず使用している。この点については、教科書では「ne」の脱落があまり言及されていないことや、学習者の母語話者との接触の少なさから「ne」の脱落をあまり耳にしたことがないことなどによる影響が考えられるが、更なる原因究明が必要であろう。

一方学習者は、「que c'est」というトライグラムを最も多く使用していた。この表現は、全体の「c'est」の使用の約4分の1を占めている。したがって、学習者は母語話者に比べて「c'est」の出現回数自体が少ないことに加え、「c'est」を使用する場合に共起する語が限定的であるといえる。また、各グループの「que c'est」の使用の差についてさらに分析するため、このトライグラムにおける「que」を品詞別に分類した(表5参照)。

表5. 「que c'est」における「que」の品詞

	学習者コーパス		母語話者コーパス	
	学習者(回)	割合(%)	母語話者平均(回)	割合(%)
従位接続詞	2	3.64	21.5	49.43
接続詞の一部 (ex. parce que)	9	16.36	14	32.18
疑問代名詞	25	45.45	2.5	5.75
関係代名詞	0	0.00	5.5	12.64

以上より、母語話者は「que c'est」というトライグラムを用いる際、従属節を導く従位接続詞としての「que」を用いることが多いことが分かった。

例

je te jure que c est du bruit

non non non je crois que c est R&B

c est vrai que c est des choses qui sont super intéressantes

一方、学習者は疑問代名詞の一部として、すなわち«qu'est-ce que c'est»(これは何ですか)という表現の一部として«que c'est»を用いていた。「qu'est-ce que c'est」の出現回数の「que c'est」の出現総数に対する割合は、学習者の場合が 45.45%だったのに対して、母語話者はわずか 5.75%であった(表 5 参照)。

この疑問文«qu'est-ce que c'est」は、「c'est quoi」と同様の意味を持つ。この 2 表現の出現回数について比較したところ、以下のようになった(表 6 参照)。

表 6. « qu'est-ce que c'est »と« c'est quoi »の比較

	母語話者平均 (回)	学習者 (回)
qu'est-ce que c'est	2.5	25
c'est quoi	19.5	0

フランス語の口語における部分疑問文では、疑問詞の後に« est-ce que »を置く型と、イントネーション型を使用することが多い(Grevisse 1986 : 626)。特にフランス語の砕けた会話では、平叙文で疑問詞に対応する語が現れる位置に疑問詞を用いる (Grevisse 1986 : 656)。すなわち、イントネーション型である« c'est quoi »がそれに相当する。倒置疑問文である«qu'est-ce que c'est」よりも、イントネーション型の«c'est quoi」の方が砕けた表現方法であることにより、母語話者は『これは何ですか』を意味するこの疑問文を使用する際、カジュアルなスタイルの方を好むと言える。

5. 2. « ça »の過少使用

学習者は、もう一つの指示代名詞である« ça »も過少使用している (母語話者平均 588 回 ; 学習者 86 回)。使用方法にどのような違いがあるのかを分析するため、「ça」を含んだ 2 語連続であるバイグラムを抽出した。すると、学習者コーパス内に見られる« ça »を含むバイグラムの数が少なかったため、学習者によるこのバイグラムのうち最も出現回数が多かった« ça fait »の 34 回を基準とし、両コーパスで 34 回以上出現したバイグラムを以下に示した(表 7 参照)。

表 7. « ça »を含むバイグラム

学習者コーパス上位バイグラム				母語話者コーパス上位バイグラム					
バイグラム	母語話者 平均 (回)	割合 (%)	学習者 (回)	割合 (%)	バイグラム	母語話者 平均 (回)	割合(%)	学習者 (回)	割合(%)
ça fait	31	5.27	34	39.53	comme ça	62	10.54	0	0
					ça c	61	10.37	0	0

«ça»を含むバイグラムも他の場合と同様に、母語話者グループと学習者グループで種類が異なっている。また、学習者による«ça fait»の使用割合が約40%と高いことから、学習者は«ça»という単語自体の使用頻度が少ないのに加えて、「ça」と共起する語が«fait»という単語に限定されていると言える。これらは、「c'est」の過少使用で見られた傾向と酷似している。

一方、母語話者によって用いられたバイグラムである«comme ça」と«ça c'est)»は、学習者グループによって一度も使用されていない。コーパス内で一度も出現しない語に関しては、学習者がこれらの表現を習得しているのか否かを明言するのは不可能である。この点に関しては、追加研究が必要となるだろう。

6. 結論と今後の課題

本論文では、学習者コーパスと、母語話者コーパスである『パリコーパス』と『エックスコーパス』を比較し、学習者と母語話者間における話しことばの使用語彙の差について分析した。それぞれのコーパスに含まれる語彙の頻度情報を基にクラスタ分析を行った結果、学習者の使用語彙は母語話者の使用語彙とは異なることが明らかになった。また、学習者グループと母語話者グループの使用語彙の差をk-means法を使用して比較したところ、学習者による過剰使用語と過少使用語があることが分かった。さらに、過少使用語に関して、学習者と母語話者では使用頻度差に加えて、共起する語も異なることが明確になった。学習者による過少使用語には指示代名詞である«c(ce)»と«ça»が含まれており、また、学習者がこれらの過少使用語と共に用いる語は限定的であった。

今後の課題として、学習者の使用語彙をさらに綿密に分析するためには、使用語彙の単なる差だけでなく、語彙の豊かさや使用語彙の難易度などの指標を使って、学習者と母語話者とではどのように使用語彙が異なるのかを明らかにする必要がある。また、分析結果を教室活動や教材に還元するため、その方法や効果なども研究していくことが求められる。

註

1) 東京外国語大学大学院 21 世紀 COE 「言語運用を基盤とする言語情報学拠点」の研究成果として公開されたコーパスである。それぞれ詳細は、パリコーパス(http://www.coelang.tufs.ac.jp/multilingual_corpus/fr2/index.html?menulang=ja)、エックスコーパス(http://www.coelang.tufs.ac.jp/multilingual_corpus/fr/index.html?menulang=ja)を参照のこと。

2) 構築方法およびコーパスの内容の詳細な説明については、杉山香織(2010)「日本人フランス語初級学習者スピーキングコーパスの構築」、『ふらんぼー』第35号、東京外国語大学フランス語研究室, pp.19-33.を参照のこと。

4) フリーソフトウェア。 <http://www.antlab.sci.waseda.ac.jp/software.html> からダウンロード可。

参考文献

- De Cook, S. and Granger, S. (2004) "Computer Learner Corpora and Monolingual Learners' Dictionaries: the Perfect Match". In W. Teubert. and M. Mahlberg (eds.) *The Corpus Approach to Lexicography. Special issue of Lexicographica 20*. pp. 72-86.
- Granger, S. (1999) "Computer Learner Corpus Research: Current Status and Future Prospects". In Connor, U. and Upton, T. A. (Eds.) *Language and Computers, Applied Corpus Linguistics -A Multidimensional Perspective-*. Rodopi. pp.123-145
- Grevisse, M. (1986) *Le Bon Usage -Grammaire française-*. 12^e édition. Duculot.
- McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-Based Language Studies: an Advanced Resource Book*. Routledge.
- Nesselhauf, N. (2004) "Learner Corpora and Their Potential for Language Teaching". In J. Sinclair (ed.) *How to Use Corpora in Language Teaching*. John Benjamins. pp.125-152.
- Shirato, J. and Stapleton, P. (2007) "Comparing English vocabulary in a spoken learner corpus with a native speaker corpus: Pedagogical implications arising from empirical study in Japan". *Language Teaching Research 11(4)*. pp.393-412.
- Tono, Y. (2007) "The roles of Oral L2 Learner Corpora in Language Teaching: the Case of the NICTJLE Corpus". In M.C.Copoy. and M.J. Luzón (eds.). *Spoken Corpora in Applied Linguistics*. Peter Lang. pp.163-179.
- Tsui, A. B. M. (2004) "What Teachers Have Always Wanted to Know — and How Corpora can Help". In J. Sinclair (ed.) *How to Use Corpora in Language Teaching*. John Benjamins. pp.39-61.
- 石川慎一郎 (2008) 『英語コーパスと言語教育—データとしてのテキスト』大修館書店
- Council of Europe (2008) 『外国語の学習、教授、評価のためのヨーロッパ共通参照枠』, 初版第 2 刷, 吉島茂、大橋理枝(訳、編) 朝日出版社.

参考資料：『パリコーパス』における上位 50 語と各語の総語数に対する割合

頻度順 位		パリ (%)	エックス (%)	学習者 (%)	頻度順 位		パリ (%)	エックス (%)	学習者 (%)
1	est	4.87	2.80	2.36	26	en	0.92	1.30	0.64
2	c	4.38	2.55	0.84	27	d	0.91	0.90	0.55
3	je	2.18	1.85	5.64	28	à	0.83	1.23	3.08
4	oui	2.18	1.43	0.19	29	l	0.79	0.80	1.02
5	la	2.03	1.08	1.67	30	tout	0.78	0.42	0.05
6	mais	2.03	1.18	0.63	31	euh	0.76	2.70	0.17
7	pas	1.98	1.71	1.17	32	qui	0.67	0.66	0.26
8	ça	1.94	1.80	0.30	33	sais	0.65	0.43	0.05
9	de	1.80	2.26	2.86	34	quoi	0.65	0.49	0.01
10	mmm	1.75	1.27	0.29	35	là	0.64	0.67	0.05
11	non	1.64	0.49	0.35	36	ils	0.63	0.63	0.03
12	il	1.63	1.76	1.19	37	une	0.63	0.68	0.51
13	et	1.55	2.13	2.38	38	dans	0.60	0.42	0.21
14	ouais	1.49	1.42	0.00	39	elle	0.60	0.62	1.37
15	un	1.36	1.37	0.85	40	moi	0.60	0.43	0.72
16	a	1.32	1.23	0.91	41	hein	0.57	0.42	0.00
17	le	1.24	1.33	2.30	42	heu	0.57	0.00	0.00
18	que	1.22	1.61	1.31	43	comme	0.54	0.30	0.08
19	ah	1.17	0.78	0.13	44	accord	0.53	0.38	0.04
20	y	1.11	0.85	0.49	45	même	0.51	0.45	0.02
21	les	1.10	1.00	0.58	46	qu	0.50	0.72	0.49
22	des	1.08	0.96	0.79	47	bien	0.48	0.40	0.20
23	on	1.07	1.14	0.34	48	du	0.47	0.41	0.89
24	tu	1.05	1.94	0.62	49	pour	0.46	0.56	0.61
25	j	0.95	1.19	2.76	50	fait	0.46	0.78	0.38