

博士学位論文（東京外国語大学）  
Doctoral Thesis (Tokyo University of Foreign Studies)

氏名	杉山 香織
学位の種類	博士（学術）
学位記番号	博甲第 175 号
学位授与の日付	2013 年 11 月 6 日
学位授与大学	東京外国語大学
博士学位論文題目	フランス語学習者の発話における使用語彙分析

Name	Sugiyama, Kaori
Name of Degree	Doctor of Philosophy (Humanities)
Degree Number	Ko-no. 175
Date	November 6, 2013
Grantor	Tokyo University of Foreign Studies, JAPAN
Title of Doctoral Thesis	Lexical Analysis on Learners' Oral Production of French

博士学位論文

フランス語学習者の発話における使用語彙分析

杉山香織

東京外国語大学大学院博士後期課程

地域文化研究科博士論文

フランス語学習者の発話における使用語彙分析

杉山香織

## 目次

序章.....	1
1. 本論文の目的.....	1
2. 本論文の構成.....	1
第一章 コーパス言語学.....	4
1.1. 言語学におけるコーパスとは何か.....	4
1.1.1. 言語コーパスの位置づけ.....	4
1.1.2. コーパスと代表性.....	6
1.2. コーパス言語学の歴史.....	8
1.2.1. チョムスキーとコーパス言語学.....	8
1.2.2. 電子コーパス以前からコーパスの普及まで.....	10
1.2.3. コーパス基盤型とコーパス駆動型.....	13
1.2.4. 近年のコーパス.....	14
1.2.5. 話し言葉コーパス.....	17
1.3. コーパス分析ツール.....	26
1.4. 学習者コーパスとプロジェクト.....	27
1.4.1. コーパス登場以前の間言言語研究.....	28
1.4.2. 学習者コーパスの枠組み.....	31
1.4.3. 代表的な学習者コーパス.....	32
1.4.4. 話し言葉の特徴と評価.....	35
1.4.5. 学習者コーパスの利点と欠点.....	37
1.4.6. フランス語学習者話し言葉コーパスプロジェクト.....	38
1.4.6.1. French Learner Language Oral Corpora.....	38
1.4.6.2. Progression in Foreign Language Learning.....	39
1.4.6.3. Langue et communication (LANCOM).....	40
1.4.6.4. InterPhonologie du Français Contemporain (IPFC).....	41
1.5. コーパスの利点と欠点.....	42
1.6. まとめ.....	43
第二章 語レベルの研究 –Word Units 研究–.....	45
2.1. 『一語』とは何か.....	45
2.2. 語彙リスト開発の歴史.....	46
2.3. 語彙と第二言語の能力.....	47
2.4. 語彙能力とはなにか.....	49
2.4.1. 受容語彙と産出語彙.....	49
2.4.2. 語彙の深さと広さ.....	49
2.5. 伝統的な語彙の測定.....	50



2.5.1.	測定対象	50
2.5.2.	テスト形式	51
2.6.	語彙の豊かさとその測定法	52
2.7.	語彙の豊かさの問題点	54
2.8.	頻度情報を使用した語彙の測定法	56
2.8.1.	<i>Lexical Frequency Profile</i>	58
2.8.2.	<i>VocabProfil</i>	59
2.8.3.	<i>LFP</i> への反論	60
2.9.	過剰使用と過少使用	62
2.10.	学習者の語彙研究	63
2.10.1.	フランス語学習者の語彙研究	64
2.10.1.1.	Debrock et al. (1999)	64
2.10.1.2.	David (2008)	65
2.10.1.3.	Treffers-Daller (2010)	66
2.10.1.4.	Bulté et al. (2008)	67
2.10.2.	<i>VocabProfil</i> に基づく研究	68
2.10.2.1.	Goodfellow et al. (2002)	68
2.10.2.2.	Ovtcharov et al. (2006)	69
2.10.2.3.	Thomas (2008)	69
2.10.2.3.	Sugiyama (2012)	70
2.11.	まとめ	71
<b>第三章</b>	<b>Word Units から Multi-Word Units へ</b>	<b>72</b>
3.1.	単語以上の単位 - Multi-Word Units -	73
3.2.	フレイジオロジーの歴史的背景	76
3.2.1.	伝統的アプローチ	76
3.2.2.	頻度アプローチ	78
3.2.3.	二つのアプローチの融合	80
3.3.	MWUs の測定法	80
3.3.1.	コロケーション分析	80
3.3.2.	N-grams 分析	81
3.4.	中間言語とフレイジオロジー	81
3.4.1.	理論的背景	81
3.4.2.	中間言語データに基づく MWUs の分析	84
3.4.2.1.	De Cock et al. (1998)	84
3.4.2.2.	De Cock (2004)	85
3.4.2.3.	De Cock (2007)	85
3.4.2.4.	Myles et al. (1998)	86
3.4.2.5.	Myles et al. (1999)	86
3.5.	まとめ	87

<b>第四章 研究方法</b> .....	<b>88</b>
<b>4.1. 研究目的と研究背景</b> .....	<b>88</b>
<b>4.2. コーパスデータ</b> .....	<b>88</b>
4.2.1. タスクに基づくコーパス(学習者コーパス).....	89
4.2.1.1. 対象者.....	89
4.2.1.2. タスク.....	90
4.2.1.3. 録音システム.....	93
4.2.1.4. 転写.....	94
4.2.2. タスクに基づくコーパス(母語話者コーパス).....	95
4.2.2.1. 対象者.....	95
4.2.2.2. タスク.....	97
4.2.2.3. 録音システム.....	99
4.2.2.4. 転写.....	99
4.2.3. 自由会話コーパス( <i>IPFC-JP</i> コーパス).....	100
4.2.3.1. 対象者.....	100
4.2.3.2. 会話の主題.....	100
4.2.3.3. 録音システム.....	101
4.2.3.4. 転写.....	101
4.2.4. 自由会話コーパス(フランス語(エックス)多言語話しことばフランス語コーパス).....	101
4.2.4.1. 対象者.....	101
4.2.4.2. 会話の主題.....	101
4.2.4.3. 録音システム.....	102
4.2.4.4. 転写.....	102
<b>4.3. 使用語彙分析の構造</b> .....	<b>102</b>
<b>4.4. リサーチクエスション</b> .....	<b>103</b>
4.4.1. 語彙の豊かさ.....	103
4.4.2. 特徴語.....	103
4.4.3. コロケーション分析.....	104
4.4.4. N-grams 分析.....	104
<b>4.5. 分析方法</b> .....	<b>105</b>
4.5.1. 使用語彙研究の分析方法.....	105
4.5.1.1. 語彙の豊かさの分析.....	105
4.5.1.2. 特徴語の分析.....	106
4.5.1.3. コロケーション分析.....	106
4.5.1.4. N-gram 分析.....	107
<b>第五章 語彙の豊かさの分析</b> .....	<b>108</b>
<b>5.1. タスクに基づくコーパス</b> .....	<b>108</b>
5.1.1. 記述統計量.....	108

5.1.2.	相関関係 .....	111
5.1.3.	主成分分析.....	112
<b>5.2.</b>	<b>自由会話コーパス .....</b>	<b>115</b>
5.2.1.	記述統計量.....	115
5.2.2.	相関関係 .....	118
5.2.3.	主成分分析.....	119
<b>5.3.</b>	<b>まとめ .....</b>	<b>124</b>
<b>第六章</b>	<b>特徴語分析 .....</b>	<b>125</b>
<b>6.1.</b>	<b>タスクに基づくコーパス .....</b>	<b>125</b>
6.1.1.	過剰使用語.....	126
6.1.2.	過少使用語.....	128
<b>6.2.</b>	<b>自由会話コーパス .....</b>	<b>131</b>
6.2.1.	過剰使用語.....	131
6.2.2.	過少使用語.....	134
<b>6.3.</b>	<b>まとめ .....</b>	<b>138</b>
<b>第七章</b>	<b>コロケーション分析 .....</b>	<b>140</b>
<b>7.1.</b>	<b>タスクに基づくコーパス .....</b>	<b>140</b>
7.1.1.	過剰使用語との共起語 .....	140
7.1.1.1.	<i>voudrais</i> .....	140
7.1.1.2.	<i>elle</i> .....	144
7.1.1.3.	<i>NP</i> .....	148
7.1.1.4.	<i>ne</i> .....	154
7.1.1.5.	<i>je</i> .....	155
7.1.1.6.	<i>prenez</i> .....	162
7.1.1.7.	<i>japonaise</i> .....	163
7.1.1.8.	<i>vous</i> .....	164
7.1.1.9.	<i>peux</i> .....	168
7.1.1.10.	<i>abord</i> .....	171
7.1.1.11.	<i>football</i> .....	172
7.1.1.12.	<i>université</i> .....	173
7.1.2.	過少使用語との共起語 .....	174
7.1.2.1.	<i>HESIT</i> .....	174
7.1.2.2.	<i>on</i> .....	177
7.1.2.3.	<i>donc</i> .....	179
7.1.2.4.	<i>peu</i> .....	180
7.1.2.5.	<i>bien</i> .....	181
7.1.2.6.	<i>c'</i> .....	182
7.1.2.7.	<i>vraiment</i> .....	184
7.1.2.8.	<i>tout</i> .....	185

7.2.	自由会話コーパス .....	186
7.2.1.	過剰使用語との共起語 .....	186
7.2.1.1.	<i>INT</i> .....	186
7.2.1.2.	<i>oui</i> .....	190
7.2.1.3.	<i>NP</i> .....	193
7.2.1.4.	<i>FW</i> .....	195
7.2.1.5.	<i>très</i> .....	196
7.2.1.6.	<i>visité</i> .....	198
7.2.1.7.	<i>beaucoup</i> .....	199
7.2.1.8.	<i>BC</i> .....	201
7.2.1.9.	<i>allée</i> .....	202
7.2.1.10.	<i>français</i> .....	203
7.2.1.11.	<i>japonais</i> .....	204
7.2.1.12.	<i>allé</i> .....	204
7.2.1.13.	<i>pendant</i> .....	205
7.2.1.14.	<i>est</i> .....	205
7.2.1.15.	<i>c'</i> .....	209
7.2.1.16.	<i>musique</i> .....	213
7.2.1.17.	<i>j'</i> .....	214
7.2.1.18.	<i>difficile</i> .....	218
7.2.1.19.	<i>je</i> .....	219
7.2.1.20.	<i>comment</i> .....	225
7.2.2.	過少使用語との共起語 .....	226
7.2.2.1.	<i>enfin</i> .....	226
7.2.2.2.	<i>hein</i> .....	228
7.2.2.3.	<i>qui</i> .....	228
7.2.2.4.	<i>quoi</i> .....	230
7.2.2.5.	<i>les</i> .....	232
7.2.2.6.	<i>ça</i> .....	234
7.2.2.7.	<i>des</i> .....	236
7.2.2.8.	<i>ils</i> .....	238
7.2.2.9.	<i>fait</i> .....	239
7.2.2.10.	<i>ouais</i> .....	241
7.2.2.11.	<i>qu'</i> .....	243
7.2.2.12.	<i>vois</i> .....	245
7.2.2.13.	<i>là</i> .....	246
7.2.2.14.	<i>bon</i> .....	247
7.2.2.15.	<i>tout</i> .....	249
7.2.2.16.	<i>avoir</i> .....	250
7.2.2.17.	<i>leur</i> .....	251

7.2.2.18. <i>même</i> .....	252
7.2.2.19. <i>sais</i> .....	254
<b>7.3. コロケーション分析から見る学習者の使用語彙特徴.....</b>	<b>256</b>
7.3.1. 人称代名詞.....	256
7.3.2. 否定辞.....	257
7.3.3. 短縮形・俗語.....	257
7.3.4. 曖昧性付加句.....	258
7.3.5. ディスコースマーカー.....	259
<b>第八章 N-grams 分析.....</b>	<b>261</b>
<b>8.1. タスクに基づくコーパス.....</b>	<b>261</b>
8.1.1. 繰り返し使用される N-grams 数の比較.....	261
8.1.2. 特徴的 N-grams と語彙分析の関連.....	270
8.1.2.1. 特徴的 N-grams と頻度層.....	270
8.1.2.2. 特徴的 N-grams と特徴語の関係.....	284
8.1.3. 特徴的 N-grams の質的分析.....	290
8.1.3.1. 特徴的 N-grams の傾向.....	290
8.1.3.2. WUs レベルが影響を与えた特徴的 N-grams.....	300
8.1.3.2.1. 人称代名詞.....	300
8.1.3.2.2. 語彙の選択.....	302
8.1.3.3. MWUs レベルが影響を与えた特徴的 N-grams.....	303
8.1.3.3.1. <i>beaucoup de</i> .....	303
8.1.3.3.2. 疑問文.....	304
8.1.3.3.3. 否定辞 <i>ne</i> の保持と脱落.....	305
8.1.3.3.4. <i>qu(e)</i> .....	306
8.1.3.3.5. ディスコースマーカー.....	307
8.1.3.3.6. エラー.....	308
<b>8.2. 自由会話コーパス.....</b>	<b>310</b>
8.2.1. 繰り返し使用される N-grams 数の比較.....	310
8.2.2. 特徴的 N-grams と語彙分析の関連.....	318
8.2.2.1. 特徴的 N-grams と頻度層.....	318
8.2.2.2. 特徴的 N-grams と特徴語の関係.....	326
8.2.3. 特徴的 N-grams の質的分析.....	331
8.2.3.1. 特徴的 N-grams の傾向.....	331
8.2.3.2. 語彙レベルが影響を与えた特徴的 N-grams.....	337
8.2.3.2.1. 語(句)の繰り返し.....	337
8.2.3.2.2. 中性代名詞 <i>en</i> .....	338
8.2.3.3. 統語レベルが影響を与えた特徴的 N-grams.....	338
8.2.3.3.1. 否定辞 <i>ne</i> の保持と脱落.....	338
8.2.3.3.2. 疑問文.....	340

8.2.3.3.3. <i>qu(e), qui</i> .....	340
8.2.3.4. 談話レベルが影響を与えた特徴的 N-grams .....	342
8.2.3.4.1. 順接関係を表す語.....	342
8.2.3.4.2. 方略的使用の N-grams.....	343
8.2.3.4.3. ディスコースマーカー .....	345
<b>第九章 結論 .....</b>	<b>347</b>
<b>9.1. 本研究のまとめ.....</b>	<b>347</b>
9.1.1. 語彙の豊かさの分析.....	347
9.1.1.1. タスクに基づくコーパス .....	347
9.1.1.2. 自由会話 コーパス.....	348
9.1.2. 特徴語分析 .....	348
9.1.2.1. タスクに基づくコーパス .....	348
9.1.2.2. 自由会話コーパス.....	349
9.1.3. コロケーション分析 .....	349
9.1.3.1. タスクに基づくコーパス .....	349
9.1.3.2. 自由会話コーパス.....	350
9.1.4. N-grams 分析.....	351
9.1.4.1. タスクに基づくコーパス .....	351
9.1.4.2. 自由会話 コーパス.....	351
9.1.5. リサーチクエスチョンに対する回答 .....	352
<b>9.2. 今後の課題 .....</b>	<b>354</b>
<b>参考文献.....</b>	<b>356</b>
<b>謝辞.....</b>	<b>370</b>

# 序章

## 1. 本論文の目的

本論文の目的は、話し言葉における日本人フランス語学習者の使用語彙について記述を行うことである。より包括的な記述を行うために、母語話者コーパスと学習者コーパスを使用して2コーパスを比較する。

実際に、日本人フランス語学習者がフランス語で会話を行う際に、どのような使用語彙の特徴があるのかについて、実証に基づいた研究はまだ行われていない。日本人学習者の話し言葉がどのような語彙に特徴付けられているのかは、教師の経験と勘に従って語られてきた。さらに、フランス語教育ではこのような教師の経験と勘に基づいて指導が行われているのが実情である。

コーパスデータに基づいた分析を行うことで、これまで教師間で主観的に共有されていた学習者の使用語彙の特徴について、その正当性を客観的に証明することができる。あるいは、これまでの認識を質的、計量的観点から否定することも可能となる。または、これまで見過ごされてきた新しい使用語彙特徴も明らかにすることができる。

本論文では、語彙は全ての言語的要素の根幹となるものであるという立場を取る。したがって、使用語彙を研究することで、語彙レベルだけではなく、統語レベルや談話レベルでの分析も行う。このように、本論文は質的観点と量的観点の双方から学習者の使用語彙特徴の記述を行い、学習者の話し言葉の特徴を明らかにするものである。

## 2. 本論文の構成

本研究ではコーパスに基づいて分析を行うため、コーパス言語学とはどういう分野であるのかについてまず第一章で論じる。「1.1. 言語学におけるコーパスとは何か」では、言語コーパスの特徴や、コーパスに必須の概念である代表性について説明を記す。「1.2. コーパス言語学の歴史」では、コーパス言語学がどのような状況下で誕生をしたのかや、どのように発展の過程を辿ったのかについて述べた上で、近年のコーパスについてまとめる。その後「1.3. コーパス分析ツール」では、コーパス分析を可能とする分析ツールについて紹介する。「1.4. 学習者コーパスとプロジェクト」では、中間言語の説明を行った後、コーパス分析によって中間言語分析にどのような変化がもたらされたのかを明らかにし、学習者コーパスとは何かについて定義を行う。そして、代表的なフランス語学習者コーパスプロジェクトについて概観する。そして、「1.5. コーパスの利点と欠点」で、コーパス研究についての注意点を再考する。そして「1.6. まとめ」で第一章のまとめを行う。

第二章では、語レベルの研究についてまとめる。まず「2.1. 『一語』とは何か」で様々な語の単位について整理する。「2.2. 語彙リスト開発の歴史」で、語彙研究が言語教育分野で発展する礎となった語彙リストの開発の歴史について俯瞰する。「2.3. 語彙と第二言語の能力」で、第二言語能力において語彙が中心的役割をすることを確認し

した後、「2.4. 語彙能力とは何か」では、受容語彙と産出語彙、語彙の深さと語彙の広さという特に 2 つの対立関係についての説明を行う。「2.5. 伝統的な語彙の測定」では、これまで行われてきた語彙研究におけるの測定対象と測定方法について整理を行う。コーパスの登場とともに注目を集めるようになった語彙の豊かさについて、「2.6. 語彙の豊かさとその測定法」で、これまで考察が行われてきた語彙の豊かさという概念の説明とその測定方法について整理を行い、「2.7. 語彙の豊かさの問題点」でその測定方法の欠点を指摘する。「2.8. 頻度情報を使用した語彙の測定法」では、英語における産出言語の語彙の測定で用いられてきた *Lexical Frequency Profile* についての説明を行い、そのフランス語版の *VocabProfil* の紹介とその問題点を提示する。「2.9. 過剰使用と過少使用」では、コーパスに基づいて行われてきた語レベルの研究の一つである過剰使用と過少使用研究についてまとめる。「2.10. フランス語学習者の語彙研究」では、語彙レベルから行われてきた学習者の使用語彙に関するの研究例と *VocabProfil* を使用した研究例を挙げる。「2.11. まとめ」では、第二章のまとめをする。

第三章では、語レベルの研究よりも大きな単位である **Multi-Word Units** についてのまとめを行う。「3.1. 単語以上の単位 -Multi-Word Units-」では、単語以上の単位として使用されてきた様々な名称について整理する。「3.2. フレイジオロジーの歴史的背景」では、**Multi-Word Units** 研究であるフレイジオロジー研究について、その伝統的アプローチと頻度アプローチの特徴についてまとめる。「3.3. MWUs の測定法」について、その主な分析であるコロケーション分析と **N-grams** 分析について測定方法がまとめられている。「3.4. 中間言語のフレイジオロジー」では、学習者が使用する **MWUs** についてその理論的背景と先行研究を概観する。「3.5. まとめ」では、第三章をまとめる。

第四章では、研究方法について述べる。「4.1. 研究目的と研究背景」では、改めて本研究の目的について確認を行う。「4.2. コーパスデータ」では、研究対象であるコーパスデータについて詳細に説明を行う。「4.3. 使用語彙分析の構造」では、本研究で行う分析の相互関係についてまとめる。「4.4. 研究背景とリサーチクエスション」では、各分析のリサーチクエスションを述べ、「4.5. 分析方法」では各分析の分析方法を説明する。

第五章では、語彙の豊かさの分析を行う。「5.1. タスクに基づくコーパス」では、記述統計、相関分析、主成分分析を行い、「5.2. 自由会話コーパス」でも同様の分析を行う。「5.3. まとめ」では第五章のまとめを行う。

第六章では、特徴語の分析を行う。「6.1. タスクに基づくコーパス」と、「6.2. 自由会話コーパス」で、それぞれのコーパスにおける過剰使用語と過少使用語の結果を述べ、それらの語の特徴について分析、考察を行う。「6.3. まとめ」では第五章のまとめを行う。

第七章では、コロケーション分析を行う。「7.1. タスクに基づくコーパス」と、「7.2. 自由会話コーパス」で、それぞれの過剰使用語と過少使用語との共起語について、分析と考察を行う。そして、「7.3. コロケーション分析から見る学習者の使用語彙特徴」では、タスクに基づくコーパスを用いたコロケーション分析と自由会話コーパスを用いた分析を通して明らかになった学習者の使用語彙特徴について、まとめを行う。



第八章では、N-grams 分析を行う。「8.1. タスクに基づくコーパス」と、「8.2. 自由会話コーパス」で、繰り返し使用される N-grams 数の比較を行う。また、過剰使用 N-grams や過少使用 N-grams と語彙の頻度層との関係、そして特徴語との関係をまとめる。最後に、特徴的 N-grams の質的分析を行う。

第九章では、第五章から第八章までに行った分析や考察を結論としてまとめる。そして、これから本研究がどのようにフランス語教育研究につながるのか、そしてどのようにフランス語教育へ還元できるのかについてを述べる。今後の展望や課題などについても言及する。

# 第一章 コーパス言語学

電子コーパスの発達は、全ての語彙研究に大きな影響を与えた。たとえば、辞書編纂者はコーパスを用いて単語の定義、共起頻度の記録を行い、単語の特徴的な使用に関する情報についてもコーパスから得ており、これらの側面に関する記述にむけてコーパスを日常的に使用している。また、コーパス分析は話し言葉や書き言葉における単語の共起や 2 語以上から構成される句の機能について新たな見解を与えた。評価に関しても、コーパスに基づいて頻度、出現レンジなどの基準を考慮した単語を抽出することが可能となり、より正確な単語リストが入手可能となった。さらに、コーパス分析は学問分野別などの特定のコンテキストにおける語彙の使用について、言語の語彙的特徴の記述にも貢献した (Read 2007: 106)。このようにコーパス言語学は、コンピュータ技術の発展とともに言語記述、自然言語処理、そして言語教育の分野など、新しいタイプの言語研究の機会を与えた (Kennedy 1998: 294)。

言語学の分野におけるコーパスの重要性は以前から指摘されてきていたが、近年の科学技術の進歩により、コーパスはますます言語学者にとって身近なものとなり、コーパスに対する興味や関心は年々高まっている。コーパスを使用した研究は、ここ十年間で飛躍的に拡大した。しかし、「コーパス言語学」という名前が一人歩きをして、コーパスとは何かを理解されないまま「コーパス」と呼ばれるものに基づき言語分析が行われている場合も少なくない。本章ではまず、一概に「コーパス」と呼ばれているものがどのようなものなのかを定義する。

## 1.1. 言語学におけるコーパスとは何か

### 1.1.1. 言語コーパスの位置づけ

コーパスという用語は、ラテン語である *corpus* に由来している。この語は、もともと「からだ」を意味しており、12 世紀の後半には *corpus Deu* 「聖体」という意味でこの語が用いられていた。19 世紀に入って、*corpus* は *corpus juris* 「法典」のような資料の総体を指すようになったが、言語分析で用いられている「コーパス」の意味での使用が始まったのは 20 世紀中旬からである (Imbs 1978 :215)。

コーパスは、言語学者のみが使用するものではない。Dalbera (2002:90-91)は、コーパスデータの基本的な特徴によってコーパスを二分化している。まず一つ目のコーパスは、あるジャンルや時代を代表するテキスト、口語証言などから構成される収集物のことである。この種のコーパスは歴史学者、文献学者、法学者などが使用するコーパスである。二つ目のコーパスは、言語学者が使用する、文、語、音素等の言語単位のコーパスである。前者のコーパスは、一般的にアーカイブとして捉えることができるだろう。では、言語学におけるコーパスと、アーカイブは何が異なるのであろうか。

言語学におけるコーパス研究の第一人者である Sinclair (1996: 4)は、以下のように後者のタイプのコーパスを定義している。

A corpus is a collection of pieces of language that are **selected and ordered according to explicit linguistic criteria** in order to be used as **a sample of the language**<sup>1</sup>.

(コーパスは、言語サンプルとして使用される目的として、明確な言語的基準に従って選択され整理されている言語の構成要素の収集物である)

つまり、言語学におけるコーパスと、その他の分野におけるコーパスが決定的に異なるのは、コーパスが言語分析の目的をもって構築されているという点である。したがって、言語学の分野におけるコーパスとその他の分野におけるコーパスとを区別することは、重要であると言えよう。以下より、特別な言及がない限り「コーパス」は言語学におけるコーパスのみを指し示すこととする。

コーパスの中にもまた、多様な区分が存在する。まずは、コーパスがあらゆる言語事象を代表するサンプルであるのか、それともある分野に特化したサンプルであるのかという区分があげられる。前者は、言語についての包括的な情報を提供するために設計されたものである。対象となる言語に関連するすべての変種や特徴となる語彙について、十分に代表できるほどの規模を目標に構築されている。この種のコーパスはしばしば汎用コーパスと呼ばれており、文法、辞書、シーソーラスや他の言語参照資料の確かな基盤となりうる (Sinclair 1996: 10)。そして、語彙、文法、談話構造における特定の質問に対する答えを探すため、言語学者が分析するテキストの総体でもある。汎用コーパスは、話し言葉と書き言葉、私的なものと公的なものといったように様々なジャンルや分野からのテキストを含んでおり、一般的にバランスよく設計されている (Kennedy 1998: 19-20)。

後者は、特殊コーパスと呼ばれる、ある特定の目的に限定して作成されたコーパスである。特殊コーパスには、たとえば、通時研究を行うために様々な時代のテキストを集めて作られたコーパスがある。実際、言語変化に関する研究のほとんどはコーパスに基づくものである。したがって、1990年代中旬においてコーパスに基づく研究が最も活発に行われた分野が英語の通時的研究であったということは、驚くべきことではない。このタイプのコーパスの例として、トロント大学によって構築された *Complete Corpus of Old English* が挙げられる。これは、3022 のテキストから構成されており、現存する古英語のテキストの全体をカバーするものである。このコーパスは 1981 年に *Dictionary of Old English* の決定版の基礎として出版され、現在は電子形式で入手可能である (Kennedy 1998: 38)。

ある観点を共時的に比較するため、コーパスを使用して英語の様々なバリエーションを比較研究したプロジェクトもある。国際英語における比較研究に関するコーパスプロジェクトで最も大掛かりなものは、1996 年に Sidney Greenbaum によって発表された *International Corpus of English (ICE)* である。1998 年に、当時 University College London の *Survey of English Usage* の研究所長であった Greenbaum は、世界中の英語の地域変種の話し言葉と書き言葉に関する比較研究を行う目的で、巨大コーパスの開発を提案した。ICE プロジェクトは、20 のサブコーパスから構成されている。イギリス、アメリカ、カナダ、オーストラリア、ニュージーランドなどの英語が支配的もしくは主要な

---

<sup>1</sup> 強調部分は筆者によるものである。

第一言語である国と、インド、ナイジェリア、シンガポールまたはカリブ海域諸島のような英語が公用語や人口の多くの第二言語である国で、少なくとも中学校まで英語で教育を受けてきた 18 歳以上の人々が対象である。それぞれのサブコーパスはこれらの英語話者が産出した 100 万語から構成されている (Kennedy 1998: 54)。

また、第一言語研究のみならず、第二言語習得研究においてもコーパスは注目を集めている。第二言語の学習者が産出したテキストからなるコーパスは一般的に学習者コーパスと呼ばれている。学習者コーパスは、外国語学習者ならび第二言語学習者によって実際に口頭形態および筆記形態によって産出された言語を体系的に集積したテキストの集合体である (Granger 2003: 538, Granger 2004: 124, Nesselhauf 2005: 40)。たとえば、Clive Perdue が 1993 年に作成したコーパスである *European Science Foundation Second Language Databank* がその例である。これは、オランダ語、英語、フランス語、ドイツ語、スイス語を学ぶパンジャブ人、スペイン人、トルコ人、フィンランド人、イタリア人、アラビア語話者のモロッコ人の移民を長期的に調査した会話を転写したものである (Kennedy 1998: 42)。その他には、特に英語学習者コーパスが多く構築されている。学習者コーパスについては、一章 1.4. で詳述する。

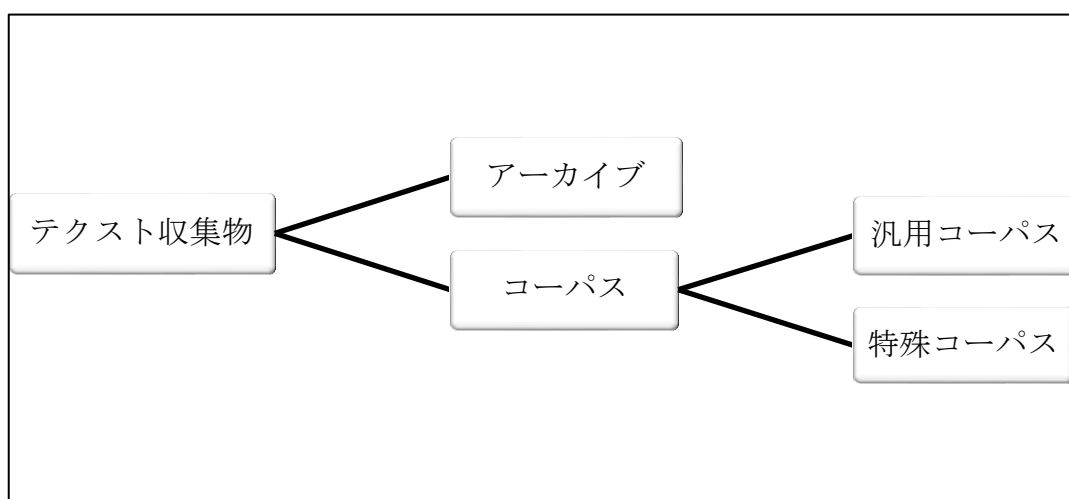


図 1 コーパスの位置づけ

### 1.1.2. コーパスと代表性

コーパスの定義として、忘れてはならないのは代表性についてである。Leech and Fligelstone (1992: 116)は、コーパスの収集方法について以下のように述べている。

They (corpora) ... are often assembled to be (informally speaking) representative of some language or text type.

(コーパスはしばしば、言語やテキストタイプにおいて代表的であるように収集されている。)

このデータの代表性は、コーパス言語学者にとって最も重要な問題の一つである (Jacques 2005: 29)。なぜなら、コーパスを用いて研究を行う場合、全ての基礎をコーパスに置くからである。研究の基盤となるデータであるコーパスが、チョムスキー派による批判にもあったような「歪んだ」ものであれば、言語の正しい様相を明らかにすることは不可能である。したがって、汎用コーパスはその性質上、あらゆる言語的観点において代表的でなければならず、代表性は特殊コーパスにも考慮されるべき点である。

ほとんどのコーパスの代表性は、以下の 2 つの要因によって大部分が決定される (McEnery et al. 2006: 13-21)。

#### 1) バランス

コーパスに含まれる分野の範囲によって、コーパスの代表性が決定される。また、目的によってバランスも異なる。汎用コーパスでは特に、対象となる言語や言語バラエティにとって代表的であると考えられるような広いテキストカテゴリーを包括することが求められ、バランスのとれたコーパスであると見なされる基準の一つである。

#### 2) サンプルング

各ジャンルのテキストをどのように選択するのかが、代表性で重要なもう一つの観点である。全ての自然言語を記述することは不可能であるため、リサーチクエスチョンに最も適合するバランスや代表性となるようにサンプルする必要がある。

しかし Dalbera (2002: 94)が指摘するように、いくらある言語や分野を代表するサンプルから成る代表性を持つコーパスでも、コーパスは実世界におけるデータの単なる部分集合なのではなく、既に加工されたサンプルであることに留意しなければならない。なぜなら、データを選択し文脈から切り離れた時点で、データに手を加えたことになるからである。また、ある集団における言語特徴を扱うコーパスの場合、代表性の問題はさらに複雑化する。たとえば、学習者コーパスにおける代表性の問題について Leech(1998 :xix)は、以下のように述べている。

The methodology associated with learner corpora is to isolate a variable in terms of which two otherwise comparable corpora differ: particularly the variable of the learners' native language. But it is very difficult to be sure that what the two corpora represent differs only with respect to this variable: for example, two English learner corpora of native language X and Y may provide evidence that differences in English language performance are due to the NLs concerned. But it is conceivable that particular effects may be due to other factors such as differences in cultural context or in educational institutions and practices.

(学習者コーパスに関連する方法論は、2つの比較可能なコーパス間に見られる異なる変数、特に学習者の母語の変数を取り出すことである。しかし、2つのコーパスが代表するものの違いが、母語の違いだけであると断定することは大変難しい。たとえば、母語 X と Y の 2つの英語学習者コーパスは、英語のパフォーマンスの違いは母語の違いによるものであるという証拠を提示するかもしれない。しかし観察効果が文化的文

脈の違いや、教育機関や教育法の違いといった他の要因に起因するということも観察されるかもしれない。) )

この代表性の問題の対処法として、コーパス分析の結果から一般的な推論を導く際には十分な注意を払い、データ収集やサンプル抽出時における隠れた変数の影響にも注意を払うべきだと Leech(ibid: xix)は続けている。

現在、コーパスに基づく研究は言語学研究の幅広い分野で行われている。したがって、音声学、統語論、意味論、語用論のような言語学の独立した一学問というよりは、方法論であると考えべきである。音声学、統語論、意味論、語用論といった分野は、言語使用のある側面について記述し、説明するものであるが、コーパス言語学は言語の特定の側面に限定するのではなく、むしろ言語研究のほぼ全ての分野において使用されることができると捉えることが最も妥当であると言える(McEnery et al. 2006: 7-8)。そしてこの方法論は、言語研究の新しいアプローチを誕生させ、理論とデータを一致させるような新しく有益な方法を生み出してきた (Leech 1998: xiii)。次節では、コーパス言語学がどのように発展してきたのかを俯瞰していく。

## 1.2. コーパス言語学の歴史

### 1.2.1. チョムスキーとコーパス言語学

言語学者の多くは、言語学理論や言語記述のゴールとして、チョムスキーによる見解の妥当性を一般的に認めてきた。つまり、言語学に固有のゴールは言語能力や言語知識を十分に説明し、規則や制約に関するこの能力をモデル化することである。 *Syntactic Structures* (Chomsky 1957) が出版された後、変形生成文法の支持者は徐々に言語理論の目的は言語行為を記述することではなく、言語使用者の知識について記述し十分に説明することであると考えようになった。この動きは、言語運用の記述研究から言語能力のモデル化へと多くの言語研究の方向性を変えた。チョムスキーはとりわけ、コーパスの使用と言語運用能力の統計に基づく確率モデル化を批判した。第二言語習得の分野においても、この生成主義のパラダイムが大きな影響を与えていたため、言語学習における頻度役割などはほとんど扱われず、頻度について取り扱われたとしても、否定的に扱われるか、上辺だけでしか扱われなかった。このような状況が、チョムスキー以降 20 世紀後半まで続いていた (Leech 2011: 10)。このような批判は特に、当時のコーパスの性質に起因するものであった。というのも当時のコーパスは、電子化されたコーパスではなかったため、大量の言語データについて全て手動でコーパスの収集、分析を行うことは不可能であった。結果的に、当時のコーパスは「歪みのある(skewed)」コーパスにならざるを得なかったのである (McEnery et al. 2006: 4)。

このように、コーパス的手法に批判的立場をとっていた生成主義であるが、チョムスキー派の言語学者が言語記述のために用いた証拠は、例文の正当性の内省的判断であることが多かった。さらに内省的言語学では、言語を均一なシステムとしてとらえているが、この均一性は本来一種の人工物であり、言語学者が作り出したものである (Kennedy 1998:

23-24, 271, Jacques 2005: 23)。Cori and David (2008 :115-116)はチョムスキーへの批判を以下のようにまとめている。

1) 文法性と容認可能性 :

特に、文法性と容認可能性の判断について批判がされている。それは、以下の三つの観点による批判である ;

- a) 判断の二分法 : 判断は二分されるべきではなく、連続体と考えるべきである。特に生成文法の支持者である言語学者は、判断基準を二項対立させている(cf. 容認可能/文法的 : 容認不可能/非文法的)。しかし、この方法では、多くの現象については分析を行うことはできないか、又は正しい説明を与えることができていない (Bod et al. 2003:1)。
- b) 判断の主観性 : 判断は言語学者の内省によって行われている。しかし、内省は信頼性の高い手段ではない。既に、個人差や個人内の差が存在することが指摘されている。
- c) 判断対象の不明瞭さ : 文脈上の要因や、適当な文脈であると判断する能力が干渉する以上、判断を何に基づいて行うかについて詳細なことは分かっていない。

2) 能力とパフォーマンスの概念 :

能力とパフォーマンスの間の境界線の線引きや、能力しか研究対象として認めないという立場における言語学の対象自体に批判が行われている。

- a) 能力とパフォーマンスの概念の線引きの基準は変化し続けている。生成文法学者からパフォーマンスだと考えられていた事象も、生成文法で扱えることが分かったとたんに能力の事象として考えられることもあった。
- b) 能力とパフォーマンスの概念の区分が曖昧である。
- c) 均一のコミュニティに属する理想的な話者という概念を持っている。
- d) 言語学の対象自体が定義し直されている。

確かに、巨大コーパスを分析する技術が確立される以前は、母語話者の言語に関する勘は基本的に正確であると一般的に見なされていた。したがって、実際の言語テキストを用いて例証することや、テキストを収集することは、勘を用いることと対して大きな違いがないと考えられてきた (Wray 2002: 21)。しかし、現在では巨大コーパスも入手しやすくなった。先行研究からも、コーパスに基づく記述は、言語の証拠もしくは言語理論の拠り所として言語行為に基づいているため、これまで重視されていた言語の普遍性よりも広く焦点を当てることができるということが分かっている (Kennedy 1998: 271)。

コーパスを用いるアプローチと内省的アプローチとの 2 つのアプローチの研究目的や研究の焦点が一般的に異なっても、これら是对立しているというよりは、相補的であると考えることができる。内省法とコーパスに基づく分析の両方を使用することによって、言語分析や言語記述に大きく貢献する。コーパスは、言語がどのように機能するのかについてすべてを提示することはできない。どの構造やどのプロセスが不可能であるのかを示すために、コーパスを使用することはできない。コーパスと直感的な判断が、2 つのアプ

ローチ間を埋めるために必要となる (Kennedy 1998: 271)。チョムスキーが批判してきたコーパスに基づく言語分析の復活によって、内省的言語学が貶められたのではない。コーパスによって研究の視点が変わったと言える (Jacques 2005: 21)。Leech (1991: 14)は、以下のように見解を述べている。

Neither the corpus linguist of the 1950s, who rejected intuition, nor the general linguist of the 1960s, who rejected corpus data, was able to achieve the interaction of data coverage and the insight that characterize the many successful corpus analyses of recent years.

(1950年代に直感を否定していたコーパス言語学者も、1960年代にコーパスデータを否定していた一般言語学者も、近年のコーパス分析に特徴付けられるような幅広く集められたデータと研究者の直感を相互につなぎ合わせるができなかった。)

コーパスに基づく分析と、言語学者の内省的判断を融合させることによって、言語の新たな側面を明らかにすることができるかと期待できる。

#### 1.2.2. 電子コーパス以前からコーパスの普及まで

電子的、機械可読コーパスに基づく研究が行われ始めたのは 1960 年代初頭であるが、これ以前にも主に 5 つの学術分野でコーパスに基づく様々な言語分析の伝統があった。それらは、1. 聖書および文学研究、2. 辞書学、3. 方言研究、4. 言語教育研究、5. 文法研究<sup>2</sup> の分野である。

##### 1) 聖書および文学研究：

言語学に関連するコーパスに基づく最も重要な研究の一つは、解説や批評を行うために聖書をコーパスとして使用する研究であった。18 世紀から、聖書の大部分が互いに事実と一致しているということを示すため、聖書で使用されている単語リストやコンコーダンスが作成された。

##### 2) 辞書学：

辞書学におけるコーパスの利用は、少なくとも 17 世紀初頭までさかのぼる。Samuel Johnson は 150 年以上前の前任者の仕事に基礎を置き、英語の単語の意味と使用を説明するために、最も著名な作家から引用した文を細長い紙片に記録してコーパスを構築した。Johnson は 6 人のアシスタントとともに、4 万語近くの見出し語を説明するために 15 万もの実例の引用を集め、*Dictionary of the English Language* に収録した。また、*Oxford English Dictionary (OED)* もまた同様に、コーパスに基づいた辞書である。1928 年に出版された第 12 版でかつ最終版でもある *OED* は、主に約紀元前 1000 年からの英語の名作を 71 年もの間継続して集めたコーパス制作の集大成であった。後の増補では、初版が出版された 1884 年から収集されてきた何百万もの引用紙片を利用することができた。特筆す

<sup>2</sup> この 5 つの分野に関する引用は、特記しない限り Kennedy (1998:13-19)から引用した



べき点は、コンピュータによる迅速で正確で網羅的な助力なしに、辞書編纂が行われたことである。

3) 方言研究：

方言研究のためのコーパスの利用は、19世紀にさかのぼる。この時代に地域方言の言語変異への興味が起こり、言語学者は多くの種類のコーパスを収集し解釈を行った。研究の多くは綴り字と発音の両側面から分析を行い、特定の概念を示す単語の選択に関する語彙の変異や特定の単語に可能な変異形を調査したものであった。データに基づくフランス語の方言学研究も特に19世紀後半から活発に行われていた。たとえば、スイス人の Jules Gilliéron が中心となって編纂した *Atlas linguistique de la France* は1902年に発表された。これは、方言学者 Edmond Edmont が1897年から1901年に行った方言調査をもとに、1400の調査項目をもうけて語の実地調査を行ったものである (Séguy 1973: 66-67)。

4) 言語教育研究：

コーパスに基づく教育的目的を持った研究は、20世紀前半におこった。これは最も影響を与えたコーパスに基づく研究の一つである。Edward Thorndike は当初、アメリカの英語母語話者へ行う識字教育のより良いカリキュラム作りに向けた語彙頻度リストを作成するために、41の異なる出典から成る450万語のコーパスを構築した。コーパスの4分の3は聖書や19世紀に出版された英語のフィクションの古典作品から引用されたものであった。また、手紙、新聞そして教科書もコーパスに引用されていた。1930年代の間に、Thorndike のコーパスは Irving Lorge の協力のもとアップグレードされた (Thorndike and Lorge 1944)。このコーパスは古典作品に偏りがあるものの、広い範囲の文献から引用され1800万語に至った。このコーパスの語彙分析や Thorndike や Lorge の研究から派生した出版物である *Theachers' Word Book of 30,000 Words* は、その後30年以上世界の広い地域で行われる英語教育に非常に大きな影響を与えた。

5) 文法研究：

第二言語としての英語の分野における多くの文法学者は、20世紀末までにコーパスが文法研究を根本的にかえるだろうという見解で一致していた (Conrad 2000: 548)。コーパスは長い間、記述文法を構築する典拠でもあった。20世紀前半の英語の記述文法の多くは、文法的特徴や文法構造を記述するために例として新聞や小説を使用していた。後にアメリカで Charles Fries によってより体系的で系統だったコーパスが構築された。これは教育的社会的バックグラウンドの記述に関する先駆的研究である。Fries は異なる教育的社会的バックグラウンドを持つ人々がアメリカ政府へむけて書いた手紙を、社会階級による言語使用の違いを説明するための基礎として使用し、*American English Grammar* (1940)で発表した。

再びコーパスの意義が見直され、コーパスが構築され始めるようになった1960年代初頭は、話し言葉の社会的語用的機能、談話単位研究への興味の高まりと一致する (Kennedy 1998: 174)。コンピュータ処理が可能な電子形式によるコーパスの構築は、アメリカとイギリスでそれぞれほぼ同時に行われていた。アメリカのロードアイランドにあ

る Brown University では、1961 年に Nelson Francis と Henry Kuçera が、アメリカ英語で書かれた出版物の 100 万語規模のコーパス構築に着手した。これは、*Brown University Corpus* または *Brown Corpus* と呼ばれるパイオニア的コーパスである<sup>3</sup> (Sinclair 2001: vii-viii)。このコーパスが重要なのは、言語研究に向けて構築された最初のコンピュータコーパスであるからだけではない。チョムスキーが広めたアメリカにおける言語学の新しくそして徐々に優勢となっていたパラダイムの常識を支持する人々から全く関心を示されないどころか、徹底的に敵対心を持たれていた状況で構築されたからである。このような時代背景があったことから、Francis と Kuçera はアメリカで出版された書き言葉の英語を代表するおおよそ 100 万語の共時的コーパスを構築するために、膨大で非常に大変な作業を行わなければならなかった。*Brown Corpus* の構造で特徴的なのは、2000 語ずつ 500 サンプル から構成されている点である。おおよそ総計 101 万 4300 語に及ぶ(Kennedy 1998: 23-24)。

イギリスでは *Brown Corpus* 構築の少し前から、University College London の Randolph Quirk が *Survey of English Usage Corpus (SEU Corpus)* の構築を始めていた。これは、1959 年から構築が始められたコーパスである。プロジェクト設立当時にはこのコーパスは電子コーパスではなく、コンピュータ化されていないコーパスに基づく記述と、現代のコーパス言語学の間での過渡的段階であったと言える。この調査を立ち上げた際の目的は、イギリスにおける英語の話し言葉や書き言葉の サンプルを各 5000 語ずつ 200 集め、学識のあるイギリス英語の母語話者の文法や言語使用を記述するための基礎として使用される 100 万語のコーパスを作成することであった。*SEU Corpus* の話し言葉のサンプルの収集、転写、コンピュータ処理は 25 年以上要し、1989 年に完成した。これが、通称 *London-Lund Corpus* と呼ばれるものである。*London-Lund Corpus* については、本章 1.2.5. で詳述する。

*Brown Corpus* と *SEU Corpus* はアメリカとイギリスで別々に構築されたコーパスであったが、カテゴリーや下位区分の構成には共通点も多い (Kennedy 1998: 27)。実際に、*Brown Corpus* や *SEU Corpus* のサイズや構成は高い評価を受けている。たとえば、*Lancaster-Oslo-Bergen Corpus of British English (LOB Corpus)* が構築された際、*Brown Corpus* の構成を採用した (McCarthy and Carter 1997: 21)。*LOB Corpus* は 1970 年から 1978 年に University of Lancaster と University of Oslo が、ノルウェーのベルゲンにある Norwegian Computing Centre for the Humanities と協力して *Brown Corpus* のイギリス版を作成するために構築したものである (Kennedy 1998: 28)。このコーパスは、言語の多様性について比較研究を行う目的を明らかにしたコーパスである (Aarts 1998: ix)。イギリス英語とアメリカ英語と比較するため、*LOB corpus* は *Brown Corpus* と同じ年代にイギリスで発行された出版物から構成されている (Johansson et al. 1978)。*LOB Corpus* は *Brown Corpus* と同じ 1961 年の出版物を収録しているが、構築は *Brown Corpus* よりも 10 年以上後に行われている点が、*LOB Corpus* と *Brown Corpus* との決定的な違いである。

---

<sup>3</sup> 正式名称は *Brown University Standard Corpus of Present-Day American English* (McEnery et al. 2006: 61)

その他にも、特に *Brown Corpus* の構造やサイズをモデルにしたコーパスはブラウンファミリーと呼ばれており、インド英語のコーパスである *Kolhapur Corpus of Indian English* (Shastri 1988)、ニュージーランド英語のコーパスである *Wellington Corpus of Written New Zealand English*、オーストラリア英語のコーパスである *Australian Corpus of English* (別名 *Macquarie Corpus of Written Australian English*)、カナダ英語のコーパスである *Corpus of English-Canadian Writing* などが、ブラウンファミリーとして知られている (Kennedy 1998: 29-30)。

このように、*Brown Corpus* と *SEU Corpus* におけるサイズや、ある言語の変異の特徴を代表する方法は、他の多くの第一世代の機械可読型コーパスのモデルとなった (Kennedy 1998: 45)。

### 1.2.3. コーパス基盤型とコーパス駆動型

前節では、現代の基盤となるコーパスの登場までのおおよその流れを俯瞰した。近年のコーパスについて概観する前に、コーパス構築や分析における基本的な概念についてこの節で扱う。

コーパス言語学では、アプローチをコーパス基盤型 (Corpus-based) のアプローチとコーパス駆動型 (Corpus-driven) アプローチの 2 つの流派に分けることがある。コーパス基盤型アプローチは、ランカスター大学の名誉教授である Geoffrey Leech の流れを組み、コーパス駆動型アプローチは、バーミンガム大学の名誉教授である John Sinclair の流れを組み、ロンドン学派とは、Daniel Jones と Rupert Firth を祖とする言語学の学派である (表 1)。

表 1 コーパス言語学のアプローチ

ロンドン学派		
創立者	Daniel Jones (1881-1967), John Rupert Firth (1890-1960)	
第二世代	M'.A.K. Halliday (1925-)	Randolph Quirk (1920-)
第三世代	John Sinclair (1993-2007) *コーパス駆動型	Geoffrey Leech (1936-) *コーパス基盤型

(Léon 2008: 14)

McEnery et al. (2006 :8-11)は、コーパス基盤型とコーパス駆動型の主な違いについて、以下の 4 点を挙げている :

#### 1) コーパスの性質

コーパス基盤型のアプローチでは、コーパスのバランスや代表性を重視するが、コーパス駆動型のアプローチではこれらの観点を重視する必要は無いと考えられている。なぜなら、コーパス駆動型アプローチをとるコーパスのサイズは一般的

に巨大で、サイズが大きくなればなるほど代表性が累積されると考えるからである。

## 2) 既存の理論・アプローチへの態度

これは、アノテーションへの考え方と関係する。コーパス基盤型はアノテーションについて積極的な態度を取るが、コーパス駆動型はアノテーションについて強い反対の態度を示す。なぜなら、コーパス駆動型のアプローチは、コーパスデータに全て基礎をおいた言語カテゴリーを前提としており、既存の理論や直感に基づかない立場を取るからである。したがって、コーパス駆動型アプローチではアノテーション自体が既存の理論の枠組みに基づくと考えられている。一方、コーパス基盤型アプローチでは、既存の理論を起点とし、コーパスから得られる証拠から既存の理論の見直しや訂正を行うため、アノテーションは一般的である。

## 3) 研究の焦点

コーパス駆動型のアプローチは、語彙、統語、語用、意味、談話の間の区別を全く行わない。なぜなら、この区別自体がコーパス以前の概念に基づくからである。そのため、特に機能的に完全な意味単位や言語パターンの単位についての一つのレベルについての言語記述しか行わないのが一般的である。

## 4) パラダイム

コーパス基盤型アプローチはコーパス駆動型アプローチほど急進的ではない。コーパス駆動型アプローチは全ての言語事象について記述できる新しいパラダイムであると主張している。

しかし、コーパス基盤型アプローチとコーパス駆動型アプローチのこのような完全な対立はなく、実際のところ区別は曖昧である(McEnery et al. 2006 :11)。

### 1.2.4. 近年のコーパス

1970年代で既に、語彙や意味分析には100万語規模のコーパスでも規模が十分ではないということが明らかになったため、より多くのテキストを集積し、保存し、管理し、分析する必要があった。1980年代までには、*SEU Corpus*や*Brown Corpus*のような100万語規模のコーパスは、時代遅れと考えられるようになった。幸いにもテキスト集積やテキストの保存技術が発達し、より大きなコーパスを構築することが可能となった。そのため、1990年代には1億語以上の規模のコーパスが入手可能となった(Kennedy 1998: 45-46)。コーパスとサイズは切っても切れない関係にあり、一般的にコーパスの規模が大きければ大きいほど研究に役立つと考えられている。「規模」という概念は、コーパスが構築された年代によって変わっていく。数年前に大規模コーパスと呼ばれていたものが、今日では小規模と見なされる。コンピュータ技術の発達とメモリの増加といった技術面の進歩、そして電子形式で簡単に大規模テキストが入手可能となったことから、かつては大規模とみなされていた百万語規模のコーパスも現在では小規模とみなされるようになった。現在入手可能な最大規模のコーパスは *the Bank of English (BoE)* で、5億語超の規模で

ある。このコーパスは、雑誌や書籍などの書き言葉に加えて、テレビニュースやラジオニュース、自由会話といった話し言葉も扱っている。このコーパスのうち約2億語について、1993年から1995年にかけてヘルシンキ大学の計量言語学の研究チームが形態レベルと統語レベルの情報タグ付けを行った (Järvinen 2003: 43)。BoEは、Collins Publishersによって支援されている国際英語コーパスプロジェクトであり、バーミンガム大学のCOBUILDプロジェクトチームが指揮をとった。

COBUILDプロジェクトは、*Birmingham Collection of English Text*を利用したものであり、大手出版社であるCollins Publisherとバーミンガム大学英語学部の研究チームとの合弁事業であった。COBUILDという名前は*Collins Birmingham University International Language Database*に由来している。COBUILD Corpusの中核を成す構造や内容は、以下の規則に従って設計されている(Kennedy 1998: 46)。

- 1) コーパスの25%は、話し言葉データを含むものでなければならず、専門的言語よりも広く一般的な言語を反映していなければならない。
- 2) 1960年以降に通用されているもので、ドラマの台本ではなくむしろ最新の自然発生的(naturally occurring)テキストであることが好ましい。
- 3) フィクションを含む散文体は含むが詩は含まない。
- 4) 話者および筆者は16歳以上の中から選ばれる。なお、地域語や方言話者ではない標準のイギリス英語話者70%とアメリカ英語話者20%そして他の地域変種の話者が10%であった。
- 5) 書き言葉のテキストはランダムに選択するのではなく、幅広く読まれた著書から選ぶ。なお、著者の性別の構成は75%が男性で25%が女性であった。
- 6) 話し言葉のテキストは、ラジオ放送、大学の口頭試験のアーカイブ、そして講義を含む転写をもとにする。

COBUILDプロジェクトは新しい領域を切り開いたと言われているが、それはコンピュータに基づくコーパスのサイズの大きさのみならず、コーパスに基づく文法辞書や言語教育コースを作るために、商業的な研究や開発プロジェクトとコーパスを結びつけたことにある(Kennedy 1998: 47)。

今後コーパスに規模はさらに拡大されていくことが予想できる。今後十数年後には数十億語のコーパスですら小規模コーパスと見なされる時代が来るかもしれない(Sinclair 2001: viii-ix)。近年の有名なコーパスでは、*Cambridge International Corpus*が挙げられるが、これは1億語規模である。このコーパスは*Cambridge International Dictionary of English*の基礎となったコーパスである(McCarthy and Carter 1997: 21-22)。そして、*British National Corpus (BNC)*もまた、1000万語の話し言葉テキストを含む1億語規模のコーパスである(McCarthy and Carter 1997: 21)。これは、*Brown Corpus*の100倍の規模に相当し、書き言葉と話し言葉における多くの種類を代表するものである。たとえば、書き言葉データは、イギリスで発刊された小説から説明的文章である科学記事までの幅広いデータを含んでいる。話し言葉コーパスは、自由会話とタスクに基づくパートの二部で構成されている。自由会話コーパスの多くは15歳以上の成人127

名の発話データである。また、年齢層、社会階級、性別も比較的偏りが無いようにデータ収集されている。タスクに基づく会話は、教育(講義、教室活動)、ビジネス(セールス、取引、会議)、公的機関(宗教、政治、法廷)、そしてよりカジュアルな場面(スポーツの解説、トーク番組)などから録音されたデータを含む。そして、*Brown Corpus* との大きな違いは、データがより最新であることである(Leech et al. 2001)。

英語の大規模コーパス構築プロジェクトは、このように半世紀以上にわたって進化を続けてきた。日本語における大規模プロジェクトについても、近年になって盛んに行われるようになった。国立国語研究所による「現代語書き言葉の均衡コーパス」は、2006年から2010年にかけて開発が行われたものである<sup>4</sup>。これは、日本語学における従来のコーパス言語学的分析の問題点を解決すべく、設計されたコーパスである。設計の基本方針として、以下が挙げられている。

- 1) 現代日本語の実態のできるだけ正確な縮図となるコーパス、いわゆる均衡コーパス(balanced corpus)とするため、統計的母集団からのランダムサンプリングを実施する。
- 2) 母集団が狭すぎないように、国立国語研究所がこれまでに集めた新聞、雑誌九十種、中学・高校教科書、TV放送だけでなく、一般書籍やインターネット上の文書までも加えて、現代日本語の書き言葉全体をひとつの母集団とみなす。そのような母集団を用いることによって、日本語の言語の多様性を幅広く把握できる汎用コーパスを設計することができる。
- 3) 著作権処理を実施して、公開可能なコーパスを構築する。
- 4) 国立国語研究所がこれまでに公開したコーパスともできるだけ互換性を保持する。

このような方針をもとに、現代日本語書き言葉均衡コーパスは、3つのサブコーパスから構成されている。2001年から2005年までに発行された書籍、新聞などが含まれる3500万語の生産実態(出版)サブコーパス、1986年から2005年までに発行された書籍で、公共図書館の収蔵図書をもとめた3000万語の流通実態(図書館)サブコーパス、そして、白書、法律、教科書、議事録などの3500万語からなるその他のサンプルのサブコーパスである。これらはすべてランダムサンプリングで集められたものである。

これらのコーパスは現在、オンライン公開されている。また、このコーパスの分析ツールである少納言<sup>5</sup>は一般公開されており、より詳細な分析を行うことができる中納言<sup>6</sup>は登録後に使用可能である。また、DVDの配布も行われている。

フランス語に関しては、*BNC* や *BoE* などのようなプロジェクトが今のところ存在しない(Verlinde and Selva 2001: 594)。フランス語における同様のプロジェクトは英語ほど広まりを見せていないのが事実である。フランスの参照コーパスに近い形として最も知られているのは、*FRANTEXT* であろう。*FRANTEXT* は、*Trésor de la Langue*

---

<sup>4</sup> 「現代語書き言葉の均衡コーパス」については、<http://www.ninjal.ac.jp/kotonoha/>を参照した。

<sup>5</sup> <http://www.kotonoha.gr.jp/shonagon>

<sup>6</sup> <https://chunagon.ninjal.ac.jp/login>

*Française(TLF)*という辞書開発に際して、例文を提示するために構成されたフランス語の巨大コーパスである。*TLF*は、Paul Imbs の指揮のもと 1957 年に考案され 1960 年に作成が始められた。Imbs は *Oxford English Dictionary* に匹敵するフランス語版の辞書を作ることを目的に *TLF* の編纂を始めた (Martin 2000 : 969-971)。現在では、1950 年以降のテキストが多くを占めているが、1180 年から 2009 年までのテキストを含んでいる<sup>7</sup>。

#### 1.2.5. 話し言葉コーパス

以上の参照コーパスの大半は、書き言葉に基づくものである。しかし実際のところ、一般的な原理に基づく、最も基本的な言語媒体は話し言葉であり、書き言葉はそのあと起こったものである (Čermák 2009 : 113)。書き言葉コーパスの構築が主に行われてきた理由について、Blanche-Benveniste (2005 :57)は以下のように述べている。

L'introduction de la langue parlée s'est généralement faite un peu plus tard que celle de la langue écrite en raison de la difficulté et du coût beaucoup plus considérable des recueils de données.

(話し言葉(コーパス)の導入は、一般的に書き言葉(コーパス)よりもやや遅れて行われた。それは、構築が難しいことと、データ収集により膨大な費用がかかるからである。)

たしかに今日では書き言葉コーパスの構築はスキャンを使うことで大量の書き言葉を一億語にも及ぶ語にアクセス可能なデータベースへと変換できる。このように、書き言葉コーパスの構築と比較すると、話し言葉コーパスの構築には数多くの困難があり、それが話し言葉コーパスの数が少ないという点に帰結する。Bruxelles et al. (2009: 6)は、話し言葉コーパスの数が少ない点について以下のように説明している :

La rareté des corpus oraux dans les grandes banques de données tient à des spécificité qui affectent tout le processus de recueil, constitution, traitement des données : nécessité de faire du terrain pour les enregistrer en situation écologiques diversifiées, exigences éthiques et juridiques en matière de respect de la vie privées des personnes concernées, complexité des données primaire (sous forme d'enregistrement multiples, sur différents supports) et de leur transformation en données secondaires (posant notamment des problèmes de conventions de transcription très différentes d'une discipline voire d'une théorie à une autre, ainsi que des problèmes de représentation graphique de l'oral), complexité des choix en matière d'annotation et d'alignement et des solutions informatiques pour y répondre. En outre, les techniques permettant de traiter ces spécificités de la langue orale sont encore très peu automatisées, demandant un travail colossal à effectuer à la main.

(大規模なデータバンクとしての話し言葉コーパスが数少ないのは、収集、構築、データ処理の全過程に関わる特異性に起因する : 多様な生態的状况で録音するためのフィールドワークの必要性、関係者のプライバ

---

<sup>7</sup> [http://www.frantext.fr/dendien/scripts/regular/7fmr.exe?CRITERE=QUI\\_FTX\\_QUEST\\_CE\\_DONC;ISIS=isis\\_bbibftx.txt;OUVRIR\\_MENU=8;s=s0e2119ec;ISIS=isis\\_bbibftx.txt](http://www.frantext.fr/dendien/scripts/regular/7fmr.exe?CRITERE=QUI_FTX_QUEST_CE_DONC;ISIS=isis_bbibftx.txt;OUVRIR_MENU=8;s=s0e2119ec;ISIS=isis_bbibftx.txt) を参照

ートの尊重に関わる倫理的・法的な要求、一次データの複雑性(多様な録音形態、様々な媒体)と一次データを二次データへ変換する複雑性(特に専門分野や、さらに理論ごとに異なる転写規則の問題や、話し言葉の描写の問題)、アノテーションやアライメント方法の選択の複雑性と、これらの問題に応えるための情報科学の解決法。これらに加えて、これらの話し言葉の特異性を扱うことのできる技術は、ほとんど自動化されておらず、手動で行わなければならない膨大な作業が要求される。)

しかし、話し言葉と書き言葉では、語彙や文法の特徴が異なっているため、コーパスが書き言葉のみに偏重していることは、言語記述を行う際に問題となる。実際、話し言葉に関する見解は、主に話し言葉コーパスの発達に伴い実に大きな変化を遂げてきた(McCarthy and O'keeffe 2004: 26)。たとえば、Fraisie and Breyton (1959: 61)は、第一言語における書き言葉と話し言葉の違いについて、以下の三点を挙げている。

- 1) 書き言葉を習得するかなり前に話し言葉を習得する。多くの人にとって、書くという行為よりも話すという行為の方を多く行っている。
- 2) 書き言葉と話し言葉の違いは、表現形態にある。書くことは時間がより長くかかる。書くために必要な考える時間をより長く持つことができる。反対に、話し言葉は条件がどうであれ、前に進み続け、いったん始めた文を終わらせ、次から次に文を続けていかなければならない。
- 3) 書き表現と口頭表現は状況が異なる。ものを書く際、言語産出者は紙の前にいるが、話す際には言語産出者は直接誰かを相手にしなければならない。これは前者が個人行為で後者が社会行為であり、異なっている。このことによって、パーソナリティなどの別の変数を考慮しなければならない。

さらに、話し言葉コーパスの登場によって発見された言語項目や言語構造は数多く、話し言葉コーパスの重要性はますます認識されるようになってきた(Čermák 2009: 113)。そのため、更なる話し言葉コーパスの構築が望まれている。

話し言葉コーパスは、スピーチコーパスと区別して考える必要がある。一般的に話し言葉コーパスは非公式の即興会話データであると考えられているが、口頭形式で行われたいかなる形式の言語データも話し言葉コーパスとしてとらえられている場合もある(Sinclair 1996: 8)。しかし、口頭形式で行われた会話にもいろいろな場合が想定できるため、口頭形式の全ての言語データを話し言葉コーパスの範疇に分類するのは問題となるだろう。たとえば、テレビやラジオのニュースで読まれた原稿は書き言葉であり、これを話し言葉として扱うには無理が生じる。スピーチも同様である。スピーチも一般的に、書き言葉を出発点として行われた口語形態のものである。したがって、スピーチを純粋な話し言葉として見なすことはできない。

書き言葉コーパス構築の陰で、話し言葉コーパスは以前から構築されていたが、影響力は書き言葉コーパスには到底及ばなかった。現在までに、様々な話し言葉コーパスが構築されてきたが、話し言葉コーパスの多くは、英語母語話者の話し言葉コーパスである。代表的な英語母語話者話し言葉コーパスについて、以下にいくつか例を挙げる。



1) *Oral Vocabulary of the Australian Worker (OVAW)*

初期のパイオニア的話し言葉コーパスの一つとして *OVAW* というプロジェクトが作成したコーパスがある。*OVAW* プロジェクトは、約 50 万語の話し言葉データを集め、特徴語や話し言葉特有の句などについて分析し、記述した (McCarthy and Carter 1997: 22)。

2) *American Davis-Howes Count of Spoken English*

*OVAW* の約 10 年後に Davis Howes が *A Word Count of Spoken English* (Howes 1966) を発表した。このコーパスには大学生や病院の患者とのインタビューを転写した約 26 万語が含まれている (McCarthy and Carter 1997: 22)。

3) *London-Lund Corpus*

初期の話し言葉コーパスと知られているコーパスの一つである。このコーパスは、教育を受けたイギリス英語が母語である成人話者の自然会話を集めたコーパスである。このコーパスは前述の通り、University College of London の Quirk の指揮の下行われた *Survey of English Usage (SEU)* に端を発している。*SEU* のうち、話し言葉に関するコーパス構築は、1975 年に University of Lund の Jan Svartvik によって指揮され、構築が始まった。これは、1960 年代から 1970 年代中甸までの話し言葉に関するデータである。各 5000 語から成る 100 の話し言葉テキストで構成され、計 500 万語の規模である。サンプルは、様々な度合いの形式性や非形式性、モノログ、ダイアログに加えて、会話、インタビュー、レクチャー、セミナー、会議や電話などの様々なコンテキストに現れる様々なトピックを反映する幅広いジャンルをカバーしている。話し言葉のセクションは 1980 年から電子形式で出版することが可能となった (Kennedy 1998: 17-19, McEnery et al. 2006: 63)。

4) *Santa Barbara Corpus of Spoken American English (SBCSAE)*

1990 年代に構築された重要なコーパスの一つは 1991 年に Wallace L. Chafe などによって構築された *SBCSAE* である。これは、前述の *International Corpus of English (ICE)* のアメリカ英語のパートであり、アメリカ全土から様々な年齢層、職業、民族、社会的背景の幅広いバラエティの成人話者によるデータから成るコーパスである。数百もの自然会話から構成されており、アメリカ英語のコンピュータ化された最初の大規模コーパスでもある。*SBCSAE* は、イギリス英語の話し言葉コーパスである *London-Lund Corpus* 以来の数少ない巨大な話し言葉コーパスの一つである。元来、20 万語を収録する予定であったが、後に 100 万語に到達した。*SBCSAE* は約 900 名の話者による会話 80 時間から構成されている。会話は、ダイアログ、モノログ、自然会話、事前準備した会話、電話による会話、議論、サービスエンカウンター、セールストーク、仕事場における会話、ビジネス会議、政治会議、カウンセリング、レクチャー、教会の説教、礼拝、トークショー、ラジオ番組における電話の会話、ニュース、スポーツ中継など多岐にわたっている (Kennedy 1998: 37, McEnery et al. 2006: 63)。

5) *Lancaster/IBM Spoken English Corpus (SEC)*

このコーパスは、1992年に Gerry Knowles らによって構築されたコーパスで、詳細なプロソディ研究のために設計されている。このコーパスは 1984年から 1997 までのイギリス英語の話し言葉おおよそ 52600 語から構成されている。主にラジオ放送の話し言葉データから構成されており、コメンタリー、ニュース、講義、ダイアログなど 11 のカテゴリーに分けられている (Kennedy 1998: 36, McEnery et al. 2006: 63)。

6) *Wellington Corpus of Spoken New Zealand English (WSC)*

これは、ニュージーランドの話し言葉に関する約 100 万語規模のコーパスである。1988年から 1994年の間に収集された 551 の話し言葉データから構成されているが、その大部分は 1990年から 1994年に収集されたものである。データの 12 パーセントは、モノログの公式スピーチ、13 パーセントはタスクに基づくモノログの半公式的スピーチ、残りの 75 パーセントは非公式のダイアログの会話である (McEnery et al. 2006 :63)。

7) *Cambridge and Nottingham Corpus of Discourse in English (CANCODE)*

このコーパスは、*Cabridge Inernational Corpus(CIC)*の一部である (McEnery et al. 2006 :63)。BNCや *Bank of English* のような他の現代コーパスと比較すると規模は小さいが、このコーパスは緻密に分類が行われており、さらにタグ付けも行われているため、コンテキスト間の比較に有用なリソースとなっている。主にデータ収集は 1994年から 1999年に行われ、様々な談話コンテキストやスピーチジャンルにわたる会話の収集が行われた。母集団を大きくし、幅広い社会経済層を確保するため、さまざまな年齢、性別、社会背景、教育レベルを持つ成人の話者が参加できるよう、参加者は注意深く選別された。また、公的な会話か私的な会話かの伝統的区別も、一般のガイドラインとして使用した。CANCODE コーパスに適応された発話のコンテキストタイプは、以下の 5 つに大きく 分類することができる (Adolphs and Schmitt. 2004: 41-43) :

a. 私的コンテキスト

話者間は近い関係で、多くの場合共同生活を送っている者同士である。パートナーおよび家族間の会話がこのカテゴリーに分類されている。

b. 社会・文化的コンテキスト

話者間の関係は友人関係が多く、a のカテゴリーほど話者間の関係は近しくない。また、他のどのカテゴリーにも分類されなかったテキストがこのカテゴリーに分類されている。録音環境は、集会、誕生日会、スポーツクラブ、ボランティアの会議などである。

c. 職業的コンテキスト

このカテゴリーでは、日常業務におけるインタラクションを扱う。しかし、接客業における店員と客の会話は d に分類されている。また、話題が必ずしも業務上関係ないものでも、勤務先での同僚の会話はこのカテゴリーとして分類されている。

d. 相互交流コンテキスト

話者同士が初対面である場合、このカテゴリーに属される。このカテゴリーで扱う会話は、ある交流的目的を持っている。特に、このカテゴリーに分類されたインタラクションの主な目的は財貨とサービスである。

e. 教育的コンテキスト

話者間の関係が教育的コンテキストによって定義できる場合、このカテゴリーに属される。個人指導、セミナー、講義、教授会議などが含まれている。

英語以外でも、これまでに日本語についても話し言葉コーパスの参照コーパスが構築されている。日本語話し言葉コーパス(*Corpus of Spontaneous Japanese: CSJ*)は、国立国語研究所、情報通信研究機構(旧通信総合研究所)、東京工業大学が共同開発した、質量ともに世界最高水準の話し言葉である<sup>8</sup>。このコーパスは、日本語の自発音声に多くの研究情報タグが付加されている。音響、言語モデルを構築することを目的とするため、音声認識にとって主要な対象となるモノログを集中的に収集している。音声のタイプは、学会講演、模擬講演、講演インタビュー、タスクに基づく対話、自由対話、再朗読、朗読などであり、話者は約 1400 名、録音時間は 661 時間にわたる。なお、再朗読は学会講演や模擬講演の音声転写テキストを同一の話者が朗読したものであり、朗読は書き言葉テキストや話し言葉テキストを朗読したものである。総語彙数はおおよそ 752 万 5 千語である(前川 2006: 1-6)。CSJは 2004 年から頒布が開始されており、音声・音韻、形態、語彙、統語、談話、自動言語処理など多岐にわたる分野で分析が行われている。

フランス語に関しても話し言葉コーパスが作成されている。フランス語の場合、1990 年代中旬より話し言葉コーパスの開発が盛んになったが、これは英語を始め、他の国際的な話し言葉コーパスの開発からは遅れをとっていたということを意味している(Bruxelles et al. 2009: 2, Bert et al. 2010: 17)。1990 年代以前にも話し言葉コーパスの構築は行われていたが、Cappeau et Seijido (2005: 5-6)はこれを «corpus d'historiques» と呼び、代表的なコーパスとして以下の 3 つを挙げている。

1) *Archives de la parole*

1911 年に Ferdinand Brunot によって構築されたものである。表音化文字などの当時の最新技術を駆使することによって、話し言葉を録音し保存することが可能となった。

2) *Français fondamental*

1951 年から 1955 年にかけて構築されたコーパスであり、話し言葉については 275 の録音からなる計 30 万語を有する。

---

<sup>8</sup> 明記しない限り、CSJ については <http://www.ninjal.ac.jp/csj> を参照した。

### 3) *Corpus d'Orléans*

1960年代末から作成されたコーパスで、オルレアンに住む話者を録音の対象としている。このコーパスの一部は、後述する *ELILAP* などのコーパスにも収録されている。

今日では、話し言葉フランス語の巨大コーパスの構築やコーパスプロジェクトも立ち上げられるようになった。以下に代表的なコーパスやプロジェクトを紹介する。

#### 1) *Groupe Aixois de Recherche en Syntaxe (GARS) - Description Linguistique Informatisé sur Corpus(DELIC)*<sup>9</sup>

それほど大規模ではないが、幅広い社会言語的背景を持つ話者によって構成されたコーパスとして *DELIC* が挙げられる。このコーパスは 36 時間以上の会話に相当する 44 万語を含んでいる。*GARS* は 1998 年に Claire Blanche-Benvenist の指揮により、プロバンス大学のコーパスチームによって発足された。2000 年より、このプロジェクトは Jean Véronis の指揮のもと *DELIC* チームが引き継いでいる。このコーパス構築の目的は、フランスの主要都市における今日の話し言葉を観察することであり、言語学者、研究者そして教師にデータを提供することである。したがって、一般的で日常の使用と定義することができる話し言葉を代表するようなデータがなによりも優先して集められた。また、発話者の特徴や発話状況を反映するよう、場面の選択を行った。録音地点は、北部の 27 地方と南部の 10 地方とパリの地域である。最終的な録音数は、北部で 50、南部で 57、そしてパリ地域で 27 となり、合計録音数は 134 に至った。*DELIC* チームは以下の 3 つの録音状況を設定している。一つ目は私的な会話であり、特にアンケートの枠組みとして行われたインタビュー形式で行われた。旅行に関する話題や子供の時の思い出など生活に関連することと、職業などのノウハウを説明するものの二種類の産出状況がある。二つ目は、職業に関する会話であり、インタビュー形式で行われた。しかし、会話は職業訓練中か仕事場で録音された。三つ目は公的な会話であり、聴衆がいる常にいる状況で話さなければならないという点で上記の二つの会話状況とは異なっている。一部はインタビュー形式で行われ、その他は時事、インタビュー、会議、電話による討論などのラジオ放送であったり、授業や学会、市議会、組合会議、地区会議などの政治会議であったり、美術館のガイド、ワインの試飲などのより特定の状況に関するものであったりした。コーパス構築の目的は、リラックスした状態で自然会話を録音することであったので、録音は MD の録音機能を使って行い、マイクやヘッドホンなどは使用しなかった。そのため、詳しい音声研究などには適さないかもしれない。転写に用いられた規則は、*GARS* の方法から着想を得た。これは、標準正書法を用いるものであるが、句読点や感嘆符や疑問符を用いない。また、ポーズや完全に発音されない単語も表記する。また、発話者に録音許可証や配信許可証への記述も義務づけており、また個人が特定される固有名詞や会社名などは偽名するなどの配慮を行い、法的な観

<sup>9</sup> DELIC に関する記述は特記しない限り Equipe DELIC (2004)に基づいている。

点にも細心の注意を払っている。このコーパスによって全ての言語使用を反映することは不可能であるが、フランス語研究やフランス語教育への資料として使用することは十分に可能である。実際に、このコーパスは様々なアクセント、表現、言い回しを含んでおり、現代都心部のフランス語の詳細な言語イメージを反映している。なお、1つの会話の平均時間は16分48秒である。発話者は134名で、様々な教育背景を持っており、会話のシチュエーションの種類も豊富である。男女比は男性が56%で女性が44%である。また、発話者の教育レベルも、三段階に分けられており、小中学校レベルの発話者が24.77%、バカロレアレベルが41.28%、そして高等教育レベルが28.44%であった。

## 2) *Projet CIEL-F*<sup>10</sup>

このプロジェクトは、コーパス言語学、変異言語学、相互行為言語学、機能言語学の理論と手法から着想を得たものである。コミュニケーションエリアの分類に基づき、さまざまな生態的行為の種類をサンプリングしており、革新的なアプローチであると言える<sup>11</sup>。*CIEL-F*の構築の目的は、以下の主な3つの点である：

1. 言語学者によって作られた環境下ではなく、言語が使用されている自然な環境下で、言語データを構築すること。自然会話はコミュニケーションの機会、話者の能力によって変わるもので、収集データを歪曲してはいけない。
2. ある一つの、もしくは限られたフランス語圏に限定するのではなく、様々なフランス語圏を代表する口頭サンプルを含むこと。
3. 比較可能なデータを提供すること。様々なフランス語圏の場所、様々な会話の状況を文書化する。
4. 様々な種類の分析に対応できるコーパスを構築すること。

食事中の会話、仕事の会議もしくはサービスや商業的なインタラクション、電話における会話もしくは地元のラジオのディベートの録音といった3種類のインタラクションのシチュエーションのもと録音が行われた。録音は主にフランス、スイス、ベルギー、アンチル諸島、カナダ、アルジェリア、エジプト、コートジボワール、カメルーン、セネガル、レバノン、モーリシャス、レユニオン島といったコミュニケーション域で行われた。このプロジェクトは以下の2つの専門的業務を行っている：

1. コーパスベースで国際的な専門業務を行う；リヨン大学で開発されたコーパスである *CLAPI*、Louvain 大学で開発されたコーパスである *VALIBEL*、Freiburg 大学と Louvain 大学による共同開発のオペレーティングシステムである *MOCA* に基づいている。
2. 共同研究チームごとに作業を行う；相互行為言語学(*ICAR* 研究室)、社会言語学(*VALIBEL* センターと Paris Nanterre 大学の *MoDyCo* 研究室)、

<sup>10</sup> 特記しない限り、Bert et al. (2010)および Gadet et al. (2012)を参照した。

<sup>11</sup> <http://www.ciel-f.org/>を参照した。

接触言語学(Halle 大学の文化接触の研究グループ及び Freiburg 大学のロマンス語セミナー)の分野の研究が行われている。

以下、このプロジェクトの核となっている 2 つのコーパスである *CLAPI* と *VALIBEL* についてまとめる。

a. *CLAPI (Corpus de Langues Parlées en Interaction)*

このコーパスの構築は、リヨンの *UMR GRIC (Groupe de Recherches sur les Interactions Communicatives)* によって 1990 年代後半に始まった。今日では、*UMR ICAR (Interactions, Corpus, Apprentissages, Représentations)* と名前を変え、運営が行われている。このコーパス構築の目的は、研究者および学生が作成したコーパスを目録化し、利用できるようにすることであった。第一版である *CLAPI 1.0* は 2000 年にインターネット上で公開されており、第二版である *CLAPI 2.0* が 2005 年に公開された際には、第一版のデータは全て電子化され、メタデータなどがさらに付け加えられた。また、ここにはフランスにおけるインタラクション研究で使用されたコーパスの祖とも言える *Mode* や *La Dame de Caluire* といったコーパスも含まれている。同時に、*Fonds Bielefeld* など *ICAR* が開発して来た別のコーパスも *CLAPI* に統合された。600 時間の録音がアーカイブ化され、135 時間分がコーパス化されている。

このコーパスの特色は、録音状況が多様であることであり、自由会話、職場での会話といった典型的な状況における会話の他にも、これまでのコーパスではあまり扱われてこなかった状況における会話(歯科、離婚裁判)や、多様な社会実践をカバーする状況における会話(接客、商業取引)も含まれている。

b. *VALIBEL (Variété Linguistiques du français en BELgique)*<sup>12</sup>

このコーパスは、1988 年後半に行われた話し言葉フランス語に関する 2 つの異なる研究に着想を得て、構築されたものである。一つ目が、先述した統語研究である *GARS* であり、もう一つはケベックの社会言語学的研究である。このコーパスが構築された当時、ベルギーのフランス語圏では規範文法の伝統から少しずつ解放されてきた時期であり、ベルジシズムに否定の立場を取る純粋主義から、実使用の記述への興味へと研究の対象が移ってきた時代であった。このような背景のもと、ベルギーのフランス語圏では、フランス語の言語変異を文書化するための信頼の高い資料を準備することが必要課題であった。そこで、*VALIBEL* はワロン地方とブリュッセルにおけるベルギーの言語変異を文書化することとなった。

1987 年から 1995 年にかけて 533 名のベルギー人を対象に録音が行われた。これは 373 時間の録音と約 390 万語に匹敵する。会話の形式は様々で、インタビュー形式、自由会話、授業、ディベートなどがある。インタ

---

<sup>12</sup> 特記しない場合、Dister et al. (2009)を参照した。

ビューは修士論文の審査を録音したものが多い。テーマも同様に多様であるが、ベルギーフランス語に関する話題が多い<sup>13</sup>。

### 3) *Etude Linguistique de la Langue Parlée(ELILAP)*<sup>14</sup>

フランス語母語話者による話し言葉を収集したコーパスで、最も古いものの一つが、*ELILAP* である。このコーパスは、1980年から1983年にかけて Leuven 大学によって構築された。なお、このコーパスは3つのデータから構成されている。

#### a. 1968年から1971年までに集められたオルレアンの社会言語研究 (*Etude Sociolinguistique sur Orléan:ESLO*)のデータ

Essex 大学のフランス語教師が Bureau pour l'Etude de l'enseignement de la Langue et de la Civilisation françaises de Paris (BELC)の協力のもと行ったアンケートが元となったコーパスである。話者は、オルレアンの都市部の住人を、Institut National de Statistique et d'Etudes Economiques (INSEE)のリストより無作為抽出した。コーパスの規模はおよそ450万語で315時間の録音に相当する。録音はコミュニケーションの状況から、対面式会話、職場などにおける非公式な会話、電話会話、インタビュー、会議やディスカッション、公共の場の会話(レストランやスーパーなど)、心理教育医療センターでの会話といった、7つのグループに分けられている。現在までに、80時間相当の約90万語の転写が行われている。

#### b. 1974年に収集されたトゥールコーパス (*Le livre parlé de Tours*)

ロンドンにある College of Education のフランス語教師が、トゥールにある Institut Français du Royaume-Uni (IFRU) で行われたインターンシップを機に作成したコーパスである。職業や年齢が様々なトゥール出身者193名の録音を120時間行った。13のテーマで、形式的な対面式の会話が行われた。このうち、4時間に相当する3万6千語が転写されている。

#### c. 1976年に構築されたオーベルニュコーパス (*Voix d'Auvergne*)

Clermont-Ferrant の IFRU で行われたインターンシップに参加したフランス語教師などによって構築されたコーパスである。インターンシップの目的は、現代フランス社会と話し言葉フランス語の多様性をより良く理解することであり、インターンシップの一環として録音が行われた。録音は計52時間で、17時間分の録音転写されている(およそ17万7千語)。

これら3つのコーパスを全て合わせると、500時間近い録音、計100万語超のフランス話し言葉コーパスとなっている。

<sup>13</sup> <http://www.uclouvain.be/81834.html> を参照した。

<sup>14</sup> 特記しない場合、<http://bach.arts.kuleuven.be/elicop/projetELILAP.htm> を参照した。

#### 4) *Phonologie du français contemporain : usages, variétés et structure(PFC)*

現在、質的量的観点から最も発達しているフランス語コーパスプロジェクトの一つは *PFC* である<sup>15</sup>。この *PFC* プロジェクトは、Université Toulouse II の Jacques Durand、Université Paris Ouest Nanterre La Défense の Bernard Laks、Université d'Oslo の Chantal Lyche によって始められた音韻研究プロジェクトである。以下の目的のために、話し言葉フランス語の巨大データベースが構築された：

- a. 話し言葉フランス語の単位や多様性について、より正確な姿を提供する
- b. 共時的観点と通時的観点と両観点から音韻モデルを検証する
- c. 共通の手法を使用して話し言葉フランス語の巨大データベースを構築する
- d. 音韻知識を応用し、話し言葉の自動処理ツール開発を行う
- e. フランス語教育やフランス語学のためのデータを拡大し、更新する

発足から 10 年間、*PFC* は世界中の 60 名超の研究者や学生によって、プロジェクトが共同で遂行されている。現在では、フランスに限定せず 50 以上のフランス語圏のアンケート地点において、500 名以上の話者の録音が行われており、その半数は既にインターネットのサイト<sup>16</sup>からアクセスすることができる。

*PFC* は共通のプロトコルがあり、データの比較を行うことができる。各アンケート地点でできるだけ年代の異なる 10 名の代表者が以下のタスクを行って、それについて録音した：

- a. 話者の音韻体系を引き出すと考えられる *PFC* 単語リストの音読
- b. 他のタスクで観察された話者に特有の音韻体系と、シュワやリエゾンなどを観察することができると考えられる *PFC* テキストの音読
- c. アンケート調査者との方向付けられた会話
- d. 自由会話

*PFC* から派生したプロジェクトも 2 つある。それは、*PFC* のデータをフランス語教育へ応用するのを目的に、2006 年から行われている *PFC-EF* (*Enseignement du Français*) プロジェクトと、学習者のフランス語のデータベース構築と学習者の中間音韻を研究するのを目的に始められた *IPFC* (*InterPhonologie Français Contemporain*) プロジェクトである。中間音韻とは、中間言語<sup>17</sup>から派生した用語である。*IPFC* プロジェクトについては、本章 1.4.6.4. で詳述する。

### 1.3. コーパス分析ツール

コーパスを構築するだけでは、言語分析を行うことはできない。分析するツールも不可欠であるし、そのツールを使いこなせる知識もまた分析者に要求される (Nesselhauf

<sup>15</sup> *PFC* に関する記述は特記しない限り Detey et al. (2010) に基づいている

<sup>16</sup> <http://www.projet-pfc.net>

<sup>17</sup> 本章 1.4.1. を参照のこと



2005: 41)。まず一般的な分析機能は、テキスト内の単語数を数える機能である。これは、テキスト内の数多い言語項目の頻度を比較するのに必須の機能である (Granger 2004: 127)。特に広く使用されているのが、*WordSmith Tool* (Scott 1996)である。*WordSmith Tool*には大きく分けて3つの機能がある。

1) *Concord* :

*Concord* という機能は、コーパス内の特定の語や文の共起リストであるコンコードダンスの分析を行うための機能である。コンテキスト内における形態素、単語、文の用例を得ることができ、とても優れた機能である。*Concord* でコンコードダンスを調べることによって、対象の語がどのような語と共に現れるのかという観点だけでなく、頻度情報も入手することができる (Scott 2001: 50-51)。

2) *WordList*

語のリストを作成するこの機能は、コーパスの中の共通の語を特定するのに役立つ。また、どの語を教えるのか、または教えなくても良いのかを決定するときや、新しい語彙をどれくらいの頻度で教科書に入れるべきかを決定する時に役立つ情報を提供する (Scott 2001: 54)。

3) *KeyWords*

キーワードの研究は、コーパス言語学のアプローチによって恩恵を受けたものの一つであると言っても過言ではないだろう。ある単語について、最少の頻度として規定された以上の回数でテキスト内に生起する場合、また、参照コーパスの頻度と比較するとテキスト内の頻度の統計的確率が規定値以下の場合、キーワードと見なされる。キーワードリストは通常、統計的に最も有意なキーワードから並び、*Keyness* の順序で表示される。研究者はキーワードリストを使用して、特殊性の記述を行うことができる。キーワードによって、研究者のバイアス抜きに、迅速かつ効率的にテキストにおける異常に頻度の高い要素や頻度の低い要素に目を向けることができる。言語事象のより複雑な分析を可能にした (Baker 2004: 346-348)。

たしかに、コンピュータによる自動化によって、手動処理では起こらなかったような弊害もでてくる。しかし、コンピュータの自動処理と従来の手動処理を組み合わせることで、言語学の分析が大きな広がりを見せたことは事実である。

#### 1.4. 学習者コーパスとプロジェクト

これまで、主に参照コーパスとして分類可能なコーパスについて概観してきた。本節では、学習者コーパスについて扱う。学習者コーパスは、学習者の言語産出データの記録である。学習者コーパスを用いて、学習者がある状況下で実際に発した書き言葉や話し言葉について分析することができる (Nesselhauf 2005: 41, Aijmer 2004: 173)。学習者コーパスは、英語母語話者コーパス構築から30年以上遅れること、1990年代後半から構築が始まり、広がりを見せたのは2000年に入ってからである。したがって、学習者コーパス

研究は、比較的新しい分野であり、学習者コーパスを基盤にして行われた研究は広がり始めたばかりである (Nesselhauf 2005: 40)。

#### 1.4.1. コーパス登場以前の間言言語研究

言語を学ぶ過程にある第二言語学習者や外国語学習者が産出する言語のタイプは中間言語と呼ばれている。中間言語という概念は、Selinker が 1969 年と 1972 年に提唱したものである。この概念は、中間言語は母語にも目標言語にも属さない流動的な言語体系であり、独自の言語体系を持つという考えに起因している。そして、学習者と母語話者が同じ意味について言語産出する際、多くの場合学習者によって産出された言語は、母語話者によって産出された言語とは異なっている (Selinker 1972: 214)。また、外国語学習の過程は、中間言語の連続体上における学習者の発達であると考えられている。つまり、外国語に対する知識が全くないという段階から、目標言語の母語話者と同じとは言わなくともそれに近い能力を持っているという段階までの連続体の上で、学習者の発達が記述されることが多い (Laufer 1998: 255)。

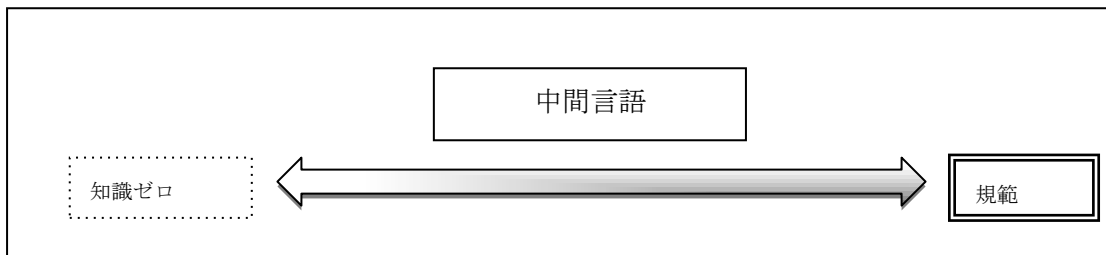


図 2 中間言語

中間言語研究はもちろん、学習者コーパスの登場以前にも行われていた。中間言語研究を行うためには、学習者の産出言語を分析する必要がある。Ellis (1994: 670)は、中間言語研究におけるデータの種類について、以下のような分類を行っている：

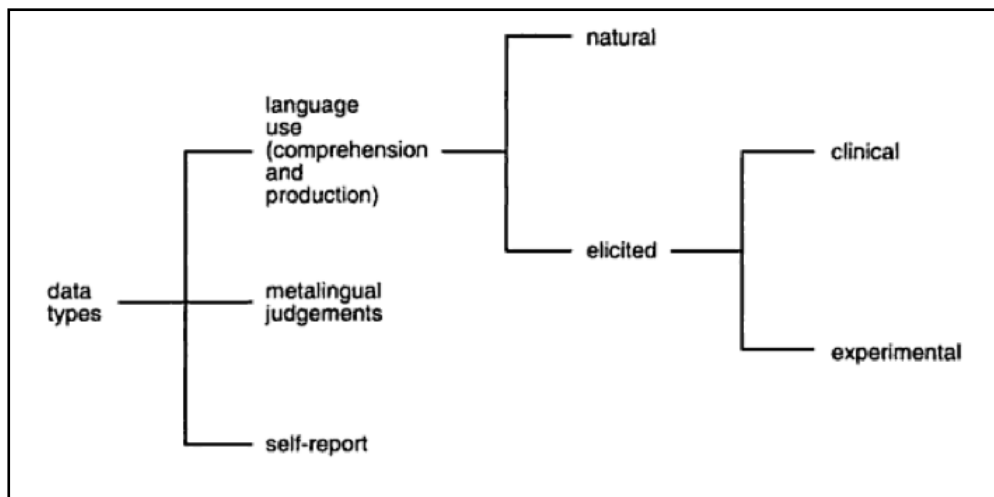


図 3 中間言語分析におけるデータの種類 (Ellis 1994:670)

まず、「言語使用データ」、「メタ言語判断データ」、そして「セルフレポートデータ」の3つに分類されている(Ellis 1994: 670)。

1) 「言語使用データ」

学習者の理解面、産出面の両方における目標言語の使用データを指している。さらにこのタイプのデータは、「非誘導型」と「誘導型言語使用」に分けられている。前者は、学習環境ではなく、学習者が目標言語でコミュニケーションを遂行している際の言語使用を指す。後者は「臨床的抽出」と「実験的抽出」に下位分類されている。「臨床的抽出」とは、インタビュー、インフォメーションギャップタスク、ロールプレイングなどを通して学習者にデータを産出させる方法で、「実験的抽出」は研究者の興味がある特定の箇所について、学習者にデータを産出させる方法である。

2) 「メタ言語判断データ」

文法性判断テストなどによって抽出される。たとえば、学習者に文法性を判断させ、学習者の外国語における直感を引き出す。

3) 「セルフレポートデータ」

学習者が使用するストラテジーや解決法について質問紙、インタビュー、思考表出法などのタスクによって抽出されるデータである。

学習者の言語産出データには、このようにいろいろな種類があるが、近年の SLA 研究で一般的に収集されてきたデータは内省データであるだろう。これは、Ellis の分類の「メタ言語判断データ」と「セルフレポートデータ」と「言語使用データ」の「誘導型」データに相当する(Granger 1998: 4)。

このような手法を用いて 1960 年代後半から 1970 年代には、特にエラー分析が頻繁に行われていた。しかし、このような内省や誘出型のデータによるエラー分析には限界がある。まず、このような方法ではあらかじめ設定されていた問題点についてしか分析を行うことができない。そしてエラー分析のデータ収集は多くの場合、体系的に行われておらず、学習者についての詳細やデータ収集が行われた環境についての詳細はほとんど記録されていなかった。また、データ収集はエラーの収集としてしか見なされておらず、エラーを誘発できた後にデータが消去されてしまうことが多かった。したがって、テキストタイプ、テキストの長さに関する情報は失われ、エラーが起こったコンテキストに関しても、多くの場合は入手不可能であったため、分析結果の確認や再分析は不可能であった。そして、学習者が正しく発したものに関しては通常研究対象にはならず、エラーが絶対的で、学習者がどれくらいの頻度で正しい構造や正しい要素を発しているかという現象とも比較することができなかった。また、伝統的なエラー分析で使用されるテキストは、コンピュータ化されていなかった(Nesselhauf 2005: 40-41)。なによりも、学習者のエラーを分類することは、時間がかかる上に無益なことが多い。なぜなら、エラーを分類する方法は多数あり、研究目的や関連する理論によるからである。また、エラー分類は依拠する理論にのみしか適応されないことも多い。そして、人によってエラーに関する観点が異なるため、採点者間信頼性も低くなる(Tono 2003: 801)。エラー分析だけでは、学習者言語の外国語

らしさについて明確に説明できないことも多い。たとえば、学習者によるフランス語会話はしばしば、明確な文法エラーがないのにも関わらず、母語話者に違和感を抱かせることがある。その外国語らしさの理由について母語話者が明らかな説明を行うことも困難な場合がある。たとえば Debrock et al. (1999 : 47)は、エラーではない外国語らしさについて以下のような例を挙げている。

- 1) 口頭による会話なのにもかかわらず、学習者が書き言葉で話す場合
- 2) インタクションが足りない場合で、その理由が目標言語の言語規則を習得しているか否かという問題によらない場合
- 3) 会話が直接的すぎる場合

さらに、実験による言語環境は人工的で、学習者が普段自然に使う言語とは異なる言語を産出する場合がある。そして、制限された実験的な誘出法のため、SLA 研究者は限定的な実験的研究を行う傾向にあり、限られた被験者から結果を一般化することはできない。そのため、良質で大規模な言語データが必要となる(Granger 1998:5)。

このような言語データとして挙げられるのが、学習者コーパスである。学習者コーパスは「自然言語データ」として分類できると一般的に考えられる。前述の通り、「自然言語データ」は積極的に SLA 研究に用いられてこなかったが、これには以下のような理由に起因している(Granger 1998:4-5)。

- 1) 興味の対象となったある特定の言語特徴のなかには頻度が低いものがあることより、誘出せざるをえなかったため。
- 2) 言語使用に影響を与える変数が統制できないことより、体系的にこれらの変数の影響を観察することができないため。
- 3) 学習者は、最も自信のある表現しか使わないので、自然言語データは学習者の言語能力の全体像を明らかにすることはできないため。

しかし、上記の問題点が指摘されているにせよ、枠組みのしっかりした規模の大きな学習者コーパスであればこれらの批判にはある程度対応できる。この点については後述する。したがって、「自然言語データ」に基づいて研究を行う意義は大きい。なにしろ、「自然言語データ」を使用することの大きな利点の一つは、学習者が目標言語でコミュニケーションを行う際に、目標言語の知識を使用して実際に産出しているデータを入手することができるということである (Ellis 1994 : 671)。

しかし、注意しなければならないのは、コーパスは純粋な「自然言語データ」とは言えない場合もあるということである。たとえば、外国語学習環境において日常生活で目標言語が使用されていない場合、そもそも目標言語で会話を行うことは「不自然」である。また、会話のトピックやタスクが制約されている場合も、完全な「自然言語データ」であるとは言えない。つまり、コーパスの構築環境によっては、「自然言語データ」と「臨床的抽出データ」の中間として、コーパスデータが分類されることもあるだろう。また、学習者コーパス内のデータに基づいた学習者の能力に関する分析結果は推論にすぎないと言う

点も注意すべき点である。つまり、学習者のパフォーマンスのみが調査可能であり、学習者の知識や能力について調査することはできない。ある特性が見られた場合、それが学習者の能力の現れであるのかどうか、学習者が意図したものを正しく発することができなかった時、そのことに学習者自身が気づいているのか、疑問に思っているのかどうかについては、学習者コーパスデータからは断定することができない。また、学習者コーパスデータ分析では、産出面に関する分析を可能にするが、受容面については分析を行うことができない。そして、学習者コーパス分析は誘因テストと比べて問題の焦点が鮮明ではない。研究者が興味のある特性の使用について、コーパスに十分にその特性が含まれないことにより、有効な議論を行えないこともありうる(Nesselhauf 2005: 41-42)。そのような不確かさを排除するためには、コーパスで何を分析することができ、何ができないのかについてを知り、コーパス分析において正しい測定方法を選択する必要がある。

#### 1.4.2. 学習者コーパスの枠組み

Granger (2002 :10-11)は、既存の学習者コーパスの分類に関連する 4 つの二項対立を紹介している。

Monolingual	⇔	Bilingual
General	⇔	Technical
Synchronic	⇔	Diachronic
Written	⇔	Spoken

4 つの二項対立のうち、左側のものが現在の学習者コーパスの主流である。学習者コーパスの多くは、単一言語によるものであり、バイリンガルコーパスである翻訳コーパスはあまり構築されていない。また、既存の学習者コーパスのうち、ビジネスコーパスなど、ある目的に特化されたものは少ない。そして、縦断的データに基づくコーパスも少なく、ある一定の期間に限定したデータ収集を基にしたコーパスが多い。最後に、話し言葉よりも書き言葉を対象とした学習者コーパスの方が多い。

既存の学習者コーパスを以上の観点から分類できるということは、学習者コーパスは少なくとも上記のような観点から構築されているということである。学習者コーパスデータに基づく研究を行うには、研究目的に応じてあらかじめ明確な枠組みを決定しておく必要がある。なぜなら、学習者言語には様々な変数があるからである(Granger 2002 :9)。

枠組みに用いる基準として、Tono (2003: 800)は 3 つの基準とその観点を詳細に提示している。

##### 1) 言語に関連する基準

モード(書き言葉/話し言葉)、ジャンル(手紙/日記/フィクション/エッセイ)、スタイル(物語/議論型)、トピック(一般的/レジャー)

2) タスクに関する基準

データ収集(横断的/長期的)、誘出法(自発的/事前準備あり)、レファレンスの使用(辞書/資料)、時間制限(固定/自由/宿題)

3) 学習者に関連する基準

内的認知(年齢/認知スタイル)、内的感情(動機/態度)、母語背景、第二言語の環境(ESL/EFL、学校のレベル)、第二言語の熟達度(標準テストのスコア)

Types of feature		
language-related	task-related	learner-related
mode [written/spoken]	data collection [cross-sectional/longitudinal]	internal-cognitive [age/cognitive style]
genre [letter/diary/fiction/essay]	elicitation [spontaneous/prepared]	internal-affective [motivation/attitude]
style [narration/argumentation]	use of references [dictionary/source text]	L1 background
topic [general/leisure/ etc]	time limitation [fixed/free/homework]	L2 environment [ESL/EFL]/ [level of school]
		L2 proficiency [standard test score]

(Tono 2003: 800)

これらの 3 つの基準について、コーパス構築時にあらかじめ明確にしておくことで、研究目的を最も反映するコーパスを構築することができる。

1.4.3. 代表的な学習者コーパス

英語学習者コーパスのうち、商業目的のコーパスとして有名なのは、*Longman Learners' Corpus* と *Cambridge Learner Corpus* である。*Longman Learners' Corpus* は、11 の母語背景を持つさまざまなレベルの学生によって書かれたテキストで構成され、約 1000 万語規模となっている。学習者が授業内に書いたもの、テストで書いたもの、辞書を使ったもの、辞書を使わなかったものなど、テキストの性質は様々である。このコーパスは辞書編纂者、教材開発者などによって有意義な情報を提供している (McEnery et al. 2006: 66)。*Cambridge Learner Corpus* は、世界中の英語学習者の書き言葉を 2000 万語以上集めたもので、規模は拡大中である。このコーパスは *Cambridge ESOL English* という試験を受けた学生の試験答案で構成されている。そのためこのコーパスは、*Cambridge University Press* の関係者および *Cambridge ESOL* の関係者のみに使用が限定されている (McEnery et al. 2006: 66)。

教育機関によるコーパスで最も広く知られているものは、ベルギーのルーバンカトリック大学の Sylvianne Granger が率いる *Centre for English Corpus Linguistics (CECL)* という研究チームが構築した *ICLE (International Corpus of Learner English)* という学習者の書き言葉コーパスであろう。*ICLE* の学習者の特性はほぼ均一である。これらの学習者は非英語圏で英語を専攻する 20 歳前後の大学 2 年生から 4 年生までの学生である。レベルはおおよそ上級に分類できる (Granger 2003: 539)。男女比はおおよそ 2 : 8 で女性学習者が大半を占めている。タスクのトピックのほとんどが論証文であり、一つのトピック

につき文の長さの平均は 700 文字程度である (Granger 2003 : 540)。現在入手可能なデータは、ブルガリア語、中国語、チェコ語、ドイツ語、フィンランド語、フランス語、ドイツ語、イタリア語、日本語、ノルウェー語、ポーランド語、ロシア語、スペイン語、スウェーデン語、ツワナ語、トルコ語の計 16 の母語の学習者データで、370 万語である<sup>18</sup>。この学習者コーパスの参照コーパスともいえるのが、イギリスの高校生や大学生、そしてアメリカの大学生による論証文で構成されている *LOCNESS* である。また、*ICLE* は 1995 年から話し言葉コーパスである *LINDSEI (Louvain International Database of Spoken English Interlanguage)* の構築および研究も行っている。このコーパスは、11 の異なる母語をもつ学習者によって産出されたおよそ 80 万語のデータから成る (Gilquin et al. 2010)。

また、ある一定の母語話者に限定した学習者コーパスもある。たとえば、香港科学技術大学 (HKUST) が構築した *HKUST Corpus of Learner English* である。これは、中国人英語学習者の書き言葉を集めたもので、約 1000 万語から構成されている (McEneaney 2006 : 66)。

日本語学習者に特化した学習者コーパスも存在する。日本では学習者コーパス研究は比較的早い時期から行われてきた。その草分け的な学習者コーパスであるのが、*JEFLL (Japanese EFL Learner) Corpus* である。このコーパスは、日本人英語学習者の異なる学習段階の自由英作文データを、できるだけ広範囲に大量に収集・コーパス化することを目的としている (投野 2007 : 7)。このコーパスに含まれるデータは、日本の中学生と高校生の英語学習者であり、2007 年 4 月までに約 67 万語を公開している (ibid : 7)。

現在、学習者コーパスの主要プロジェクトでは、上級者のデータ収集が主に行われている (投野 2007 : 9)。上級英語学習者による書き言葉コーパスは、SLA の分野において確固たる地位を築いたことには疑いの余地はない。上級者レベルでは、特に難易度の高い語彙、談話、単語、語用を扱う必要があり、上級学習者を受け持つ大学の教師にコーパス研究成果は大きな影響を与えている (Myles 2005 : 388)。一方、海外の学習者コーパスから、日本人の英語力のレベルが低いことが分かる。日本人学習者について知りたいと思う語彙や文法の発達過程は、海外のデータからはほとんど分からない (投野 2007 : 10)。そのため、*JEFLL Corpus* のような日本人学習者に特化したコーパスは貴重であると言える。

日本語学習者コーパスについても 2000 年代に入ってから様々なコーパスが構築された。それらの多くは研究目的のコーパスである。たとえば、「日本語学習者による日本語作文と、その母語役との対訳データベース」は、国立国語研究所の宇佐美洋氏が中心となって開発を行ったものである<sup>19</sup>。日本語学習者が書いた 200 字から 800 字程度の日本語作文と、その作文の対訳、そして日本語作文の添削の 3 つから構成されている。学習者はアジア 7 カ国の出身者である。

また、「上級学習者の日本語作文データベース」と「日本語学習者言語コーパス」<sup>20</sup>は、東京外国語大学の海野多枝氏が中心となり開発が行われたものである。「上級学習者の日

<sup>18</sup> <http://www.uclouvain.be/en-cecl-icle.html> 参照

<sup>19</sup> [http://www.ninjal.ac.jp/publication/catalogue/kokken\\_mado/04/05/](http://www.ninjal.ac.jp/publication/catalogue/kokken_mado/04/05/) を参照した。

<sup>20</sup> <http://cbll.tufts.ac.jp/llc/ja/index.php?menulang=ja> を参照した

本語作文データベース」は、東京外国語学部日本語専攻の留学生で、日本語検定一級を持つ上級日本語学習者の約 150 名が執筆した作文をデータベース化し、執筆者から承諾の得られた 80 名分の 326 の作文をデータ収録したものである。作文の主題は、「友人紹介文」、「自己紹介文」、「意見文」、「描写文」、「レポート」の 5 種類であった(鈴木・海野 2006)。「日本語学習者言語コーパス」は、海外の教育機関で日本語学習者による書き言葉のデータ収集を行ったものである。e ラーニングを活用した作文データ収集によるコーパス構築を行うことと、作文データ収集およびオンライン誤用辞書を開発することの 2 つの活動が実施された。データは、台湾、イギリス、ウクライナで収集された。また、作文データにパラレルデータとして、日本語母語話者のデータも収集されている。データ収集の際、被験者は機能タスクと日記タスクを行った。e ラーニングを活用した作文データの規模は 87000 字、誤用辞書のための作文データの規模は約 20 万字である。

フランス語学習者コーパスは、英語や日本語と比べると数は圧倒的に少ない。フランス語学習者コーパスの一つとして、*The Lund CEFLE Corpus (Corpus Ecrit de Français Langue Etrangère)*が挙げられる。これは、スウェーデンでフランス語を学習する高校生 85 人による書き言葉テキストのコーパスである。このコーパス構築の狙いは、書き言葉の産出における形態統語的発達を分析することである。コーパスは、習熟度の異なるスウェーデン語を母語にするフランス語学習者が書いたテキストと、統制グループとしてフランス語母語話者が書いたテキストと 2 タイプのテキストで構成されている。学習者は英語の上級学習者でもあり、フランス語は学習者の第二、第三言語にあたる。2003・2004 年度を通しておおよそ 400 テキストが集められ、総語数は約 10 万語にのぼる。1 年の間で、2 ヶ月おきに各学習者はフランス語で 4 つのテキストを書いた。なお、第一回のライティングタスクを行った際、学習者のフランス語学習歴は 1 年であった。統制群の 22 名のフランス語母語話者は、スウェーデン人学習者と同年代で、学習者と全く同じタスクを行った。タスクは、全参加者からできるだけ自然なデータを引き出せるよう、設計されている。連続する絵をもとにして描写するタスクと、個人の経験をもとにした作文タスクの 2 種類のタスクタイプがある。描写タスクには、「L'homme sur l'île」と「le voyage en Italie」という題材を使用し、作文タスクには「Moi, ma famille et mes amis」と「un souvenir de voyage」というテーマを与えた。学習者はプレーンテキスト形式を使用して、パソコンで作文を書いた(Granfeldt et al. 2006: 565)。Granfeldt et al. (2006)は、イタリア旅行に関する作文タスクを基に学習者の横断的言語分析を行い、*Direkt Profil(DP)*を使用して学習者を分類した。DP とは、形態統語的観点によって stage 1 (初級)、stage 2 (初級後期)、stage 3 (中級)、stage 4 (上級)と 4 つにレベル分けされた評価基準から学習者が発するテキストの習熟度を診断するシステムである。DP は、テキストのトークン化、句や文の文法注釈の付与、レマや品詞などのタグ付与、学習段階に特徴的な構造数のカウントなどの機能を持つ。



#### 1.4.4. 話し言葉の特徴と評価

話し言葉は、20世紀初頭から一部の言語学者にとっての興味の対象であったが、本格的に話し言葉の記述が行われるようになったのは1960年代からであった(Gadet 1996: 14)。

第二言語学習者の書き言葉コーパスは、話し言葉コーパスと比べて入手することが容易である。この理由の多くは、母語話者コーパスの場合と同様に、書き言葉コーパスの構築がより簡単で低予算で済むことに依拠している。特に学習者コーパスの場合、書き言葉コーパスの多くは学習者のエッセイで構成されており、その大半は学習者が授業の一環として行う課題である。そしてそれらの課題は、始めから電子フォームで行われる場合が多い。一方、学習者話し言葉コーパスはまだまだ数が少ない(Myles 2005: 375)。

ここ数十年の間で、どの外国語教育においてもコミュニケーションのための言語が重要視されるようになってきた(McCarthy and Carter 1997: 38)。教育においてコミュニケーションを重視する理由は、学習者が書き言葉を通じた知識や文化経験のためだけに外国語を学ぶのではなく、会話におけるインタラクションを目的として外国語を学ぶ必要性が、20世紀後半から高まったためである。第二言語習得理論では少なくとも、言語教師の中心的役割は言語に関する知識を与えることから、タスクを組み立てることや学習機会を与えることへと移行している(Kennedy 1998: 281)。このように、話し言葉は語学教育において重要性を増すようになった。言語教育を改善するために、学習者が実際に外国語で行う口頭産出能力について研究することは、極めて重要なことである(Nesselhauf 2005: 41)。

実際に母語話者による言語産出では、書き言葉と話し言葉は異なる特徴を持つ。書き言葉コーパスである *Cambridge International Corpus* と話し言葉コーパスである *CANCODE* を比較すると、書き言葉の語彙リストは代名詞、限定詞、前置詞、法動詞、助動詞、接続詞などの機能語によって構成されていたが、話し言葉コーパスの語彙リストには機能語だけではなく内容語も多く含まれていた。また、話し言葉コーパスの語彙リストに見られる内容後のいくつかは、*you know* や *I think* などの対人マーカーか、もしくは *well* や *right* のような一語のディスコースマーカーであった(McCarthy and Carter 1997: 24)。さらには、書き言葉では三人称が優勢であるが、話し言葉では一人称と二人称を多用する(Andersen 2006: 57)。

また、Beeching (1997: 384-387)は、フランス語の母語話者コーパスの分析の結果、フランス語の話し言葉の特徴として、以下の5つの点を挙げている。

- 1) 句+c'est、もしくは句+c'est +関係代名詞のような分裂文によって、トピックを強調する  
例：un des gros problèmes ici, c'est le vent
- 2) 関係代名詞と on a, il faut を伴う分裂文によって、トピックを強調する  
例：il faut un minimum qui est 15 à 20 centimètres
- 3) 前方照応代名詞を伴う分裂文によって、トピックを強調する  
例：ce taux, on le monte

- 4) *c'est-à-dire (que)*、*ce qu'on appelle*、*disons (que)*などを用いて、オウム返し、ためらい、冗長性の高い文を発する  
例： *on n'arrive pas à trouver disons les causes exactes*
- 5) *on* や *ça* の多用や、否定文の *ne* を省略など、「非標準的フランス語の文法特徴」(Lambrecht 1981 :6-7)が見られる

話すという行為は日常的に行われているため、我々はこのごく当たり前の行為であるとみなしている。しかし、母語でも外国語でも口頭産出するためには、なぜ、どうやって、いつコミュニケーションを行うのかという微妙で細かい知識を増やさなければならず、また質問をしたり、ターンをとったりといった、インタラクションをうまく遂行するために必須の複雑な技術を高めることも求められる (Burns and Seidlhofer 2002: 211)。会話は一般的に事前に準備されるものではない。長いポーズや短いポーズ、繰り返し、不完全な文法構造などに見られるような、認知的、プロセス的な制約下のもとで行われる (Aijmer 2004: 174)。外国語で会話を行うには、このような外国語の口頭産出にまつわるサブスキルとして、レンジや種類を適宜選択する必要があり、これは困難なことである。実際、外国語における話し言葉の社会語用論的規範に関する教授法では、コミュニケーションストラテジー、談話の組み立てと構造、*small talk* などの日常会話、会話の決まり文句、発話行為(要求、棄却、賛辞、質問)なども組み込まれている (Hinkel 2006: 114-116)。

第二言語における口頭産出の評価は、言語テスト分野においても大きな関心が寄せられている (Lorenzo-Dus and Meara 2005)。スピーキングテストで現在主流となっているテスト形式はコミュニケーションテストであるが、テスト受験者の評価は一般的に記述的評価基準に従って主観的に行われている。評価基準とは、語彙エラー数や語彙使用のレンジなどテスト受験者によるパフォーマンスの量的観点と、表現の明確さや語彙選択の適切さなどの質的観点とを組み合わせたものが多い。したがって評価者は、各受験者のパフォーマンスの評価が信頼性の高いものとなるよう、スピーキングテスト時に極めて複雑な仕事を行う必要がある (Read 1997: 310-311)。しかし、実際スピーキングテストは主観的な印象によって採点されることも多い。そこで、語彙に基づく採点方法が注目を集めている。語彙に基づく採点は客観性を高めるため、語彙は試験でより重要な役割を果たす (Milton 2009: 251)。さらに、言語能力における語彙の占める役割の高さが明らかになり、近年では語彙に対する興味が高まっている。

語彙の側面に限定しても、学習者の語彙能力に関して、重厚な情報を引き出すようなテストを作成することは極めて難しい。信頼性と妥当性があるテストを作るために、Milton (2009: 20)は考慮すべき2つの大きな問題を挙げている。一つ目は、語彙の測定や試験などにどの単語を選択するかという問題である。二つ目は、学習者がただ当該の単語を知っているのか、それとも学習者はその単語を使用することができるのかを確認するために、どのような手法を使用するかという問題である。一つ目の問題については、研究者の間であるコンセンサスにたどり着こうとしている。それは、語彙テストやその他のテストで語彙の頻度データを用いることである。ところが、二つ目の問題に関してはコンセンサスが得られていない。なぜなら、語彙の異なる側面や知識をテストするには、異なる手法が必要となるからである。特に、受容知識と産出知識のテストには全く異なるア

アプローチが必要となる。なお、受容知識と産出知識についての詳細は、第二章 2.4.1. で述べる。

そこで、スピーキング時の語彙使用をできるだけ効果的に分析する方法として学習者コーパスが役立つ。母語話者コーパスとは異なり、厳密に言えば多くの場合学習者コーパスはいわゆる「自然発生的」なテキストであるとは言えない。この理由は、学習者コーパスが構築される外国語教育の環境のほとんどにおいて、学習者がこの種のテキストを発することは稀であるからである。しかしながら学習者コーパスとして成立するためには、コーパスに含まれるテキストはわずかな統制下のもとで発せられなければならない。多くの場合、書き言葉コーパス構築の場合はトピックのみが与えられるエッセイ形式であり、話し言葉コーパス構築の場合はインタビューがいくつかのトピックを与えるインタビュー形式が使用される (Nesselhauf 2005: 40)。また、自然会話形式の学習者スピーキングコーパスも少なくない。自然会話を収集する際、学習者は答えを考えるために会話を中断することはできず、実際のコミュニケーションで起こりうるストレスと同様のストレスを受ける (Ovtcharov 2006 :120)。この点から、話し言葉学習者コーパスとスピーキングテスト、さらに自然会話との類似性が伺える。

#### 1.4.5. 学習者コーパスの利点と欠点

中間言語はそれ自体多様性があり、他変種との比較を行わずとも中間言語について研究ができる。しかし理論的、実践的目的により、中間言語と他の言語変種を比較することによってその特殊性を明らかにすることは有効である。コンピュータ学習者言語コーパスを扱う分野では、この比較アプローチは比較中間言語分析 (CIA: Contrastive Interlanguage Analysis) と呼ばれており、2つの比較方法を提唱している。母語話者と外国語学習者との比較と (L1 vs. L2) と、異なる中間言語の変種間の比較 (L2 vs. L2) である (Granger 2004: 127-128)。前者は、母語話者データから学習者データを弁別するような使用パターンを明らかにするのに有効である。この使用の差は、エラーのような質的な相違と過剰使用や過少使用のような量的な相違の 2つのカテゴリーに分けることができる (Granger 2003: 541)。後者は、レベルの違う学習者の比較や、母語が異なる学習者間による比較を行うことができ、横断的研究に有効である。特定のグループやレジスターに最も際立った特徴を自動的にプロファイリングすることによって、研究者はある学習者グループの中間言語の全体像を素早く手に入れることができる。さらに、今後の研究へ興味深い示唆を与える (Granger and Rayson 1998)。

しかし、中間言語は母語に比べてとても可変的である。様々な言語的、状況的、心理言語的要因によって影響を受けてしまい、中間言語研究における研究結果の信頼性を大きく低下させてしまう原因の一つにもなる。しかし、これらの要因を完全に統制することはできない (Granger 2004 : 125)。

#### 1.4.6. フランス語学習者話し言葉コーパスプロジェクト

英語学習者話し言葉コーパスに関するプロジェクトは、1990年代から始まっている。たとえば、ベルギーのルーバンカトリック大学の Sylvianne Granger が中心となって開発が進められた *Louvain International Database of Spoken English Interlanguage (LINDSEI)* コーパスは、1995年から構築が始まっており、学習者話し言葉コーパスの草分け的存在であると言える。これは、世界中の大学との共同プロジェクトであり、日本語を含む 20 言語の学習者によるサブコーパスが存在する<sup>21</sup>。また、日本人英語学習者の話し言葉コーパスに特化したものでは、*NICT JLE* コーパスがある。これは、独立行政法人・情報通信研究機構が主体となって作成したものであり、一対一の対面式インタビューテストである *Standard Speaking Test (SST)* の受験者 1281 名のデータが収録されている。このコーパスは約 200 万語の規模で、SST の評価結果が付加された能力別サブコーパスから構成されている(和泉他 2004)。

また、日本語学習者話し言葉コーパスも構築が進められている。上村隆一氏は 1991 年より日本語会話コーパスプロジェクトを始めた。このコーパスでは、ACTFL(American Council on The Teaching of Foreign Languages)の OPI(Oral Proficiency Interview)に準拠したインタビュー形式に則って学習者の言語データが収集されている。学習者は主にアメリカ人と韓国人である。また、「KY コーパス」は、鎌田修氏と山内博之氏が中心となって構築したコーパスであり、90 名分の ACTFL の OPI の試験テープを文字化したものである。学習者の母語は、中国語、英語、韓国語で、それぞれ 30 名ずつ抽出され、母語ごとに初級 5 名、中級 10 名、上級 10 名、超級 5 名で構成されている<sup>22</sup>。

上記のコーパス以外にも、英語やフランス語の話し言葉コーパスは現在までに数多く構築されている。フランス語学習者コーパスについても、構築が進められている。ここでは、主なフランス語学習者による話し言葉コーパスプロジェクトを概観する。

##### 1.4.6.1. French Learner Language Oral Corpora

Southampton 大学の研究チームは *French Learner Language Oral Corpora (FLLOC)* という電子データベースを構築している。ここに含まれるデータは、イギリス教育システムの 9 年生から 13 年生、すなわち 13 歳から 18 歳までのフランス語学習者に関する横断的研究データで、異なる習熟度レベルを代表するフランス語学習者の中間言語コーパスである。このコーパス構築の目的は、第二言語としてのフランス語の全レベルの代表させることである。参加者は 9 年生までで既に 150 時間のフランス語教育を受け、13 年生までの学習時間はおよそ 600 時間に達する。タスクは、学習者とリサーチチームの研究者による一対一のセミガイド式インタビューで、全てフランス語で行われた。タスクは二部構成となっており、前半は若者たちが様々なアクティビティをしている 2 枚の写真を見て、学習者が写真についてできるだけ多くの情報を挙げていく形式のものである。後半は、

<sup>21</sup> <http://www.uclouvain.be/en-cecl-lindsei.html> を参照

<sup>22</sup> [http://opi.jp/shiryo/ky\\_corp.html](http://opi.jp/shiryo/ky_corp.html)

研究員が学習者に最近興味があること、家族について、そして上級者には過去のバカンスについて、また将来についての展望について訊ねた。なお、このコーパスデータは研究機関向けに公開されている<sup>23</sup>。コーパスはデジタルオーディオファイル、CHILDES によってフォーマット化された転写、そして品詞タグが付与されたファイルから成っている。全転写とオーディオファイルはダウンロードすることができる。これらのデータに基づき、これまでに形態統語的、語彙的な発達側面や、決まり文句の使用、モノログとダイアローグの談話の発達側面など、様々な問題について分析が行われている (David 2008, Rule 2004)。

#### 1.4.6.2. Progression in Foreign Language Learning

*Progression in Foreign Language Research Project* は、1993年から1996年までの3年にわたってイギリスの2つの中等教育機関で行われていたプロジェクトである。フランス語学習者60名について、初年度である7年生から3年目である9年生までの3年にわたって長期的研究を行った。学習者は3年間のあいだ、6つの口頭産出タスクを行った。学習者同士でペアになって行うタスクもあれば、研究者と行うタスクもある。学習者の発達がより明確になるよう、いくつかのタスクは時期をずらして同じものが複数回行われた。このプロジェクトデータはアナログ録音で計200時間、プレーンテキストで650の転写ファイル数となった (Rule et al. 2003: 669)。5年後に新たに立ち上げられた *Linguistic Development in Classroom Learner of French* という研究プロジェクトは、*Progression in Foreign Language Research Project* を基にしたプロジェクトである。このプロジェクトは Myles によって指揮され、イギリスの経済社会研究委員会 (Economic and Social Research Council) によって支援された。イギリスの中等教育における9年生から11年生までのフランス語学習者が教育活動を通してどのように言語発達をするのか、またどのような経験を学ぶのかについて文書化することを目的としている。さらに、非分析的な発話の機械的学習を行う教室の役割についても検証することも目的である。また、前述の *Progression in Foreign Language Research Project* で集められた7年生から9年生までの話し言葉中間言語データを拡張することも狙いの一つである。学習者の数は、それぞれ9年生から11年生までの各20名であり、性別や学力は均整がとれている。各学習者は4つの口頭タスクを、母語話者および母語話者レベルの話者と一対一のインタビュー形式で行った。各学年間でのパフォーマンスが比較できるよう、全ての被験者は全く同じタスクを行った。そのうちの3つは *Progression in Foreign Language Learning* で使用したタスクと同じものであり、前プロジェクトのデータとも比較可能である。このプロジェクトで集められたフランス語の話し言葉の録音は約50時間であり、前回のプロジェクトと合算すると計250時間にも及ぶ。2つのプロジェクトを合わせることで、中等教育の初頭から中等教育修了試験までどのようにフランス語が発達していくかについての全体像を明らかにすることができると考えられている (Myles et al. 1998: 329, Marsden et al. 2002: 3-4, Rule et al. 2003: 670-675, Rule 2004: 354)。これらのコーパスを用いて、中級学習者

---

<sup>23</sup> <http://www.floc.soton.ac.uk>

の常套語の使用と学習者言語システムの発達における常套語の役割について(Myles 2005a)、動詞句の発達(Myles 2005b)、初級者の否定文の使用の発達(Rule and Marsden 2006)、談話管理と話術の発達(Myles 2003)などの研究も行われている。

#### 1.4.6.3. Langue et communication (LANCOM)

Leuven 大学のフランス語学研究チームは、1994年から2001年にかけて *LANCOM (Langue et Communication)* というコーパスの構築を行った。これは、1993年の連邦制への移行といったベルギーの政治的背景と関連したプロジェクトである。フラマン語共同体では、当時フランス語教育の必要性が再考されていた。そこで Leuven 大学は新しい教育プログラムで行う言語教育に有効なデータ研究について教育機関の了承を得て、教育と研究の双方で利用できる中間言語コーパス構築に着手することを決定した。このコーパスの特徴は、教育への応用が第一義的なものとされていることである。したがって、コーパス構築の目的は、新しいマニュアル開発の基盤となるデータの収集と、教室活動で利用できるような様々なロールプレイングのタスクに基づくコーパスを作成することであった。タスクは録画および録音されている。場面は、電話による会話、仕事の面接、招待、海外旅行の計画、ホテルの予約、航空券の予約など多岐にわたっている。また、フランス語母語話者も学習者と同じタスクを行い、それを録画した。加えて、フランスでの日常生活の場面も録画した。したがって、大半は一つの場面につき、母語話者によるバージョンと学習者によるバージョンの2種類の録画が存在する。このアプローチには2つの利点がある。一つ目は教育上の利点で、フランス語母語話者による実際の日常会話を含むコーパスであることより、教材開発者に有用な話し言葉フランス語のデータを提供できる点である。二つ目は大学の研究に関する利点で、学習者のエラーやぎこちなさを研究することができる点である。このように中間言語研究が目的となっているため、言語エラーに関するタグ付けも考案され、様々なタグが既に付与されている。1994年以降、データは断続的に追加されており、39時間分の録音が完了しており、19時間分に相当する160593語が既に転写されている(Debrock et al. 1999:48, Flament-Boistrancourt 2001: 5, Mertens 2002: 4)。そして、このコーパスに基づく研究も行われている。このコーパスによって非母語話者のエラーやぎこちなさなどを言語学的に発見することができ、フランス語におけるインタラクションの働きについてこれまであまり研究されてこなかった点について焦点を当てることができた(Debrock et al. 1999:46)。

なお、前述の *ELILAP* コーパスと *LANCOM* を Mark Debrock が再構築したものが *ELICOP* である。情報技術の発達とインターネットの普及によってこれらのリソースの大半を公開することが可能となった。*ELICOP* は <http://bach.arts.kuleuven.be/elicop> よりアクセスすることができる。

#### 1.4.6.4. InterPhonologie du Français Contemporain (IPFC)<sup>24</sup>

このプロジェクトは、コーパスに基づいた中間言語の音韻研究を主たる目的として発足された。フランス語の中間音韻研究に適した話し言葉コーパスは特に少ない。コーパスに基づく第一言語の音韻研究は、前述の *PFC* プロジェクトによってかなり前進したと言える。*PFC* プロジェクトが誕生した 10 年後の 2008 年に、非母語話者を研究対象にする *PFC* のサブプロジェクトとして *IPFC* プロジェクトが発足された。

*IPFC* プロジェクトの目的は、外国語としてのフランス語研究に向けたコーパスを構築し、公開することである。タスクに関しては、同じ母語を持つ学習者には同じプロトコルが使用され、様式は学習者に適応されているが、*PFC* プロジェクトに近いプロトコルが用いられている。*IPFC* プロジェクトの独創的な面として、以下の 5 つの点を挙げるができる(近藤・川口 2008: 54)。

- 1) *PFC* のツールを利用し、話し言葉フランス語の統語的・語彙的情報を収集しながら、学習者の言語能力、とくに彼らの音声と音韻に焦点を定めている。
- 2) 諸地域の学習者データを比較することができるように、調査項目に一貫性を与えている。
- 3) 既存の *PFC* データベースを参照することで、単一言語および多言語環境における言語変異を比較研究することが可能である。
- 4) *PFC* で培われたネットワークを継承し、研究協力者の国際的なネットワークを構築している。
- 5) 中間音韻論としての研究的側面だけではなく、*PFC* 教育プロジェクトとしての教育的側面における探求も行う。

データ収集の際には、5 つのタスクを行った：

- 1) 母語話者によって発音された単語を聞いて、発音する。学習者全体に共通する単語リストと、学習者の母語によって問題となる単語リストの 2 つから成る。
- 2) 上記のリストと、*PFC* で使用した語彙リストの音読
- 3) *PFC* で使用したテキストの音読
- 4) 学習者のレベルや、学習背景に応じた閉ざされた質問と開かれた質問から成る方向付けられたインタビュー
- 5) 学習者同士による自由会話

*IPFC* プロジェクトは、日本語母語話者とスペイン語母語話者の学習者のコーパス構築から始まった。なお、日本語母語話者であるフランス語学習者を対象とした体系的なコーパスは、この *IPFC* コーパスが唯一である。

現在までに、日本語とスペイン語母語話者によるサブコーパスの他に、ドイツ語、カナダ英語、ギリシャ語、イタリア語、オランダ語、ノルウェー語、ブラジルポルトガル語、

---

<sup>24</sup> 特記しない限り、*IPFC* プロジェクトについては Racine et al. (2012)を参照した。

スウェーデン語、トルコ語を母語に持つ学習者のサブコーパスが構築されている<sup>25</sup>。IPFCのデータを用いて様々な研究が行われているが、日本人学習者データに基づく研究としては、音韻論的分析 (Marushima et al. 2010, Racine et al. 2010)、語彙的分析 (Sugiyama 2011、杉山 2011a、杉山 2011b)などが行われている。

### 1.5. コーパスの利点と欠点

1960年代以来、コーパス分析やコーパスの記述によって、語彙、文法、談話に関する多くの事実が明らかになり、話し言葉や書き言葉のレジスターや様々なジャンルにおける多くの言語的特徴についても新たな発見をもたらした (Kennedy 1998: 288, Hinkel 2006: 112)。

コーパスが登場して大きな変化が起こった点は、コンピュータを利用した分析が可能となったことである (Myles 2005: 376)。このことによって、テキスト処理を行うソフトウェアを使用でき、言語分析の一部を自動的に行うことができるようになった (Altenberg and Granger 2001: 176)。そして手動で行う言語分析よりも正確に、一貫性を持ち、そして何よりも素早く分析結果を提示できるようになった。また、コンピュータ化の一番の意義は、データを広く配信できることである。データ収集者のみならず、別の研究者もデータを共有することが可能となり、研究結果を検証したり比較したりすることも可能となった (Nesselhauf 2005: 41)。

また、Leech (2011 :7)はコーパスの利点について、以下のように述べている。

If asked what is the one benefit that corpora can provide and that cannot be provided by other means, I would reply 'information about frequency'.

(コーパスが与えることができ、別の手段によっては得られないものは何かとも尋ねられたら、私は『頻度に関する情報』と答えるだろう。)

頻度情報によって、伝統的言語学で行われてきた内省法や勘から引き出すことができなかつた観点が明らかになった。言語学は、言語の性質、要素、構造、機能についての理論に関する証拠や、言語で何が可能であることを言明するための基盤としての証拠を常に必要としてきた。多くの場合、勘や内省、実験や誘出法、そして話し言葉や書き言葉のテキストの言語生起の観察に基づいた記述からこのような証拠は導きだされてきた。一方、コーパスに基づく研究の場合、証拠は直接テキストから引き出すことができる。この意味でコーパス言語学は証拠を勘に頼る言語アプローチと異なっている (Kennedy 1998: 7)。このように、言語コーパスは実際の言語使用に触れることのできる最適の手段であるといえる (Jacques 2005: 24)。コーパス言語学の登場によって、自然発生的言語の規則の研究について、大規模かつ一定の原理に基づく経験的データに基づいて行うことができるようになった (Conrad 2000: 548)。自然発生的言語は、小規模の分析では見過ごされてきたが、コーパスに基づく分析によってより客観的に様々なパターンを分析できるようになった

<sup>25</sup> <http://cblle.tufs.ac.jp/ipfc/> 参照



(Baker 2004: 346)。つまり、実際に行われている言語使用の観点に再注目することになった。そしてそのことから、言語観察事象や言語記述を文脈の中に置き直すことが可能になった (Jacques 2005: 29)。具体的には、あるコンテキストにおいて頻度が高い語パターンや低い語パターンの特定や、レジスター別の特定の語句への依存についての記述などの言語使用の経験的記述である (Biber and Reppen 2002: 207)。Jacques (2005: 25-26)は、コーパスに基づく研究の利点として、以下のようにまとめている。

- 1) 勘で見落とされる言語の働きについて焦点を当てることができる
- 2) 言語の働きに関する勘を訂正できる
- 3) 言語事象について頻度の観点から示唆を得て、統計的観点から実証できる
- 4) 言語変異に触れることができ、認識できる

しかしコーパス構築で注意しなければいけない点がある。まず、コーパス構築条件により、研究結果が変わることがあるということである (Cori et David 2008: 125)。つまり、ある現象を分析する際に使用するコーパスが異なれば、導きだされる結果も異なる可能性がある。この問題は経験データを用いる全ての研究に該当するものである。

また、多かれ少なかれ、言語学者は言語事象の性質に影響を与えうるテキストをあらかじめ選択しているが、言語学者がコーパスを構築する際に言語事象についての見解や記述について言語学者は直接介入してはならない。この点は、内省的言語学へ向けられた批判と類似していることを強調しなければならない。内省的言語学は、勘によって言語例をフィルターにかけ、選択を行ってきた (Jacques 2005: 26)。しかし、コーパスに基づく研究の場合、コーパス構築時に被験者の代表性に注意することで、問題は緩和される。また、コーパスの規模を拡大することで、あるコーパスがもつデータの偏りなどや特異性のある程度解消できると考えられる。

伝統的に、コーパス研究分野はあまり資金援助されてこなかったため、コーパス収集にかかる費用の実負担が増えることも多く、コーパスの規模も拡大されず、コーパスのテーマも限定的であることが多い。また、このように困難を極めて収集されたコーパスデータは、研究コミュニティの全てに解放されているとは限らない (Myles 2005a: 375)。さらに、入手可能な学習者コーパスは研究には不十分なものが多い (Nesselhauf 1998: 40)。

コーパスの登場によって、言語分析に新たな観点が加わった。コーパスが構築され始めて50年以上が経過し、個々のコーパス規模も拡大されたり、コーパスの種類やコーパスの数も増加したりした。しかし一方で、構築されたコーパスは共有されにくく、研究も独立して行われる傾向が強い。

## 1.6. まとめ

本章では、まずコーパス言語学についてその主要概念をまとめた。コーパス言語学の発展を概観した結果、コーパス言語学は英語学を中心として発達してきたことが分かった。そして、現在では英語学だけに限らず、日本語学やフランス語学など他の言語研究にもコーパスが使用されている。また、コーパスへの興味は第二語習得の分野まで広がりが見ら

れることが分かった。英語学習者コーパスはもちろんのこと、日本語学習者コーパスやフランス語学習者コーパスも次々と構築が行われていることについて、認めることができた。しかし、体系化された日本人フランス語学習者のコーパスは、*IPFC* コーパスを除いて存在が確認できなかった。

学習者コーパス構築には新たな中間言語についての発見の期待が込められている。そして、すでに学習者コーパスは言語教育学の分野に多大な貢献をしてきた。しかし、コーパス研究の成功を喜ぶには時期尚早であると Granger (2004) は指摘している。なぜなら、特定の目的のための英語 (English for Specific Purposes) や、長期データコーパスなどといった種類のコーパスがまだあまり構築されておらず、品詞タグやエラータグなどのより洗練された自動的コンピュータ処理も改善が必要となっているからである。また、コーパスは言語学における全ての問題を解決する道具ではない。コーパスの利点と欠点を十分に知った上で研究を遂行し、言語教育現場に還元していくことが求められている。

次章では、コーパス研究で核として行われてきた語レベルの研究を俯瞰する。

## 第二章 語レベルの研究 –Word Units 研究–

前章ではコーパス言語学について概観したが、コーパスの登場によって最も利益がもたらされたのは語彙論の分野であろう。なぜなら、語彙に関する頻度情報はコーパスから最も簡単に取り出せるからである。コーパス研究とコンピュータ技術によって、将来的には語彙記述がより正確で網羅的な情報を提供し、人間の観察だけでは分からなかった言語パターンを発見することができるだろうと Stubbs (1995: 21) によって約 20 年前に指摘されていた。多くの専門家が指摘する通り、第二言語習得の分野においても語彙への興味は、少なくとも研究者の間では高まっている (Thomas 2008: 588)。

### 2.1. 『一語』とは何か

語彙研究における最も難解な問いの一つとして、「一語とは何か」という問いが挙げられる。しかし、この問いの意図によって答えが変わるうえ、十分な解答も得られていない。語彙研究が累積的で比較可能なものとなるためには、研究者は調査を行う上で語における基準について少なくとも類似する基準を使用するか、完全に同じ基準を採用すべきである。(Bauer and Nation 1993: 265)。

そもそも、「語」という用語は曖昧な用語であり、研究対象を限定する用語や定義が使用されたり、意味の最小単位である「形態素」の方が研究対象として好まれていた。フランス語の場合、一語を文字の観点から定義すると、2 つの空白や句読点で区切られた文字であると定義できる。しかし、統語・意味論的観点から定義を行うと、意味を持つ自立的な言語単位である (Cavalla et al. 2009: 12)。このように、「一語」の概念は分析観点によって大きく異なる。

コーパスにおける語の数え方として一般的な方法は、トークン(Token)とタイプ(Type)である。トークンは、テキスト内の語の出現回数であり、タイプは、テキスト内で使用された語の種類の数である (Dewaele 1993: 93)。しかしタイプを使用する際、大文字と小文字を区別するのか、または同音異義語などを区別するのかどうか、などといった問題が残る (Nation and Meara 2002: 35)。もし、本の長さを知りたい場合や、一分間にどれだけ早く読んだり話したりできるのかを知りたいければ、トークンを数えればいい。しかし時に、学習者がどれだけ多くの語を知っているのか、またはどれだけ語を使用しているのかについて興味を持つこともある。たとえば、どれだけ学習者が早く正確に語を認識できるのかを知りたい場合は、タイプを数える必要がある (Nation and Meara 2002: 35)。

また、単語家族 (word family) も語彙の単位となりうる。しかし、単語家族の大きな問題は、何を家族として数えるかを決定することである。最も保守的な方法はレマ (Lemma) を数えることである。レマは、同じ品詞に属する語幹と語尾変化から成る語で構成される。また、基礎語と共通の意味を持つ派生形を単位に含むものもある (Hirsh and Nation 1992: 692, Nation and Meara 2002: 36, Read 2007: 108)。もし学習者が接辞に関する知識があれば、既知の基本形の語尾変化や派生接辞を既知の単語として認識することができるというのが、単語家族の背景にある概念である (Hirsh and Nation

1992 : 692)。また、Laufer and Nation (1995: 312)は、単語家族こそ学習者が接触してきた単語数に最も近い定義であると主張している。たとえば、リーディングでは単語家族が基礎語を構成しており、学習者は派生形や活用形を全て別々に習得していなくても理解することができる。つまり、単語家族の概念にある重要な本質は、一度基礎語やその派生語が認知されたら、その単語家族に含まれる語の認知にはほとんどの努力を要さなくてもよくなるということである (Bauer and Nation 1993: 253)。

タイプ、トークン、そして単語家族のような様々な単語の勘定単位が混在しているが、いずれにせよ、語彙研究で重要な点は「語彙のどのような側面を測定するのか」ということである。研究目的によって、その測定したい語彙知識や語彙能力を最も反映する語彙の勘定単位を選択する必要がある。

## 2.2. 語彙リスト開発の歴史<sup>26</sup>

第一章 1.2.2.でも言及したが、20世紀前半には既に言語教育の目的にコーパスを用いて基本語彙の選定が行われていた。このように、言語教育分野におけるコーパスに基づく語彙研究は、語彙リストの作成とともに始まった。本節では、語彙リスト作成の歴史を俯瞰する。

言語教育の目的で、コーパスに基づいて言語使用に関する記述についての研究を行ったパイオニア的存在は Edward Thorndike や Ernest Horn であろう。Thorndike は、聖書や古典英語データを集め頻度研究を行い、一万語までの最頻出語を集めた *The Teacher's Word Book*(1921)や、改訂を加えて二万語までを集めた *A Teacher's Word Book of 20,000 Words* (1932)を発表した。また、Horn は、手紙や出版物をまとめたコーパスを利用して一万語までの再頻出語を特定して *A Basic Writing Vocabulary* (1926)を発表した。これらのリストを共通の統計手法を用いて再構築したのが、Faucett and Maki (1932)による、*A Study of English-Word Values Statistically Determinated from the Latest Extensive Word Counts* である。

コーパスに基づく語彙リストの作成の一方で、Harold Palmer は経験で養った直感を用いた語彙リストの作成を行っていた。Palmer は語彙の頻度研究のより実用的な応用を目指して、1930年に経験を基にして作成した3000語の頻度リストを発表した。翌年には、その訂正版と1000語までの最頻出単語リストを完成させた。また、Albert Sydney Hornby と共に *IRET (Institute for Research in English Teaching) 600-word Vocabulary for Story-telling Purposes* (1932)を執筆し、後にそれを1000語まで拡大した *Thousand-Word English* (1937)を発表した。Michael West も Palmer と同様、経験に基づいた単語リストの作成を行った。West は、Indian Education Service の役人として英語教育に携わった経験からリーディングの重要性を訴えて、語彙を統制したリーディング教材を作成した。教材作成の過程から、重要語彙を直感的に選択し1800語のリストにしたものが、*New Methods Readers* (1927)である。

---

<sup>26</sup> 特記しない限り、この節は Howatt and Widdowson (2004 : 264-293)を参考にしている。

このように、基本語彙の選定の土台は 1920 年代から 1930 年代初頭にかけて築かれたと言える。1934 年には、West の指揮のにより Faucett, Palmer, Thorndike がニューヨークに集まり、Carnegie Corporation の協力のもと *The Use of English as a World Language* という会議が行われた。ここで注目すべきは、語彙リスト作成に向けて経験を基にする研究者とコーパスを基盤とする研究者が一同に会したことである。翌年にはロンドン大学で会議を行い、語彙選択の方法や基準を決定し、語彙リストの草案を完成させた。その判断基準の元となったのは、以下の七点である。

- 1) 語彙の頻度
- 2) 構造的価値(機能語を全て含む)
- 3) 普遍性(ある特定の地域で例外となる語は除く)
- 4) 主題のレンジ(専門語は除く)
- 5) 定義語(辞書編纂用の語)
- 6) 造語の可能性
- 7) スタイル(スラングなどは除く)

その成果は *Interim Report on Vocabulary Selection for English as a Foreign Language* (1936) であり、通称 *Carnegie Report* と呼ばれているものである。1939 年にこのレポートの改訂版のための話し合いが行われ、West が改訂を行うこととなったが、戦争が起こったため、終戦を待ってから改訂作業が行われた (West 1953: vi)。そして、1953 年に *A General Service List of English Words (GSL)* (West 1953) が完成した。

1970 年代からも同様に、外国語教育の目的に向けた語彙リスト作成について 2 つの手法に分けることができる。1 つ目は、統計処理を全く行わず、学習者にとって有益と考えられる意味や概念から出発する直感的なアプローチである。その例として、*Council of Europe* の *Threshold Level* プロジェクトが挙げられる。

2 つ目は、大量のデータと統計処理を用いる語彙リスト作成である。幅広い研究を網羅するデータベースを分析でき、提供できるよう、ほぼ無限の範囲のテキストからなるコーパスの構築が出発点であった。特に *Brown Corpus* の完成後、コーパスに基づく語彙リストの作成が次第に行われるようになった。

### 2.3. 語彙と第二言語の能力

本節では、語彙が第二言語能力のどのような役割を果たすのか概観する。前節で語彙リストの作成は教育分野で長い間行われてきたということを確認した。Michael Lewis は、第二言語学習における語彙の重要性について説き、Lexical approach の有効性を訴えた。このアプローチは、Lexical items (語彙項目)を中心に学習者の言語学習を進める教育方法をとるものである。語彙の重要性について Lewis (1993: 89)は以下のように述べている：

Language consists of grammaticalized lexis<sup>27</sup>, not lexicalized grammar. Lexis is the core heart of language but in language teaching has always been the Cinderella.

(言語は語彙化された文法から成るのではなく、文法化された語彙から構成されている。語彙は、言語の中心であるが、言語教育の分野ではシンデレラのような扱われ方をされている。)

外国語教育では長い間、言語の基礎となるのは文法であり、文法体系を習得することがコミュニケーションに必須であるという大きな誤解があった。かつては、統語的形態的な技術が長い間過大評価されてきた。これは長い間、言語学習の中心は文法構造を身につけることであり、語彙の負担は必要最小限にすべきだと言う Fries の見解に起因している (Lewis 1993: 115)。しかし実際のところ、母語話者らしい選択は統語規則だけの問題ではなく、語彙も大きな要因となっている。むしろ現在では、言語能力において単語の知識が最も重要な要因であると考えられている (Vermeer 2001: 217)。たとえば、語彙文法において、伝統的構造言語学者は語彙が文法の補助的役割をすると仮定してきたが、実際には文法が語彙の補助的役割を担っているということが、コーパス研究によって明らかになっている (Howatt and Widdowson 2004: 359)。言語使用パターンを決定する要因は語彙であるという発見が、コーパス分析で得られた最も顕著な発見であるといっても過言ではない (Howatt and Widdowson 2004: 358)。

近年の言語能力モデルでは、チョムスキー派のミニマリストなども、語彙を中心的役割として認識している (David 2008: 316)。また、実践を中心に扱う言語教育理論でもここ 30 年間、言語の文法に焦点を当てる伝統的な教授法よりも、コミュニケーションを通じた学習プロセスに焦点を当てている。つまり、語彙や文法に関する学習方法から、単語を使ってどのように物事を処理するのかについてへと、言語教育の焦点が移行してきた (Kennedy 1998: 280)。そして、文法構造は語彙と関連して学習されていることが明らかになっている (Conrad 2000: 553)。つまり、語彙の習得は言語学習の根幹である。

実際に、語彙が言語学習に必須の基礎であることは教師や学習者にとって既知のことであり、学術分野でも語彙の重要性について 1980 年代から再認識されるようになった (Adolphs and Schmitt 2003: 425)。これは、第一言語習得にも第二言語習得にも当てはまる (Vermeer 2001: 217)。また、評価の観点でも語彙に注目することによって大きな利点がある。言語能力自体を測定することは難しいが、第二言語における語彙学習の側面を調査する利点の一つは、少なくとも表面的には意味のある方法で数えたり測定したりできる点である (Milton 2009: 7)。

言語能力の下位構成要素と考えられている語彙能力やさらにその下位構成要素も、直接的な観察および測定をすることができない理論的な構成概念にすぎない。これらは、さらに下層の構成要素から推測されているものである。つまり、外国語による産出や受容パフォーマンスにおける、基本的な認知的構成概念の行動的具現として捉えられているものである (Bulté et al. 2008: 279)。そして、測定方法にしても言語能力を測定することは、距離や重さをはかるように簡単なものではなく、言語知識は直接質的観察を行うことはできず、学習者に測定可能な方法で知識を提示してもらおうほかない (Milton 2009: 6)。した

---

<sup>27</sup> 文法(統語)と対比した場合の語彙 (Richards et al. 1985)

がって、外国語による語彙能力のなかでも語彙の運用に関連する側面を、ここでは語彙能力ということにする。

## 2.4. 語彙能力とはなにか

語彙能力は一般的に、「その単語を使って何をするのか」が分からないレベルから、単語をどのコンテキストで使用するのかについて詳細な定義ができるというレベルまでの連続体のうち、どこに位置するのかを基準としている (Vermeer 2001: 219)。ではどのようにして詳細に語彙能力を測定することができるのだろうか。

### 2.4.1. 受容語彙と産出語彙

語彙能力を分類する主な慣行の一つとして、受容知識 (受動知識) と産出知識 (能動知識) がある (Milton 2009: 13)。受容知識は、読むことや聞くことと関連している。インプットを理解できるかどうか、語形を認識できるかどうか、語彙の意味を探し出すことができるかどうかといった点を指す。産出知識は、話すことや書くことと関連している。表現したい単語や表現を適切に使用できるかどうかを指す (Laufer and Goldstein 2004: 404, Laufer 2005a: 231)。また、産出知識には、統制された産出知識と自由産出知識があるという見方もある (Laufer 1998: 257)。統制された産出知識は、タスクによって誘発された産出語彙知識で、自由産出知識は言語産出者の意思による産出言語知識である (ibid: 257)。

一般的に、語の多くは先に受動的に学習され、産出知識はより上級知識が要求されると考えられている。ある単語を知っているということは、理想的にはそれを使って話したり書いたりすることができるということである (Laufer 2005a: 233)。また、ある場面で単語を使用することができる人は、多くの場合その単語に付随する意味も分かる (Laufer and Goldstein 2004: 408)。つまり、産出知識は発音、綴り、コロケーション、レジスターと言った面からこれらを正しく使用するのは受容知識を獲得することよりも難しい。さらに、学習者の多くはインプットとして単語に触れる量よりもアウトプットとしてそれらを使用する機会は少ない (Laufer 2005a: 231)。つまり、同一の学習者でも受容知識と産出知識の差がある可能性は高い。また、受容知識と産出知識の区別が二項対立であるか、または連続体であるのかについての絶対的なコンセンサスはまだない。しかし、多くの研究者によってこの区別自体は認知されており、教育面ではこの区別は役立っている (Laufer 2005a: 231)。

なお、受容知識も産出知識も「知識」という用語が使用されているが、これは Chomsky による言語能力に関する「知識」の定義とは異なっている。これらの語彙「知識」は、語彙の純粋な知識というよりは、語彙の運用能力を指しているといえる。

### 2.4.2. 語彙の深さと広さ

語彙能力を測定する際に、語彙の受容知識と産出知識の区別の他にも別の基準も考案されている。Anderson and Freebody (1983)は、語彙能力の下位分類として、知識の広さ

と深さを提唱した。語彙の広さは語彙サイズとも呼ばれ、学習者が少なくともある単語の基本的意味を知っている単語数のことを指す (Laufer 2005: 230)。語彙サイズの構成概念を知ることは、学習者の語彙能力を評価するのに役立つだけでなく、広い意味での言語能力を評価する指標にもなりえると考えられている (Milton 2006: 203, Read 2007: 120)。

語彙の深さは学習者がある単語についてどれくらい知っているかを意味する。たとえば、語彙知識が深いということは、ある単語の形態的、文法的特性や、様々な概念的、感情的意味、また語用的統語的關係やその使用に際する制限を詳細に知っているということである。単語の特性を知っていれば知っているほど、その人の語彙知識は深くなると考えられる。語彙知識の深さを掘り下げることは、より言語の発達が進むということである。この区別を行うことによって、たとえば語彙リストによる機械的な学習によって多くの単語を学習してきたが、単語をどのように使用するのが曖昧な学習者と、単語が持つ意味のニュアンスや他の単語との関連も知っている学習者との区別が可能となる (Milton 2009: 13, Laufer 2005a: 233)。

## 2.5. 伝統的な語彙の測定

たしかに、上記の語彙能力を測定するための基準は一見単純であり、学習者の能力について量的記述を加えることができる方法であるが、広さや深さという用語の意味自体に多様性がある曖昧な単語であり、混乱をきたす恐れもある。たとえば語彙の広さは、意味とかけ離れた語形の受動的な認知能力を含む場合がある。また、外国語学習においてある単語の意味や翻訳ができなくても、その単語を見たり聞いたりしたことを学習者が覚えている場合、その単語を知っていることになるとも言える。また、単純な語彙の受容知識や産出知識、あるいは語彙の広さや深さといった分類自体は、語彙能力の複雑性を公正に評価する妨げとなっているという指摘もある (Milton 2009: 13-14)。

### 2.5.1. 測定対象

伝統的に語彙能力の測定でよく用いられた対象は、語彙の広さ(語彙サイズ)である。語彙サイズは、読解能力に深く関連しているため、語彙テストは伝統的に読解能力の発達に関する研究やリテラシープログラムで重要な役割を果たしてきた (Read 2007: 107)。語彙サイズは、一般的にテストスコアを推定することによって測定される。たとえば、通常辞書などの単語リストから 1000 語ごとに 1 語というように無作為抽出し、各単語をテスト項目とする。そして、テストスコアに無作為抽出した単位の語数を掛けることによって、語彙サイズの指標を得る。たとえば、1000 語ごとに 1 語を無作為抽出し、正答数が 30 であった場合、語彙サイズは三万語であると考えられる (Vermeer 2001: 219)。語彙サイズテストにおける語彙の提示方法は一般的に最も一般的な語彙から順番に並べていく形式が好まれている。しかし、母語話者用の辞書は語彙頻度に関する明確な情報が含まれていないことから、語彙サイズテストに適応することには問題が残るという指摘もある (Read 2007: 107)。そこで、電子コーパスに基づいた単語の頻度リストを使用してテストを行うことで、ある一定のインプットを得た学習者、もしくはあるパフォーマンスレベルに達



した学習者グループがどのような語彙能力を持っているのかを、ある程度立証することが可能となった (Milton 2009: 251, Read 2007: 107)。学習者は一般的に低頻度語を習得する前に高頻度語を習得するという事は、驚くべきことではない。目標言語に触れる機会は、自然環境のインプットと教室活動によるインプットの両方から得られる。そしてどちらの場合においても、より語彙項目の頻度が高いほど、より学習者がそれらの項目を目にする機会は多くなるため、頻度は重要な要因である (Schmitt and Redwood 2011: 174-175)。また、一般的に低頻度語は中級レベル以下の学生にとってテキスト理解を妨げる要因となる (Nation 2001: 33)。つまり、学習者がどれだけ多くの低頻度語を学習しているのか、また学習者が学習した低頻度語がどのようなものかを分析することができれば、学習者の産出言語の評価にとっても有効である。

たとえば、英語の典型的な高頻度語と低頻度語の区別は、単語家族を単位とし 2000 語レベルあたりで線引きできる。つまり、2000 語のほどの高頻度語があると考えられている。この線引きは多くの研究によって支持されている (Nation 2001: 32-33)。そして、この高頻度語である 2000 語を有していれば、大抵の日常会話を行うことができるという見解がコンセンサスとなっている<sup>28</sup> (Adolphs and Schmitt 2003)。一方、フランス語では約 2000 語を習得することによって、テキストの 90 パーセントをカバーできていると言われている。なお、フランス語の上位 2000 語のリスト内に既に日常的な言語使用だけでなく、学術目的にもより効果的に役立つ単語も含まれていると考えられる (Cobb and Horst 2004)。

### 2.5.2. テスト形式

一般的な語彙テスト形式としては、これまでテスト項目に対応する意味を選択する手法が多くとられてきた。つまり、語彙の受容知識の測定である。たしかに受容語彙の測定は興味深いものであり、受容語彙と産出語彙は密接に関連していると考えられるが、受容語彙の測定だけでは学習者の産出能力の正しい根拠にはならない (David 2008: 317)。たとえば、学習者がある語彙について最も使用されている意味の知識があるからといって、必ずしも学習者がそれを的確に使用できるということとは直結しない (Bogaards and Laufer 2004: x)。そこで、目標言語から母語への翻訳や、また逆に母語から目標言語の翻訳を行うことで語彙産出知識の測定を促す方法もある。しかし、翻訳を使用する語彙能力の測定は、受容認識基準のみを使用するテストに比べて評価が低くなることが分かっている (Milton 2009: 14)。

語彙サイズテストの開発や応用が進んでいるのに対して、語彙能力の質に関する測定はあまり進んでいない。なぜなら、第二言語習得分析では主として、語彙の深さよりも語彙の広さを測定してきたからである (Vermeer 2001: 218)。伝統的に行われてきた語彙の測定は、語彙の側面のごく一部しか反映していないため、語彙能力全般についてはまだ分かっていないことが多い。語彙能力の質をテストする場合、語彙の書記形と語彙の意味と

---

<sup>28</sup> しかし CANCODE コーパスや BNC などの大規模コーパス分析の結果、2000 語以上が要求される可能性も指摘されている。

を一致させる手法が多くとられているが、それらの単語がどのように発音されどのように綴られるか、形態は何か、統語的にどのように機能するか、頻度はどれくらいなのか、社会言語学的に適切に使用することができるかなど、学習者の語彙目録に実際にそれらの単語が機能的な単位として取り込まれているのかも調査する必要がある(Read 2007: 113)。

このように、語彙の測定にむけてどのように辞書から語彙サンプルを抽出するのか、語彙を知っているのか知らないのかもどのようにテストすればよいのか、という深刻な手法的問題が残っている。これらの疑問に対して十分な答えが得られていないため、語彙サイズに関する研究結果には整合性がなく、時には誤った結果が発表されていることもあった(Nation and Warning 1997: 7)。また多くの場合、語の定義が曖昧な場合も多い。学習者が必要とする語彙能力の種類を反映するような勘定単位を選択する必要がある(Nation and Meara 2002: 35)。

## 2.6. 語彙の豊かさとその測定法

過剰使用と過少使用の他にも、語彙の測定として「語彙の豊かさ」という指標がしばしば用いられる。外国語能力を測定するのに、語彙の豊かさを利用する利点は多くある。まず、言語サンプル内で使用される全ての単語が分析対象となることである。また、自然な環境に近いコンテキストで発せられたテキストについての分析が可能となる(Treffers-Daller 2010: 5)。つまり、学習者コーパスを分析することで、直接学習者の産出語彙知識を分析することができると考えられる。よって、コミュニケーション重視の学習に対して正しい評価を行うことができると期待できる。

それでは、語彙の豊かさを測定するにはどのような測定の観点があるのだろうか。

Laufer and Nation (1995: 309-310)は、語彙の豊かさとして 1) 語彙の独創性；2) 語彙密度；3) 語彙の洗練性；4) 語彙の多様性の4点を挙げている。

### 1) 語彙の独創性；

ある調査対象が使用した語彙と、その他のグループメンバーが使用した語彙との比較。

### 2) 語彙密度；

基本的にテキスト内の名詞、動詞、形容詞、副詞などの内容語の割合のことであり、内容語数を作文内の総語数で割ることによって求められる。つまり、情報を主に伝達するのが内容語であることから、内容語と機能語を合わせた総語数に対して内容語を多く含んでいれば、テキストの密度は濃いと判断される。またJohansson(2008: 65)は、語彙密度の分析と情報パッケージの概念との関連を指摘している。つまり、内容語の割合が高いテキストは、機能語の割合が高いテキストよりもより情報量が多いということである。

### 3) 語彙の洗練性；

テキスト内の「上級語」の割合。一般的に上級語は低頻度語であると考えられている。

#### 4) 語彙の多様性

テキスト内のトークン数に対するタイプ数の割合。タイプとトークンの関係を計算する方法は、時とともに進化してきた。最もよく知られている語彙の多様性の測定方法は、**Type Token Ratio (TTR)**であり、タイプをトークンで割る方法である。また内容語に限定した **TTR** をもって、語彙の多様性を検証する場合もある。

語彙の多様性についてはその用語が普及しているが、同時に用語の定義が一貫していない場合があり、混乱を招いているという指摘もなされている。「語彙の多様性」は書き手や話し手の産出能力を指し示す場合もあるが、書き言葉や話し言葉自体の質を指す場合もある。言語産出者と言語産出者によって産出されたテキストには関係があるのは確かであるが、言語産出者の語彙の多様性は実際のところ動的なものである可能性がある一方で、言語の産出物における語彙の多様性は、言語産出者の静的な現れでしかないとの指摘もある(Yu 2010: 238)。

Bulté et al. (2008 : 279-280)は以上の点の他に、5) 語彙の複雑さ；6) 語彙の産出性；7) 語彙の流暢さについても、語彙測定の観点として挙げている。また、語彙の洗練性も観点として挙げているが、定義が Laufer and Nation (1995)とは異なっている。Bulté et al. (2008) による語彙の洗練性は、意味的に特殊な語彙に関する知識を持ち、関連語の中で語用的に最も正確な単語を使用できるかについての能力を問う観点である。このことより、ここでの語彙の洗練性は、下位語、上位語、類義語、反意語のような意味知識を含むこととなる。つまり、学習者が同じ指示対象を意味する別の語彙の代替案を持っているかどうかに関連している。ここでの語彙の洗練性はそれ自体で、学習者の語彙のマクロレベルを明らかにすることができ、語彙の広さと最も関連する測定法であると言える。

#### 5) 語彙の複雑さ；

ある単語について、典型的で標準な意味の単語やコロケーション、文法、語用を使用できるだけでなく、その単語のより特殊で周辺的な低頻度の属性を理解したり使用できたりする能力である。したがって、ここでの語彙の複雑さは、Laufer and Nation (1995)をはじめとする研究者による語彙の洗練性の定義と近いが、異なる点はこの能力が学習者の語彙知識の深さと関係しているということができる。

#### 6) 語彙の産出性；

語彙産出における行動観察レベルでの構成物である。たとえば、あるコンテキストを描写する際に、ある学習者は 100 語使用するのに対して、200 語使用する学習者もいる。つまり、あるタスクを遂行するために学習者が使用するトークンの数と関連している。

#### 7) 語彙の流暢さ；

関連する語彙情報が処理化される度合いによって決定される。つまり、学習者が語彙を産出する速さや、また特に内容語を把握する早さを指している。

また、Crossley et al. (2011: 564-565)はこの他に、語彙能力の核をなす「今、ここ」の概念と関連する単語の具体性や、ある単語が他の単語とどれくらい多く関連しているのかを表す語彙の有意味性、心的イメージを素早く簡単に引き出せるかどうかに関連する語彙の想像可能性などの観点も語彙測定の観点として指摘している。

## 2.7. 語彙の豊かさの問題点

語彙の豊かさは語彙研究で良く使用される用語であるが、その定義が一定ではない。語彙の多様性の研究のみをもって、語彙の豊かさの研究としているものも多い (Johansson 2008: 62)。その他にも語彙の豊かさの測定には、まず実用的な面で問題が残る。それは、語彙の豊かさを測定する際、言語データを測定可能な形にしなければならない。つまり、転写やフォーマット化が必要となる。このプロセスは大変時間がかかるため、臨床的応用に向けた道具としては向かないかもしれない (Treffers-Daller 2010: 5)。

さらにより困難なのは測定方法に関する問題である。とりわけ、語彙の多様性の測定で使用されてきた TTR は以前支持されていたほど精度が高くないことが分かっている (Malvern and Richards 1997)。それは、語彙の多様性の測定では、TTR は短いテキストでは不安定で、テキストの長さの違いによっても影響を受けるという欠点があるからである。なぜなら、話し手や書き手が持つタイプ数は限定されているため、テキストが長くなるとタイプ数の出現機会が減るからである (Daller et al. 2003: 199)。また、発達段階によってタイプはトークンと異なる発達段階を踏むという性質により、問題が発生することも考えられる。もし、タイプとトークンの数の発達の間、比例関係があるとすれば、TTR は常に一定の値を示すことになり、発達を測定することができない。もし曲線関係があるとすれば、トークンはタイプに比べて比較的早く増加するので、TTR の値は下がってしまう。たとえば、初級段階で“a”と“the”という機能語を習得した学習者の場合、これらの機能語は頻出するのでトークン数は増えるが、2つのタイプだけが増加することになる。すると、分母であるトークンが分子であるタイプを上回り、右下がりのカーブを描くこととなる。しかし、この学習者の言語習得は進んでいることには変わりはない (Vermeer 2004: 174-176)。Pierre Guiraud は、1954年に、これらの問題の一部の解決法を打ち出した。彼は、トークンを平方根にすることで、長さの異なるテキスト間に関する問題を解決できるとした。つまり、大きなサンプルでも同じ値を得ることができるようになり、信頼性も TTR より高くなる。また、Gustav Herdan は Log TTR を発表し、タイプの対数をトークンの対数で割ることで語彙の豊かさの概念を的確に反映できると考えた。Charles Muller はコーパス内に一度しか用いられない語である hapax の影響を中和できる公式を 1967年に発表した。また Daniel Dugast は同年、Uber index という公式 ( $Uber(U) = \log 2N / (\log N - \log V)$ )<sup>29</sup> を考えた (Dewaele 1993: 93-95, Daller et al. 2003: 200)。Malvern et al. (2004)は、D という測定法を発表した。D には3つの利点がある。テキストの長さは測定に影響を及ぼさないこと、全てのデータを一つのテキストとして扱うことができること、そして TTR よりもより有益であることである。D は

---

<sup>29</sup> N=トークン、V=タイプ

*CHILDES* から入手可能な言語プログラムである *CLAN (Computerised Language Analysis)* に組み込まれている<sup>30</sup> (David 2008: 321)。D は学習者の一般的な英語能力と有意に相関し、ライティングやスピーキングのタスクの質の決定に有効であり、特に学習者のスピーキングにおける言語運用をより正確に予測することができる (Yu 2010: 253)。実際に、英語学習者の書き言葉における語彙の豊かさについて統計的に研究した Crossley et al. (2011) の研究では、語彙の多様性の指標である D が、人間による語彙能力の判定における分散の 34% を説明することを明らかにした。これは、語彙能力についての人間の判定は、書き手が産出する語彙の多様性によって最も予測されるということを示唆している。従って、語彙の多様性は、熟達した語彙使用をもっともよく反映していると結論付けている。

この他にも、Corrected TTR (Carroll 1964)、Advanced TTR (Daller et al. 2003)、Advanced Guiraud (Daller et al. 2003) といった TTR の変形も提案されている。このように TTR の派生形が多く考案されてきたが、Treffers-Daller (2009) は Guiraud 値がタスク効果やレマ化のストラテジーに影響されない点で最も有効であると指摘している (ibid: 82)。また、van Hout and Vermeer (2007: 102) は、多くの研究で Guiraud 値、もしくは単なる異なり語数が最も語彙の豊かさを示す指標として機能していると指摘されていると報告している。さらに、Tidball and Treffers-Daller (2007) の研究では、D 値、Guiraud 値、Advanced Guiraud 値、また基本語の割合が一般的な言語能力テストの結果と有意で強い相関を示したことから、これらの測定法は言語能力における語彙測定に有効であると結論付けている。

しかし、TTR から派生したより洗練された公式でさえ、学習者のエッセイのような短いテキストには向かないと考えられている。その解決策の一つは、サンプルを同じ長さにするることである。たとえエッセイの長さについて語彙数を固定することによって、語彙の多様性における敏感性の問題を解決しても、二つの問題が残る。まず、語彙の多様性は語彙の定義に依存している。もし普段この測定法で行われているように、派生語が他の単語として考えられているのならば、単語家族が一つの単語と考えられている場合よりも語彙の多様性の値は上がる。もしそれぞれの語形が異なる単語として数えられているのならば、使用された単語家族数という観点から見れば、語彙の多様性が高いと言うことが必ずしも語彙の豊かさを反映しているとは限らない。単語家族自体の数は少ないが、その派生形を多く使用する学習者と、異なる単語家族を多く使用する学習者を区別することができない。また、語彙の多様性はどんな単語の種類を使用したかということも区別しない。たとえば、総語数 300 語の作文で、200 の異なり語数が使用されている場合、2000 語の語彙を学習した人と 5000 語を学習した人とでは、その 200 語に含まれる語彙の頻度レベルが異なることが考えられる。TTR の値が高いということは、学習者が知っている単語を使って表現することができるということを示すことができるが、どんな単語の種類を知っているのかを証明するものではない。したがって、語彙の多様性は単語の質を区別するものではない。そして、どのように単語が使用されているか、どこで使用されているかなどの点も考慮されていない (Laufer and Nation 1995: 309-310, Yu 2010: 246)。

---

<sup>30</sup> <http://childes.psy.cmu.edu> より入手可能

このように TTR を始めとする語彙の多様性の測定方法に対する最も大きな反論は、語彙間に確かに存在する区分別を無視していることである。つまり、語彙の難易度が考慮されていない。ある単語の難易度が高くてそうでなくても、この測定法はただデータ内に現れるタイプとトークンの実数に基づいているだけである。しかし、語彙を出現頻度層間で区別することができれば、質的側面を加えることができる。つまり、頻度に基づく語彙測定は、上級語や低頻度語、そして特殊語や抽象語を調査でき、学習者の産出語彙能力の質的観点から分析することができる。質的側面を持つ測定法は言語能力の語彙的観点についてより多くの知見を得ることができる (Vermeer 2004: 176, Daller et al. 2003: 203, Bulté et al. 2008: 292)。この点に関しては、次節で詳述する。

語彙の多様性の問題以外にも、測定方法の問題が指摘されている。たとえば、語彙の独創性の指標は、グループ内での学習者のパフォーマンスを測定するため、もしグループが変わればその指標も変わってくる。つまり、語彙の独創性の指標は産出言語によって定義されるのではなく、グループの要因によって定義されるので指標は不安定であり、信頼性を下げている要因となっている。また、語彙密度の指標は産出言語の統語特性や結束性に関係するので、必ずしも語彙のみに焦点を当てて測定しているのではないと言える。たとえば、ある作文で機能語が少なければ、従属節、分詞句、省略が多いと考えられるが、これらは全て書き言葉の構造特性であって語彙特性ではない。そして語彙の洗練性に関しては、「上級語」の定義が研究者によって異なるといった問題がある。そのため、この測定法はきちんと「上級語」の定義がなされない場合とても不安定である。研究者によって「上級語」の定義が異なることもあるため、研究の比較が困難となる場合がある (Laufer and Nation 1995: 309)。

語彙の豊かさは多面的な概念である。研究目的やデータの種類によって、研究者はそれぞれ語彙の異なる側面に焦点を当てている (Tidball and Treffers-Daller 2007: 134)。しかし、各構成要素はそれぞれ独立しているとは限らない。たとえば、語彙の多様性と語彙の洗練性は相関関係にあると考えられる。なぜなら、語彙の多様性が高まると、低頻度語の使用も増えるということが考えられるからである (Daller et al. 2003: 202-203)。語彙の豊かさの研究には、どの指標を用いるか、そして測定方法がどのようなものであるかを明確に定義し、幅広い語彙の豊かさの側面を測定できるように多様な観点を組み合わせて分析すべきである。

## 2.8. 頻度情報を使用した語彙の測定法

人間の多くの側面は、習慣や繰り返しによって特徴付けられる。言語使用の方法についてもこれの例外ではない。人間は、前に言ったことのあることを繰り返し使用する傾向にある (De Cock 2004: 225)。つまり、どの言語も非常に多くの単語を有するが、全ての単語が全て平等に有用であるとは限らない。ある語が有用であるかどうかを決定する方法の一つは語彙頻度である。これは通常の言語使用においてどれくらいの頻度で各単語が生起するかどうかを測定するものである (Nation and Warning 1997: 8)。実際に、語彙の豊かさに関する多くの測定法では、語彙の難易度の背景にある主な要因はその頻度にあると

いう前提に基づいている (Tidball and Treffers-Daller 2008: 299)。また、頻度情報は外国語教育分野でもとても有益な情報であるといえる (Biber and Reppen 2002: 205)。

語彙の重要性を決定する典型的な方法は、出現頻度と出現のレンジを考慮に入れることである。つまり、言語使用の広いレンジで多く現れる単語は、滅多に目にしない語彙で限られた分野のレンジでしか現れない語彙と比べてより有用である。また、高頻度語は、言語使用の幅広いレンジで目にする可能性が高いので、重要であると考えられる。高頻度語は多くの場合、その言語における基礎語であることが多い。基礎語彙は、日々の生活に関連するもので、学習者や身近な事柄と密接に関係し、その言語が話されている土地の文化や社会にまつわる語彙である (Andersen 2006: 49)。一方、低頻度語は上級学習者、もしくは母語話者でさえ学識のある話者しか所有しない。したがって、語彙の豊かさを低頻度語数や、専門語数、単義語数、職業語数で定義することも可能である (Ovtcharov et al. 2006: 110)。

しかし、ある言語における語彙を高頻度語や低頻度語との二区分にしてしまうのには、問題が残るだろう。一般的に、出現頻度が高い単語はほんの一握りであり、その他の多くの単語の出現頻度はとても低いからである (Stubbs 2002: 227)。また、高頻度であるからといって、それらの語彙を理解することや産出することが容易であるとは限らない。たとえば、Altenberg and Granger (2001: 174)は、高頻度の動詞は外国語学習者にとって問題となる可能性もあるということを指摘している。詳しい高頻度動詞の特徴として、Altenberg and Granger (2001: 174)は Viberg (1996)を参照し、比較言語学的観点から以下のような点を挙げている。

- 1) 基本表現であり、様々な意味分野を支配している
- 2) ほとんどの言語で対応する高頻度の動詞がある；多義語の度合いによって特徴付けられ、以下の2種類の意味の拡大によって引き起こされる。
  - a) より一般的、抽象的で、それ自身意味をあまり持たない文法化した使用となるような普遍的な傾向をもつ。
  - b) 特別な意味、コロケーション、イディオムの使用に関連する言語に特有の傾向を持つ。

このうち、特に 2b については、学習者が習得する際に問題となる場合が多いと考えられる。

語彙を単純に高頻度語と低頻度語に区分してしまうには問題が残るが、言語伝達能力や言語使用に関する興味教師の中で高まる中、母語話者による使用の出現回数、つまり頻度の傾向は、語彙の有用性の測定に重要であると考えられる (Kennedy 1998: 281-282)。また、学習者による単語の頻度は学習順序と関係があるので、語彙の測定の妥当性をより高めるためには、コーパス内の語彙頻度と関連づけるべきである (Vermeer 2004: 177)。

### 2.8.1. *Lexical Frequency Profile*

以上の頻度情報が持つ測定法の欠点を補うべく、Laufer and Nation (1995)は *Lexical Frequency Profile (LFP)* という新しい語彙の豊かさの測定法を提唱した。LFP は第二言語学習者の産出語彙を測定するための道具で、様々な語彙頻度レベルにおける語彙の割合を分析することができる (Laufer 2005b: 582)。

前述の通り、直感的に、頻繁に日常で目にする基本語ともいえる単語は、言語話者の大部分の人が理解できる単語であろう。しかし低頻度語や専門語など、教養のある母語話者や上級者にしかできないような単語もある。このような語彙の質の不均衡は、LFP の語彙頻度ゾーンの概念の根底にあり、後に説明する *VocabProfile* の前提となっている (Nation 2001: 32-33, Ovtcharov et al. 2006: 110)。LFP は対象言語の語彙頻度に基づいている。語彙頻度を利用することで、学習者の産出語彙を 3 つもしくは 4 つの頻度層に分けて割合を求めることができる。当初の研究では、頻度 1 位~1000 位までの頻度層、1001 位~2000 位までの頻度層、学術語彙リスト (University Word List: UWL) 内に含まれる単語、そしてリスト外の単語の 4 つのレベルを設定したプロファイルを使用していた (Read 2000: 204)。また LFP に基づく語彙測定は元来、書き言葉における語彙使用の様相を計量化することを目的としていた (Laufer and Nation 1995: 307)。

英語の LFP には、初級学習者向けの測定法と上級学習者向けの測定法の 2 つの測定法が提案されている。初級学習者は上級学習者よりも低頻度語を使用する割合が低いことから、初級学習者の場合、学習者の分類を頻度 1~1000 位レベル、1001 位~2000 位レベルとその他の単語との間で行うことができる。上級学習者は、基本語以上でないと詳細な区分ができないので、1001~2000 位レベル、UWL とリスト外単語の間で区別することができる。この学術語彙リストには、英語の頻度リスト上位 2000 位までに含まれない 836 の単語家族からなり、幅広い分野の学術書に頻出の単語を含んでいる (Laufer and Nation 1995: 311-312)。

LFP と語彙の洗練性は双方とも語彙の頻度を基にしており、学習者が使用する単語がどれだけ上級の単語かどうかについて研究を行うために使用される。そのため着眼点は似ており、双方間で差異があまりないと考えられることもある。しかし、大きな違いが 2 つある。一つ目は、語彙の洗練性はただ単に高頻度語であるか低頻度語であるかの 2 種類の区別しかないが、LFP はさらに細かい頻度レベルを設けることによって、語の種類についてより詳細な全体像を提供することができる。また、語彙の洗練性について行われた研究は、洗練された語彙についての定義が一定ではないため、学習者の環境が異なるとグループ間の比較を行うことはできない。その点、LFP は語彙シラバスに基づくのではなく、決められたコーパスの頻度によってあらかじめ頻度層の定義付けが行われている。そのため、信頼性の高い頻度統計によって順位付けされた客観的測定法であり、学習システムが異なる学習者間を客観的に比較することができる (Laufer and Nation 1995: 312-313)。

LFP の作成は、単語のトークン、タイプ、単語家族を基に行われる。語彙の豊かさの指標として最も有効だと考えられるのは、単語家族を基にした計算である (Laufer and Nation 1995: 312-313)。しかし、同じ単語家族に属する語でも、母語話者や学習者から



認知されやすいものもあれば、そうでないものも存在する。たとえば、Bauer and Nation (1993: 257)は、学習者は単語家族内にある単語間の関係を把握するためには様々な知識が要求され、様々なレベルに注意を払う必要があると続けており、以下の点を学習者に求められている事項として挙げている。

- 1) 学習者は基礎語の知識が求められている。
- 2) 学習者は既知の基礎語を認識できなければならない。つまり、2つの単語が共通の基礎語を持つとき、学習者はその2つの単語が関連していることを認識できなければならない。また、基礎語に同じ綴り字が含まれていても、2つの単語が共通の基礎語を持たないとき、2つの単語に関連性が無いことを見分けられる能力も必要とされている。
- 3) 学習者は、少なくとも接辞が単語を形成するという事について知っている必要がある。多くの接頭辞やいくつかの接尾辞においては、接辞の語彙的な意味に関する知識も含まれる。接尾辞については、学習者は接尾辞によってもたらされる統語的情報も知っていなければならない。たとえば英語の場合、*-less* は派生語を形容詞に変えるという知識である。
- 4) 話し言葉や書き言葉で接辞を生産的に使用するために、学習者は基礎語と接辞を正しく組み合わせることができなければならない。

語彙の豊かさの測定方法は他にもあるが、なかでも *LFP* は他の語彙測定法と比べて妥当性のあるデータを提供することが可能である。まず、語彙の多様性とは異なり、*LFP* はテキストの長さには依存せず、調査対象として語彙の主観的定義も排除できる。また、語彙の独創性とは異なり、学習者の学習環境にも依存せず、学習者グループが変わっても結果は一貫性を持つことができる。さらに語彙密度とは異なり、*LFP* は語彙のみに焦点を当てるため、産出言語の統語規則やテキストの結束性、文法の側面から独立することができる。*LFP* には他にも、トピックやサブトピックに関する主観性を排除できるなどの利点もある (Laufer and Nation 1995: 312-313, Ovtcharov et al. 2006: 113)。

#### 2.8.2. *VocabProfil*

*VocabProfil* は、Laufer and Nation によって作成された英語版である *VocabProfile* をフランス語版に応用したものである。ベルギーの語彙論研究者である Verlinde and Selva (2001) が作成した新聞コーパスから語彙リストを作成した。このコーパスはフランスの新聞である *Le Monde* とベルギーの新聞である *Le Soir* から 1998 年に発行された記事を抽出したものである。その後、Jones (2001) がコンピュータ処理によるアルゴリズムを使ってレマ化を行った (Ovtcharov et al. 2006: 113)。

レマ化された単語は頻度順に並び替えられ、頻度 1 位から 1000 位までを K1、頻度 1001 位から 2000 位までを K2、頻度 2001 位から 3000 位までを K3、そして頻度上位 3000 位までに現れなかった単語を Off-list と、4つの頻度層に分類した。

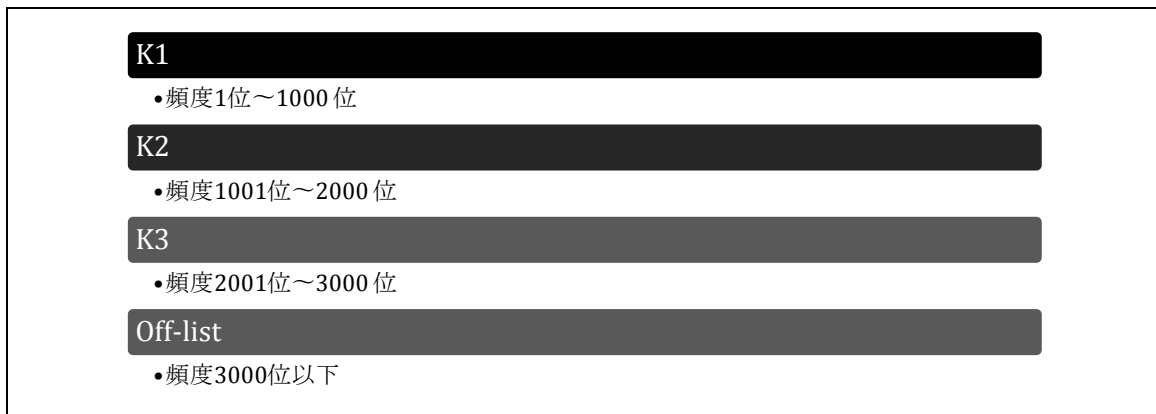


図4 VocabProfilの頻度層

*VocabProfil* は指定するテキスト内のトークンの数、タイプの数、単語家族の数を、あらかじめ設定された頻度リストと比較して自動計算し、語彙の豊かさを測定する(Thomas 2008: 590)。また *VocabProfil* を使用して、分析対象のテキストと語彙リストを比較することができ、テキスト内のどの単語がリストに入っているかを見ることや、リストでカバーされている項目がテキスト内にどれくらいの割合を占めているかを知ることができる。また、2つのテキスト内の語彙を比較し、どれくらい同じ語彙を使用しているのかや、テキスト内にどのような語彙の違いが存在するのかなども分析することができる(Nation 2001: 34)。Goodfellow et al. (2002)は *VocabProfil* を使用して初級学習者の書き言葉の語彙の豊かさについて分析し、*VocabProfil* の妥当性を検証した。

なお、Cobb and Horst(2004)によると、フランス語の上位2000語のカバー力はとても高いため、学術文献の理解を促進する *University Word List* と同様のものをフランス語で作成する必要はないと考えられていることから、英語版 *VocabProfile* に見られる *UWL* はフランス語版 *VocabProfil* には統合されていない。

### 2.8.3. LFPへの反論

*LFP* も他の語彙テストと同様に、一種の間接テストである。*LFP* は、テキスト内に現れる高頻度語や低頻度語の割合を出すことができるが、この割合はテキスト産出者が持つ語彙の高頻度語や低頻度語の割合を反映しているとは言い切れない。なぜなら、プロフィール内には正しく産出された語彙のみが反映されるため、*LFP* は学習者が言語形式と意味が正しくリンクしている高頻度語と低頻度語の割合しか示すことができないからである(Laufer and Goldstein 2004: 403)。つまり、テキストに現れないが、話者が持っていると考えられる語彙能力については、分析することができない。

また、*LFP* に基づく測定は、全て頻度情報をもとに行われる。しかし、頻度は語彙の洗練性を決定付ける唯一の要因ではないため、語彙の洗練性という観点をを用いる測定法にとってこれは大きな問題となる(Treffers-Daller et al. 2008: 274)。書き言葉の測定に限定したとしても、語彙の豊かさを決定する要因は、頻度以外にも多数存在する。たとえば、ライティングのトピックへ精通しているのか、ライティング技能はあるのかという

点である。つまり、トピックが変われば語彙の豊かさの結果も変わる可能性がある。もし、語彙の豊かさが頻度以外の要因によって強く影響を受け、それらの要因を制御することができなければ、語彙の豊かさの信頼性の高い測定を行うことはできない。同じ学習者が異なるテーマで書いた作文について、一定して信頼性の高い語彙の豊かさの測定結果を返すことができるかを示すことが必要である (Laufer and Nation 1995: 308)。また、単語の学習しやすさも考慮すべき観点であろう。母語に同起源の単語がある場合、外国語でこの単語を習得することは難しいことではない。一方、目標言語では高頻度語であるが、母語には対応する意味を持たない場合、また母語で別の語彙が使用されている場合は、理解することや習得することが困難である。このことより、言語教育において単語の頻度リストのみでは語彙選択や語彙の等級付けの唯一の拠所とはならない (Bogaards and Laufer 2004: x)。また、ある抽象的な意味においては有用な語彙ではないが、特定の人や特定の目的には有用な語彙も存在する。これは 1950 年代にフランスで *le Français Fondamental* を確立するための研究が行われた際に発展した概念である。頻度に基づく単語を補強するために、研究者は « disponibilité » という概念を提案した。これは、言語全体ではそれほど高い頻度ではないが、ある特定のコンテキストで使用される単語を考慮したものである (Howatt and Widdowson 2004: 290-291)。これらの頻度情報以外の観点については、*LFP* 単独で分析することはできない。したがって、*LFP* の欠点を補完できるような分析が個々に求められている。

その他にも *LFP* には、テキスト処理の方法についていくつかの技術的問題点や不明確な点がある。たとえば、エラーの扱い方、固有名詞の数え方、句の扱いなどが挙げられる。つまり、エラーを無視するのか訂正するのか、リストに含まれない固有名詞を低頻度語としてカウントするのか、複合語や定型句などを一語として扱うのかという問題も残っている (Meara 2005: 34)。また、*LFP* は多義語を区別しない点も指摘されている (Ovtcharov et al. 2006: 121)。しかしこの点については、実際に解決が困難であることが予測される。つまり、多義語の定義自体が不鮮明であるため、ある単語に複数の意味があるのかどうかについて、どのような判断を行えば良いのかが明らかでは無い場合があるからである。

フランス語版である *VocabProfil* に限定した問題点は、Jones (2001) のフランス語の *VocabProfil* に適応したレマ化のアルゴリズムが意図的に網羅的である点である (Cobb and Horst 2004: 27)。*VocabProfil* は、動詞の全活用形をレマ化している。フランス語教師やフランス語母語話者にとって、動詞の活用形は全て同じ価値を持たないということは明らかである。好まれて使用される活用形、あまり使われない活用形、話し言葉のあるレジスターではもう使用されることが少ない活用形もある。また、低頻度の接辞のいくつかが高頻度基礎語に付加されており、学習者が容易に認識できないような形式も高頻度語として分類されていることもある (Ovtcharov et al. 2006: 121, Cobb and Horst 2004: 27)。つまり、フランス語の *VocabProfil* では英語の *VocabProfile* ほど綿密な単語家族の定義や設定が行われていない。

最後に、話し言葉の研究において *LFP* を用いる際の最大の問題点は、*LFP* が書き言葉コーパスに基づいて作成されている点である。話し言葉コーパスに基づく *LFP* は今のところフランス語に関しては存在しない。書き言葉テキストに基づくコーパスが書き言葉に特有の単語を多く含み、話し言葉サンプルから成るコーパスには話し言葉に典型的な語を

含んでいるということは驚くべきことではない (Ovtcharoc et al. 2006: 119-120)。Treffers-Daller et al. (2008: 274)も指摘しているが、書き言葉コーパスよりも話し言葉コーパスの規模がかなり小さいことから、話し言葉における語彙頻度項目に関する妥当的な情報が得られにくいという現状であり、書き言葉コーパスの頻度データに基づく話し言葉の研究が行われているのはやむを得ない。しかし、書き言葉と話し言葉の違いを反映できていない語彙の豊かさの測定は、多くの問題に直面する。たとえば *VocabProfil* に基づくデータ分析を行う際、リスト外単語には”Bousculer” (突き飛ばす) というような非常に稀な単語が含まれるのと同時に、書き言葉では滅多に現れないが話し言葉に特徴的な単語も含まれ、この性質の異なる両者が混在する (Tidball and Treffers-Daller 2008: 299)。LFPの話し言葉分析の妥当性について肯定的な意見も見られる(Laufer 1998、Ovtcharov et al. :2006、Thomas :2008)。しかし、話し言葉コーパスに基づく LFPの開発が望まれていることも事実である。

## 2.9. 過剰使用と過少使用

前述のとおり、語彙能力の測定には語彙サイズや語彙の豊かさが多く指標として用いられてきたが、学習者コーパスの登場で過剰使用や過少使用といった面にも注目が集められるようになった。過剰使用語や過少使用語はまとめて「特徴語」と呼ばれる。この特徴語という概念は、複数のコーパスが言語的・内容的に均質であるならば、同じ語の頻度はおよそ同じであり、もしある語が一方のみで過剰に頻出しているならば、その後はそのコーパスの言語的・内容的性質を集約したものであるという考えに基づいている(石川 2012: 95)。そして、特徴語を検索する意義は、コーパスが代表する言語変種の特徴を、具体的で個別的な語のレベルに還元して、実証的手法で分析・検証するようにする点であると石川(2012)は続けている(ibid: 95)。

特徴語は一般的に、学習者コーパスと母語話者コーパスの頻度差が顕著であるかどうかを統計的な手法で検定することによって抽出される。この際、特徴性の度合いが基準となる。この特徴性の度合いの基準となるのは、主にカイ二乗統計量と対数尤度比である(石川 2008: 98)。

カイ二乗統計量は、予想される期待値と観測された実測値のずれの幅を検定の為に加工した値であり、対数尤度比もまた、実測値と期待値の差に注目したものである(石川 2008: 98-99)。この2つの公式は、以下のとおりである(石川 2008: 100)：

カイ二乗統計量：	$\chi^2 = \sum (\text{実測値} - \text{期待値})^2 / \text{期待値}$
対数尤度比：	$G^2 = 2 \times \sum \text{実測値} (\log e (\text{実測値}) - \log e (\text{期待値}))$

対数尤度比では定数 e を底とし、その値はおおよそ 2.7 である。

このように、統計的手法を用いることによって、主観ではなく学習者の特徴的な使用を抽出できるが、日本人フランス語学習者の産出言語に関する特徴語の研究は、これまで行われていない。

## 2.10. 学習者の語彙研究

前節まで、中間言語における語彙研究の流れについて概観した。語彙研究を行う上で、学習者の産出言語への評価の観点に興味の的になってきた。そして、母語話者コーパスに基づく頻度情報を用いた評価に一定の有効性が見られることが分かった。しかし、また、産出言語の評価は書き言葉の分析から発展した分野であるため、学習者の話し言葉に関する分析があまり行われていない。

日本人英語学習者の話し言葉における使用語彙について行われてきた代表的な研究の一つとして、日本人英語学習者のスピーチプレゼンテーションを英語母語話者のパフォーマンスと比較した Ishikawa (2007)が挙げられる。これは、TOEIC® 600-650 点に相当する 6 名の中上級の大学生英語学習者が、それぞれトピックを決めて英語スピーチを行った際の語彙使用の特徴について分析を行ったものである。また、英語母語話者である 3 名の教師が論理パターンをかえずに同じトピックについてスピーチを行ってパラレルコーパスを構築し、比較を行った。主に以下のリサーチクエスションについて分析が行われた。リサーチクエスションとその分析結果は以下の通りであった。

1. 日本人英語学習者はどれくらい多くの語を使用するか。また、どれくらい語彙が豊かでテキストが複雑か。

学習者の使用語彙数はトークン数、タイプ数とも話者よりも有意に低かった ( $p < .001$ )。また、TTR の値、語彙平均長さ、文の数、文の長さを分析したところ、学習者の方が母語話者よりも簡潔で短い分を使用しているが、TTR の値や語彙の平均長さの値は学習者の方が高かった。したがって、学習者の方が難しい語彙を使用している傾向にある事が分かった。

2. 高頻度後をより多く使用しているか。

それぞれのコーパスに含まれる上位 5 位から 30 位までの語彙が占める頻度を調べたところ、日本人英語学習者は必ずしも高頻度後を多く使用している訳ではなかった。

3. 日本人英語学習者によって過剰使用、または過少使用された高頻度後にはどのような特徴があるか。

学習者のプレゼンテーションには「I」が多く含まれており、「I-oriented」のスピーチであった。また、学習者はトピックに直結する語を多用していた。そして、順序を明確に示す語 (*first, second, third*)などを多く用いるため、「affirmative-centred」であると言える。しかし、*so, and, but, as*などの接続詞は過少使用していた。また、*the, they, it*などの前方照応に見られる語彙装置も使用が低かった。さらに、母語話者に比べて *be, go, do, have, think*などの基礎動詞の使用が低いことから、母語話者の方が動詞中心的な分を発していることが示唆できた。

4. 日本人学習者によって過剰使用または過少使用された品詞は何か。

学習者のスピーチは名詞中心であり、名詞を修飾する *of*などの前置詞が多用されているため、語彙レベルで言語単位に注意を払っていることが分かった。

一方、母語話者は接続詞、代名詞、定冠詞など、断片をつなぐための語を多用しており、句や文レベルで言語単位を構成していると言える。

5. 観察された使用語彙の違いは、母語による違いか、それともトピックの違いか。

コレスポネンス分析を行い、サンプルとカテゴリー間の関係を表にしたところ、語彙の分布を決定する要因はトピックの差よりも母語の差であることが分かった。

このように、コーパスを用いた分析を行うことで、学習者の使用語彙の傾向が明らかになる。しかしこの分野の研究は、英語学習者に関する研究に偏重している。

本節では、今まで行われてきたフランス語学習者の話し言葉における語彙研究を概観していく。

## 2.1 0.1. フランス語学習者の語彙研究

### 2.1 0.1.1. Debrock et al. (1999)

この研究はLANCOM (LANgue et COMmunicatio) コーパスを使用して、フランス語学習者による口頭インタラクションの実現における問題を分析したケーススタディである。データは先述のとおり、フラマン語母語話者であるフランス語学習者や母語話者がロールプレイングを行い、それを録音転写したものである。フラマン語母語話者のフランス語における発話には、以下の4点の特徴が見られた。

#### 1) 制度的特徴に起因する人工的なインタラクション

学習者が行うロールプレイングには、インタラクションにおける発話内行為の観点でフランス語話者とは異なる社会文化的振る舞いが見られた。たとえば、依頼人と請負人の関係において、フラマン語母語話者である学習者は依頼人先導で物事を決定するのに対して、フランス語母語話者は請負人が決定権を持つことが観察された。

#### 2) 会話の始め方、終わり方の慣例の違い

電話会話において、フラマン語母語話者は始めに身分を名乗るのに対して、フランス語母語話者はそれを行わず挨拶から始めた。会話の終わりにはフラマン語母語話者は別れの挨拶をするにとどめるのに対して、フランス語母語話者はここで名前を名乗り、連絡先を告げた。このように、フラマン語母語話者とフランス語母語話者の間に電話会話時における慣例の違いが見られる。つまり学習者はフラマン語における慣例をフランス語会話においても適応してしまう例が見られた。

#### 3) 談話・会話の管理

学習者は書き言葉調の話し言葉を使用するのに対して、フランス語母語話者は対人的コミュニケーションを対話者と構築している。また、学習者は厳密に事実即してコミュニケーションを行うのに対して、母語話者は状況に合わせて情報や提案を提示している。

#### 4) 対人関係の管理

学習者のポライトネスの使用に母語話者との違いが見られた。たとえば、「～しなければならない」という文を表現する際に、学習者は **DEVOIR** を使用するのに対して、母語話者は **FALLOIR** を好むことがわかった。また、疑問形における形態統語的な特徴の違いも見られた。学習者は **EST-CE QUE** 型の疑問形を用いるが、母語話者はイントネーション型の疑問形をしている。**EST-CE QUE** 型の疑問文のイントネーションは下降調であり、イントネーション型の疑問文は上昇調である。実際に、イントネーション型の上昇調イントネーションの方が対話者に対して威嚇的な印象を与えないため、学習者の発話はポライトネスが欠けているという印象を与える恐れがある。

#### 2.1 0.1.2. David (2008)

この研究は *FLLOC (French Learner Language oral corpora)* を使用して、半自然会話タスク遂行時における語彙の豊かさを記述し、学年別の比較を行ったものである。この研究の目的は、先行研究ではあまり行われてこなかった分野である口頭産出におけるフランス語の語彙に焦点を当て、中等教育における語彙の発達に関する記述を行うことである。被験者は、フランス語学習者であるイギリスの 9 年生(13 歳)から 13 年生(18 歳)である。

研究設問は以下の 2 点である。

- 1) 5 年の学習期間にわたる横断的、長期的データを通して、語彙の多様性の指標が有意に上昇したかどうか。
- 2) 産出語彙の発達の特徴は何か、また、名詞と動詞の発達ペースが同じか。

分析の結果、全体的な産出語彙量はトークン数、タイプ数共に学年が上がるにつれて有意に増加していることが分かった (トークン数:  $F(4, 95)=8.475, p<.001$ ; タイプ数:  $F(4, 95)=17.216, p<.001$ )。しかし、チューキー法を使って多重比較を多なった結果、1 学年ごとの差で有意差が確認できたのは、11 年生と 12 年生との間におけるタイプ数だけであった。

各グループの D 値の中央値を比較したところ、語彙の多様性も初級から上級まで一定に上昇傾向であることが分かった。また、1 学年ごとの差について t 検定を行った所、有意差が見られたのは 9 年生と 10 年生の間 ( $p<.05$ )、10 年生と 11 年生の間 ( $p<.01$ )、11 年生と 12 年生の間 ( $p<.05$ ) であり、12 年生と 13 年生との間には有意差が見られなかった。

また、使用語の種類に関する分析については、9 年生から 10 年生までにかけて名詞が占める割合が増えるが、それから先は徐々に減少していることが分かった。これは、増えた語彙の種類が名詞のみに偏重しなくなるからである。一方、動詞の割合はこれと反対の様相を呈する。つまり、学習者は 10 年生から動詞の割合が増えていく傾向がある。また、名詞と動詞の割合の間に負の相関があることから、この傾向を裏付けることができた ( $r=-.315, p=.001$ )。

### 2.10.1.3. Treffers-Daller (2010)

この研究は、自然タスクではなく統制タスクを使用した話し言葉コーパスに基づくものである。このコーパスは学習者の語彙使用分析を目的に作成されたもので、学習者グループは2つあり、3つの言語で構成されている。一つ目の学習者グループはブリュッセルに住むオランダ語の地域変種であるフラマン語が支配的な話者25名から構成されている。全員がフラマン語を話し、何人かはフランス語とオランダ語の標準語も話すことができる。もう一つのグループは、パリのビジネススクールに通うフランス語母語話者で、学校の教育は全て英語で受けている。彼らは全てフランス語の環境で育ち、英語やその他の言語は中等教育までに学習したのみである。学校では日常的に英語を使用しているが、このグループは明らかにフランス語が支配的と言える。このデータ収集の参加者は全てフランス語を話すことができる。フランス語のタスクとして、Mayer (1969)の《*Frog where are you ?*》という絵本を使用した。英語とオランダ語のタスクは、フランス語のタスクに翻訳による影響を与えないため、Mayer (1974)の別の絵本《*Frog goes to dinner*》を使用して、ストーリーテリングを行った。なお、このコーパスは *Brussels Corpus* として *FLLOC*の一部として収録されている。

全被験者は研究者との個人インタビューを行う前に準備期間が与えられた。全てのデータは CHAT フォーマット(Codes for the Human Analysis of Transcript)で転写され、CLAN 上で形態統語タグも付与された。タグの曖昧な点や、エラー、不整合は手動で修正した。加えて、全ての固有名詞、ポーズ、ためらい、驚き、メタ言語的なコメントについて、転写には含めたものの語彙の豊かさの測定時には排除した。

3つの言語の D 値の中央値を境界線とし、インフォーマントを低グループと高グループに分類した。この方法によって、インフォーマントは4つのサブグループへ分類することができた。

- 1) フランス語優勢のグループ(フランス語能力が高く、オランダ語もしくは英語能力が低い)
- 2) ハイレベルでバランスがとれているグループ(フランス語能力が高く、オランダ語もしくは英語能力も高い)
- 3) 低レベルでバランスがとれているグループ(フランス語能力が低く、オランダ語もしくは英語能力も低い)
- 4) オランダ語もしくは英語優勢のグループ (フランス語能力は低い、オランダ語もしくは英語能力は高い)。

この区分では、パリのインフォーマントの大多数がフランス語優勢グループに属し、10名のブリュッセルのインフォーマントがオランダ語あるいは英語優勢グループに属した。ハイレベルでバランスのとれたバイリンガルのグループは、ブリュッセル出身者で15名、パリ出身者で8名であった。ブリュッセルのインフォーマント内にはフランス語優勢グループに属するものはおらず、パリのインフォーマント内にも英語優勢グループのものはいなかった。



この分析結果から、語彙の豊かさを指標して言語の優勢を決定することができるということが明らかになった。また今回使用した方法は、異なる言語背景を持つ話者グループを区分できる有効な手段であるということも分かった。この方法を使う最大の利点は、規格化されたテストがない言語や言語変種にこの方法を使うことができるということである。この測定法は、口頭産出データなどのデータにも応用できる。

#### 2.1 0.1.4. Bulté et al. (2008)

この研究は、第二言語における語彙能力の定義と、語彙能力の発達に焦点を当てたものである。ブリュッセル在住のフラマン語話者のフランス語学習者の語彙の多様性、語彙の生産性、語彙の洗練性の発達を分析した。学習者 19 名と母語話者 19 名が Mayer (1969) の « *Frog where are you ?* » のストーリーテリングを行った。データは二年にわたって計三回収集され、学習者の年齢はデータ収集開始時で 12 歳、終了時で 14 歳である。学習者はデータ収集時に週 2 回から 4 回の 50 分のフランス語の授業を最低 4 年受けている。母語話者データは計 1 回の収集で、13 歳時に行われた。

語彙能力を測定するために合計 42 の指標が使用されたが、分析に使用されたのはそのうちの 22 の指標である。主に指標は、頻度情報、Uber Index、ギロー値が採用されており、3 つのカテゴリーに分類されている。

- 1) 標準測定  
全体のタイプとトークンの分類
- 2) 品詞測定  
名詞、動詞、形容詞に焦点を当てた測定
- 3) 頻度に基づく測定  
相対的な生起頻度に関する情報や基礎語と上級語の分類

分析の結果、それぞれのカテゴリーについて以下のことが分かった。

- 1) 標準測定  
三年間を通してすべての指標が有意に上昇しているため、学習時間と標準測定の指標は有意に相関することが分かった。しかし、この上昇は主に 1 年時に見られるものであった。語彙の多様性の指標は 1 年時に有意な上昇を見せたが、2 年時にはその傾向が見られなかった。また、母語話者は学習者よりも常に優れた値を返していた。
- 2) 品詞測定  
分散分析の結果、学習者は形容詞に関する指標を除いて統計的に上達が見られた。しかし、これは 1 年目の名詞と動詞の語彙の多様性が上昇したからであり、2 年目には有意な発達は見られなかった。また、母語話者は常に学習者よりも数値が上回っていた。

### 3) 頻度に基づく測定

学習者は時間が経過するにつれ、母語話者に近い数値に近づくことが分かった。特に、高頻度語に関する指標について顕著な発達が見られた。また、母語話者は学習者よりも常に指標が上回っていた。

例外があったものの、反復測定分散分析によってフランス語学習者はほぼ全ての測定法において、有意に言語能力が発達していることが明らかになった。また、学習者の発達は二年目よりも一年目の方が大きいことも、ボンフェローニ多重比較によって証明された。一年目は測定を行った4つの全ての語彙能力の観点から有意な発達が見られたが、2年目には低頻度語数と、テキスト長に比例する上級異なり語の多様性といった2つの語彙能力の観点のみに発達が確認できた。最後に、形容詞の品詞測定を除いて、フランス語母語話者は学習者よりも一貫して有意に高いスコアを得ていることが、t検定によって明らかになった。

#### 2.1 0.2. *VocabProfil*に基づく研究

*VocabProfil*を使用したフランス語学習者の研究は、まだまだ数少ない。本論文では、書き言葉の研究であるが *VocabProfil*に基づく先駆的な研究である Goodfellow et al. (2002)と、話し言葉の研究に *VocabProfil*を用いた Ovtcharov et al. (2006)、Thomas (2008)について俯瞰する。

##### 2.1 0.2.1. Goodfellow et al. (2002)

Goodfellow et al. (2002) は、*VocabProfil*に基づいて初級学習者書き言葉コーパスに基づく検証を行った。被験者は通信教育を行っている Open University のフランス語学習者 36 名であり、レベルは初中級である。被験者は議論型のエッセイを執筆し、それを分析した。分析には頻度 1 位から 1000 位までの K1、1001 位から 2000 位までの K2、*ELRA Parole French Corpus*<sup>31</sup>から抽出したアカデミックリスト、リスト外単語の4つの頻度層を用いた。これらの各頻度層におけるタイプ数と、教師による評価を比較した。なお、教師による評価には内容に関する項目が2つ、正確さに関する項目が1つ、そして語彙のレンジに関する項目が1つ用意されており、各項目は25点満点から得点付けされた。なお、語彙の評価のため、内容に関する評価は分析には含まれなかった。

評価と頻度層におけるタイプ数の相関を分析した所、正確さとレンジにおいて K1 には負の相関が見られ(正確さ:  $r=-.35$ ,  $p=.05$ ; レンジ:  $r=-.35$ ,  $p=.05$ )、アカデミックリストとリスト外単語においては有意な相関が得られなかったが、K2 レベルにおいてレンジと正確さに中程度の正の相関が見られた(正確さ:  $r=.42$ ,  $p=.01$ ; レンジ:  $r=.45$ ,  $p=.01$ )。

---

<sup>31</sup> European Language Resources Association のホームページ(<http://www.elda.fr>)から入手可能

#### 2.1 0.2.2. Ovtcharov et al. (2006)

英語学習者について、語彙の豊かさと言語能力の相関関係はある程度認められているが、英語以外の言語ではその研究があまり行われてこなかった点を Ovtcharov et al. は指摘し、フランス語学習者についての研究を行った。この研究は、外国語としてのフランス語の会話研究と、会話における語彙の豊かさを測定するために *LFP* を使用したパイオニア的研究である。まず、言語能力には語彙の深い知識が必要であるという仮定を立て、その仮定を検証した。そして学習者がどのように語彙知識の獲得を促進できるかを模索した。学習者の語彙能力が高ければ高いほど、特殊語や低頻度語を使用するという仮定を基に、さらに3つに細分化した仮定を立てた。

- 1) 二つの異なるレベルの学習者間では、中頻度語と低頻度語において、その使用割合に統計的有意差がある
- 2) レベルの高い学習者は、レベルの低い学習者と比較してより多くの特殊語や低頻度語を使用する
- 3) レベルの高い学習者の語彙のプロファイルは、母語話者のプロファイルと近似している。

被験者は、カナダで行われる試験である *Evaluation de Langue Seconde (ELS)* の前に9~10ヶ月フランス語の学習を行った。その後、学習者のレベルは中級グループの B と上級グループの C と大きく2つに分けられ、さらにグループ内で上下の2つに下位分類された。*Commissions de la Fonction Publique du Canada (CFP)* によって開発されたインタビューモデルの大半は、*American Council on the Teaching of Foreign Language (ACTFL)* によって制作された *Oral Proficiency Interview (OPI)* から着想を得ている。

まず仮定 1) の検証のため、頻度 1001 位から 2000 位までの K2 と 2001 位以降の K3+Off-list のトークンの比率を用いて、グループ B とグループ C について t 検定を行った所、有意差が見られた(K2 :  $t=2.92$ ,  $p=.00$ ; K3+Off-list :  $t=8.96$ ,  $p=.00$ )。仮定 2) については、K3+Off-list におけるトークンの比率について学習者の4グループの差を分散分析で検定を行った所、有意差が見られた( $F=90.65$ ,  $p=.00$ )。また K3+Off-list についてチューキーの HSD を用いて多重比較を行った所、学習者の全てのグループ間に有意差が見られ、低頻度語(K3+Off-list)の単語数と学習者のレベルには強い関係があると結論付けられた。仮定 3) について、母語話者グループと各学習者グループの比較をカイ二乗検定で行った所、グループ B と母語話者グループには有意差が見られたが、よりレベルが高いグループ C については母語話者グループとの有意差が見られなかった。したがって、レベルが上がるにしたがって、母語話者の語彙使用に近似することが分かった。

#### 2.1 0.2.3. Thomas (2008)

Thomas (2008) は *VocabProfil* を使用して、留学生の話し言葉における語彙の発達を量的質的の両観点から分析した。被験者は、カナダのオンタリオ州南部の大学生で、フラン

スでフランス語プログラムの三年に在籍している英語母語話者 48 名である。比較対象として、オンタリオ州でフランス語を学習している 39 名の学生を選択した。発話タスクは 4 つあり、インタビュー形式の口頭試問、写真を基に状況記述を行うもの、自身のカナダの大学を宣伝するための記述文、そして自身の生き立ちについて語るものである。つまり、インタラクション型のタスクとモノログ型のタスクの 2 種類が行われている。新学年の始めと終わりにプレテストとポストテストとして同じタスクが行われた。

分析には、語彙の発達を観察するため、Ovtcharav et al. (2006)と同様に 2001 位以降の K 3+Off-list のトークンの比率に着目し、使用語彙の比率が 2 つのテスト間で上昇すれば、その被験者の語彙の発達を認めることができるという仮定を基に、分析を行った。その結果、フランス留学中の学習者も、カナダで学習している学習者ともプレテストとポストテストで比率の上昇が見られなかった。しかし、全体の発話量を見てみると、フランス留学中の学習者の発話量の伸びが 11.7%と顕著だったのに対して、カナダに残留中の学習者の発話量はわずかに 1.7%上昇するのみにとどまった。

また、頻度 3000 位以降である Off-list にリスト化された単語の性質を分析した所、質的な変化が見られた。留学生はフランス語圏で生活することによって、語の短縮、平俗もしくは品のない話し方、母語話者のような表現、そして « truc » や « machin » のようなコミュニケーションの中断を防ぐために使用される万能語の使用と行った者を習得したことが明らかになった。

この研究は、外国語学習者による語彙の量的かつ質的発達の分析を行ったものであり、第二言語研究分野において、この研究の意義は大きい。さらにこの研究は、英語学習者ではなくフランス語学習者を対象としている所が重要である。なぜなら、このような研究は英語教育で多く行われているが、その他の言語ではほとんど行われていないからである。

#### 2.1 0.2.3. Sugiyama (2012)

本論文は、日本語を母語とするフランス語学習者の口頭産出における語彙の豊かさを分析したものである。Ovtcharov et al. (2006) や Thomas (2008) による先行研究では、日本人学習者よりも言語能力の高い学習者を対象にしていたが、本研究では、フランス語学習歴 1 年半の学習者から留学歴のある学習者までの多様な学習背景を持つフランス語学習者を対象に行った。また、2 つの先行研究では *VocabProfil* を用いて各頻度層におけるトークンの比率についての分析を行っており、語彙の洗練性の測定のみが焦点が当てられていたが、本研究では語彙の洗練性の測定に加えて、各頻度層における語彙の多様性、そして内容語と機能語を測定する語彙密度の三つの観点をもって、語彙の豊かさの指標としている。

フランス語学習者 28 名についてこの指標を用いてクラスター分析を行った所、学習者のフランス語能力別にグループ分けを行うことが可能であることが分かった。また、グループ間の比較を行うため、一元配置分散分析を行った所、頻度上位 1 位から 2000 位までの内容語(K1 内容語)と 3000 位以下の単語(Off-list)について全グループ間で有意差が見られた。さらに各頻度層別の相関を調べた所、K1 内容語と 1001 位から 2000 位までの内容語 (K2)に強い相関が見られたため、日本人学習者の話し言葉における使用語彙の観点か

ら学習者を分類する場合には、これらの頻度層の語彙の多様性がレベル分けの指標と成りうることが分かった。

### 2.1 1. まとめ

本章では、まず語や語彙能力の定義を行った。そして、主に第二言語としてのフランス語習得における語レベルの研究が、これまでどのような発達過程を経てきたのかを俯瞰した。そして、現在ではコーパスを利用した語彙の頻度情報を使用した研究も多く行われていることが分かった。しかし、これらの研究の多くは、カナダ、イギリス、ベルギーで行われており、日本人フランス語学習者を対象とした研究は筆者の研究以外には存在しないことが分かった。

次章では、単語レベルよりもより大きな単位である **Multi-Word Units** についてその定義を行い、**Multi-Word Units** 研究の発展、そして中間言語研究との関係を概観する。

### 第三章 Word Units から Multi-Word Units へ

第二章では、語彙レベルにおける中間言語の語彙研究を概観した。コーパス言語学の多くの研究は語レベルから始まったが、コンテキストの中で語を見るためにより広い範囲の分析にシフトしつつある (Scott 1997: 234)。

実際に、言語は単語をただ単に並べることによって構成されているのではなく、非常に多くの定型句や慣用句などから構成されている。Bolinger (1976: 1) は、この現象について以下のような比喩を用いて説明している。

Language does not expect us to build everything starting with lumber, nails and blueprint,  
but provides us with an incredibly large number of prefabs.

(言語は我々に、木材、釘や青写真から、すべてを構築するようには望んでいない。言語は非常に多くのプレハブを我々に提供している。)

Kjellmer (1994: ix)は同様に、自然言語の性質について « bloc-like »と表現し、単語は同じクラスターの中で繰り返し使用される傾向にあると指摘した。言語を習得することは、単語レベル以上のかたまりについて、その使用感覚を身に付け、調節することであると言うことができるであろう。つまり、言語学習は単に単独の語が何を意味するのかを学習するだけではない。外国語による言語産出を行うには、頻度の高く、良く知られている語の連続体を使用できなくてはならない (Ellis 1997: 129)。

母語習得の際には、どの単語とどの単語と一緒に用いられる傾向にあるのかについて、日常生活で母語に触れることによって自然と学習している。ある単語においてどの意味が多用され、どの種類の単語と一緒に使用されるのかも習得するため、母語では特に意識せずとも同様の種類や意味の単語を代用できる。これは、コンピュータの中ではなく、記憶や経験の中に個人の膨大なデータを蓄積し、コーパスを構築しているようなものであるといえる (De Beaugrand 2001: 13)。

このように、母語話者は母語のスピーチコミュニティで育っていく過程で、潜在意識下で次第に語連続に関する知識を身に付けていくが、特に教室環境のみに学習機会が制限されている学習者の多くは、このような方法で知識を習得していくことはない。なぜなら、第二言語習得では目標言語との接触機会はかなり限定されるからである。多くの場合、学習者は文法規則の学習などの教室活動でしか目標言語に触れることはない。

学習者が目標言語を使用するとき、母語話者らしい言語の選択を行うということは、ただ統語規則を遵守しているということではない。母語話者のように話すということは、頻度が高く、広く使用されている表現や慣用句を多く用いて話すことである。言語学習者が勉強しなければいけないことは、これらの語連続を習得することである (Ellis 1997: 129)。

このとおり言語能力を示す重要な指標の一つは、単語以上の単位を流暢に使用しているかどうかであると考えられる。しかし、この分野の研究はまだまだ新しい分野であるため、この単位に関する明確な用語について規範が存在しない (Stubbs 2002: 230)。また、研究目的によってこの単位の数え方が異なることもある (Nation and Meara 2002: 36)。さら

に、学習者が発するこれらの単位について、研究者が容認可能かどうかを判断する確立された基準も無く、またこれらの要素が自由に結合されたものなのか、学習者がたまたま正しい表現を選択しただけなのかを断言することはできない (Nesselhauf 2005: 42)。

本章ではまずこの「単語以上の単位」についての定義を行う。

### 3.1. 単語以上の単位 - Multi-Word Units -

単語以上の単位は様々な名称で呼ばれており、異なる用語が同じ単位に使用されていたり、同じ用語が異なる単位を示したりしている (Granger and Paquot 2008: 27-28)。たとえば Nation and Meara (2002: 36)は、以下の3つの呼称例とそれらの違いについて挙げている<sup>32</sup>：

- 1) **Preformulated language** : 複数語から成る単位がそのまま使用可能な一つの単位として蓄積されている点を強調した呼び方である。
- 2) **Formulas** : ある事柄について述べる際に新しい言い方をその度に生成するのではなく、同じ言い方を繰り返す点を強調した名称である。
- 3) **Lexical phrases** : 日常生活である機能を達成する際に典型的に使用される句に焦点を当てたものである。これらの重要な特徴は、そのかたまり内の単語を他の単語と自由に置換することができないことである。

なお、Wray (2002)は「formulaic sequences」という用語を使用している。Wray は formulaic sequences の定義として、以下のように述べている (ibid: 9)<sup>33</sup>：

a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.

(単語やその他の要素から成る連続的、または非連続的なシーケンスは、既成のものであるかそのように見える。シーケンスは、文法規則による生成や分析に従っているのではなく、むしろ既に記憶に蓄えられ、使用時に記憶から取り出されるのである。)

この定義から、Wray は Nation and Meara の 1) Preformulated language と同じような性質を持つものを「formulaic sequences」と呼んでいることが分かる。

またこの他にも、Biber et al. (1999: 990ff) は *Longman Grammar of Spoken and Written English* の中で「Lexical bundles」という用語を初めて使用しており、この語連続は自然談話で通常使用されている語の連続体であると定義している。これは、上記の 1) と 2) を組み合わせたような概念であると言える。さらに、Ellis (1997: 129) は、「fluent units」という用語を用いている。これは、会話で通常速度かそれよりも速い速度で発せられる 4 語から 10 語で構成されている節であると定義しており、特に会話

<sup>32</sup> 下線は筆者が追加したもの

<sup>33</sup> 下線は筆者が追加したもの

面における語のかたまりについての側面に焦点を当てたものである。また、心理学の分野では George Miller が短期記憶の研究に « chunking » という用語を導入した (Miller 1956)。

また、語連続を構成する語の関係性に注目したものが、« collocation » である。複数の語の内部に強力な結びつきがある場合、collocation と呼ばれる (Nation and Meara 2002: 36)。Krishnamurthy (1987: 70) は collocation についての定義として、見出し語が両端最大 5 語の語彙項目によって伴われており、平均の期待値よりも高い頻度で現れる語彙項目のかたまりであると定義している。また Schmid (2003: 239-246) は collocation の定義として 5 つの基準を提案している：

1) 少なくとも 2 語で構成されている：

自明の理であるが 1 語のみでは連語をなすことができないため、collocation を構成することはできない。このように実質的には collocation は 2 語以上で構成されているが、典型的な多くの collocation は、2 語の連合体であることも事実である。

2) 問題となる語が近接している。

3) 共起頻度が高い：

もし、あるテキスト内で近接している語が単なる偶然ではなく、頻出する語のかたまりとして使用されているとすれば、それらは collocation としてみとめられる。

4) 相互期待値 (Mutual expectancy, Firth 1957: 195) と予測可能性 (Herbst, 1996: 389)：

相互期待値は、共起頻度の欠点を補うものであり、単純に文法構造によって出現頻度が高くなる語連続や、語連続に含まれる単語同士が互いに共起するかどうか予期できない語連続を排除する基準である。ある 2 語が共起されると考えられるとき、その言語の母語話者は、2 語の内の 1 語を見るだけで、もう 1 語の出現度合いについてある程度正確に予測することができる。

予測可能性は相互期待値を再解釈したものであり、予測可能性は相互期待値と語用的によく似ている。予測可能性は言語使用者の観点から語連結を見ており、相互期待値は語自体の言語内面的観点から語連結を見ている。また、先の基準である共起頻度の高さと相互期待値および予測可能性は、互いに独立しておらず、2 つの基準間に比例関係がある。collocation の頻度が高ければ高いほど、予測可能性が上がるからである。

5) 熟合度：

この基準は完全に透明な語連結から完全に慣用語的な語連結までの段階的な基準である。

このように、語連鎖の研究においてこれらを示す用語は統一されていない。呼び名は一定ではなく、さらに時には異なる観点に焦点をあてたものであり、指示対象となるものは全く同じものではないと言える。



Wray (2000) はこの語連鎖を指し示す名称として使用されているものを、以下のようにリスト化している。

amalgams	gambits	preassembled speech
		prefabricated routines and patterns
automatic chunks	gestalt	ready-made expressions
clichés	holistic	ready-made utterances
co-ordinate constructions	holophrases	recurring utterances
collocations	idiomatic	rote
composites	idioms	routine formulae
conventionalized forms	irregular	schemata
	lexical(ized) phrases	semi-preconstructed phrases that constitute single choices
F[ixed] E[xpressions] including I[dioms]	lexicalized sentence stems	sentence builders
fixed expressions	multiword units	stable and familiar expressions with specialized subsenses
formulaic language	non-compositional	stereotyped phrases
formulaic speech	non-computational	stereotypes
formulas/formulae	non-productive	stock utterances
fossilized forms	non-propositional	synthetic unanalysed chunks of speech
frozen metaphors	petrifications	
frozen phrases	praxons	

(Wray 2000: 465)

上記のように呼称は統一されていないが、Weinert (1995: 182)はこの傾向について以下のように指摘している：

while labels vary, it seems that researchers have very much the same phenomenon in mind.

(呼び名は変われども、研究者の間ではほとんど同じ現象について理解されているように思われる。)

最も一般的な意味で中立的に使用されている用語は、Multi-Word Units (以下 MWUs) である。MWUs は、実際の使用によって認められた語彙の固定パターンや繰り返しパターンであると大まかに定義することができる (Grant and Bauer 2004 : 38)。MWUs はこのように名称が一定ではなく、サブカテゴリー化の試みが行われてきたが、研究者の見解の一致を得られているものはない。どんなカテゴリー化も周辺的なケースを含み、カテゴリー間のオーバーラップが避けられないからである (Lewis 1993: 93)。

これらの MWUs を扱う研究をフレイジオロジーと呼ぶことがある。

### 3.2. フレイジオロジーの歴史的背景<sup>34</sup>

フレイジオロジーは全ての言語分野に精通しているが、学問分野として確立したのは近年になってからのことである (Granger 2005: 165)。しかし、MWUs 自体の研究の歴史は長い。1909 年には Bully によって完全に固定された句である « unité phraséologiques » とより固定度が緩い « series phraséologiques » の差についての研究が行われていた (Granger and Paquot 2008: 27)。

その後、1 世紀近い時を経てフレイジオロジーに関する研究が再興したのは、コーパスの登場が大きいと考えられる。コンピュータ化されたコーパスが入手可能となり、コーパス言語学的手法に多くの関心が寄せられるようになった。そのことによって MWUs に関する事象や概念が脚光を浴びることとなった (Stubbs 2002: 216)。

フレイジオロジーに関する研究が目覚ましい拡大を見せる一方で、混乱も見え始めている。それは、伝統的言語学に基づくアプローチと、データ駆動型アプローチの 2 つのアプローチの対立である (Granger and Paquot 2008: 28-29)。そこで、両者のアプローチを俯瞰する。

#### 3.2.1. 伝統的アプローチ

伝統的アプローチは、« significance oriented approach » (Herbst 1996: 380) や « phraseological approach » (Nesselhauf 2004) と呼ばれているものである。このアプローチはロシアの語句研究からの影響を大きく受けている。このアプローチを支持している研究者は辞書学や教育分野で研究を行っており、主な代表的研究者は A. P. Cowie、I. Mel'čuk、F. J. Hausmann である (Nesselhauf 2005: 12)。これはコーパス基盤型の研究で、あらかじめ選択された表現について、コーパスを使用してどのようにこれらの表現が使用されているのかを分析するアプローチである (Biber 2009: 276)。

徐々に改良されてはいるが、このアプローチにおける枠組として、MWUs を語レベルの単位と文レベルの単位に分類することは初期段階から行われていることであり、この点について研究者間でコンセンサスが得られている。語レベルの単位とは、単文ないしそれ以下のレベルにおいて統語的に機能するもので、文レベルの単位とは、格言、キャッチフレーズ、会話の定型表現といった語用的に機能するものである<sup>35</sup> (Cowie 1998: 5)。

さらに、語レベルの単位について、Vinogradov (1947) は 3 つの下位分類を提示した (Cowie 1998: 4-6 から引用) :

##### 1) phraseological fusion (または idiom)

非動機的、または意味的に不透明な語の結合によって構成され、一般的に構造的に固定されている ;

<sup>34</sup> この節における引用部分の MWUs は、かならずしも原典でも MWUs と書かれている訳ではなく、それぞれの用語で書かれていた。本論文では一貫性を持たせるため、MWUs に相当する全ての用語を MWUs と記載する。

<sup>35</sup> 下線部は筆者が付加したものである。

2) phraseological unity

部分的に動機的で、その意味が元来の中立的意味のメタファー的拡張とみなされるもの；

3) phraseological combination

最も興味深いのが、範囲を決定するのが最も困難であるものである。たとえば、2つの開かれたクラスの語から構成されている単位の場合、一つが文字通りの意味であり、もう一つが比喩的に使用されている。今日では、restricted collocationと呼ばれているものである。

このように、語レベルの単位では、非動機的で形式的に固定されているイディオムから、部分的に動機的で部分的に可変的なコロケーションまでが段階的もしくは連続的に並んでいる(Cowie 1998: 6)。なお伝統的に、最もイディオム的な単位がこの研究分野の中核として扱われていた(Granger and Paquot 2008: 28)。

また、Cowie は実際のテキスト内に見られる抽象的な語連続がコロケーションであると考え、特にイディオムとその対極にある「自由連結」との境界線を定めて、コロケーションの定義を行った(Nesselhauf 2005: 14)。つまり、Cowie のコロケーションの定義では、最も不定的で透明性のある句と自由結合 (free combination) を区別することに重点が置かれている。そして、自由結合はフレイジオロジーの領域外として認識されていた(Cowie 1998: 6)。

このように伝統的アプローチにおいて、語連続は、最も不透明で固定的な語連続から、もっとも透明性のある不定の語連続までの連続体を形成しているという考え方は一般的にコンセンサスを得られている。以上の分類をまとめたものが図5である。

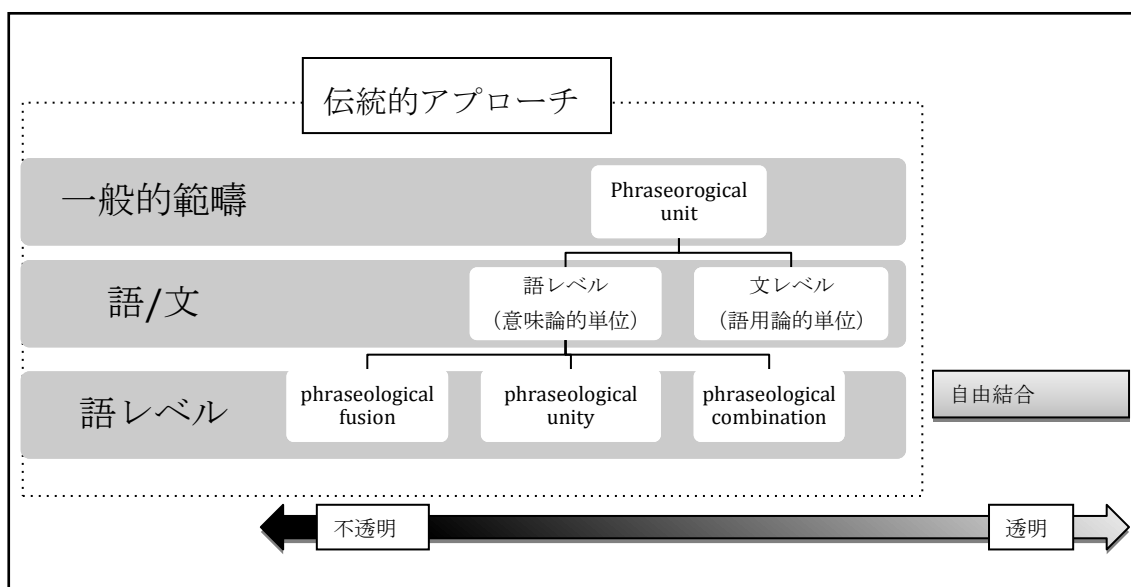


図5 伝統的なフレイジオロジーの分類

しかし、伝統アプローチの研究範囲であったイディオムやことわざといった最も制約的な MWUs は極めて出現頻度が少ないと言うことが、現在では明らかになっている(Moon 1998: 79)。

### 3.2.2. 頻度アプローチ

もう一つのアプローチは、言語内で単語が無作為に連結されるよりも高い頻度で起こる語連続が存在すると考える立場をとり、頻度アプローチと呼ばれる。このアプローチは、「statistically oriented approach」(Herbst 1996: 380)、「frequency-based approach」(Nesselhauf 2004)とも呼ばれるものである。これは遡ること J. R. Firth、特に M. A. K. Halliday や J. Sinclair が発展させた。頻度に基づくアプローチは、コンピュータに基づく連辞関係の研究に携わる研究者によって支持されている(Nesselhauf 2005: 11-12)。

頻度アプローチは、伝統的アプローチが主に扱ってきた語用的観点から、文体論や修辭論へと焦点を移行した。日常的な決まり文句の重要性を認めつつ、コーパス分析では MWUs を構成するテキストに重きを置いている。ここで分析の対象となる MWUs は、統語的、意味的な規則性を持ち、垣根表現や会話構成表現のような談話機能の範囲を満たしている。これらは主に、会話内であらわれる動詞の単位や節を構成する単位であり、書き言葉では名詞句や前置詞句も含む<sup>36</sup> (Granger and Paquot 2008: 34-35)。

頻度に基づくアプローチは、伝統アプローチが採用するトップダウン式のアプローチではない。語彙の共起関係を特定するために、ボトムアップ的なコーパス駆動型アプローチを使用する。この帰納的なアプローチによって、語連結のより拡大した範囲を説明することができる。また、伝統アプローチがとるコーパス基盤アプローチのように、あらかじめ定義されている言語カテゴリーに適合させるという手法をとらない (Moon 1998: 39, Granger and Paquot 2008: 29)。つまり、言語学の一般的な慣行では、分析に用いられる言語学的な構成概念にしたがって、あらかじめ前提をたてる。しかし、コーパス駆動型の分析では、単語の存在にしか前提を置かない。コーパス分析から引き出された単語間の共起パターンは、後の言語記述の基盤となる (Biber 2009: 276)。このように、コーパス駆動型の手法は研究目的の点から、コーパス基盤型の研究と異なる。コーパス駆動型の研究は、コーパスの機能的な分析を通して新しい言語の構成概念を明らかにすることが目的であるからである (Biber 2009: 278)。

MWUs の分析における急進的なコーパス駆動型のアプローチには 3 つの特徴がある (Biber 2009: 281):

- 1) レマではなく、コーパスに現れる実際の語形の分析に基づく。
- 2) 単語の文法的統語的状态を考慮せず、語形の連続について分析を行う。
- 3) 頻度が高く、繰り返し現れる語形の組み合わせに焦点を当てる。

---

<sup>36</sup> 前者の例として、I don't know, I think I might, what's the matter with、後者の例として the effect of, in the case of, the extent to which が挙げられている

このタイプの分析方法は母語話者や学習者の日常的に使用する MWUs の研究に良く適していると考えられている。なぜなら、それらの MWUs は良く使用され、一般的であり、心理的に顕著でないからである。また、恒常的に使用される MWUs について、広く同意されたリストがまだ存在しないため、この手法は有効である (De Cock 2004: 227)。つまり頻度アプローチでは、伝統アプローチの範囲外で周縁的なものとして捉えられてきたものに注目している。つまり、以前では興味のもたれなかった心理的にあまり目立たないような MWUs まで、フレイジオロジーのより広い視点から分析し、共起パターンを全て先験的な例外なくカバーする。これは、頻度アプローチの最大の利点である。たとえば、伝統アプローチで分析の興味外にあった自由結合でさえ、研究対象となりうる。実際、この自由結合と見なされる MWUs は一般的に言語使用の場で広く普及されている (Granger 1998: 3, De Cock 2004: 225, Granger 2005: 167)。

このアプローチでは、MWUs の抽出の方法によってコロケーション分析と N-gram 分析の2つのタイプに分けられる (Granger and Paquot 2008: 38-41)。なお、このアプローチにおけるコロケーションは、伝統的アプローチの指すコロケーションとは異なり、MWUs を構成する単語のそれぞれの意味を考慮に入れたものではなく、文字通り「共に用いられる単語」という意味でのコロケーションである：

#### 1) コロケーション分析

統計的に頻度の高い語連続を分析するものである。たとえば、ある単語の頻度が高いことと、語連続の頻度が高いことは等しくない。MWUs の頻度が高いということは、MWUs を構成する単語同士が連続して発せられる頻度が高いということである (Stubbs 2002: 235)。また Biber (1990: 990)は、この種の語連続について、熟合度や統語構造はいっさい関係がなく、ただ繰り返し現れる表現であると定義している。この単位の名称として、「コロケーション」や「コロケート」という用語が使用されている。この種の分析は *COBUILD dictionary project* の核となった分析方法である。また、ここには非連続的な MWUs も含む。

#### 2) N-gram 分析：

N-gram という用語は、自動言語処理の分野で多く用いられている。繰り返し起こる 2 語以上の語連鎖を抽出することを可能にする分析方法で、専門用語の抽出、言語変種に関する研究、中間言語研究、情報検索など、多様な目的で使用されてきた。この手法で抽出された語連続は「N-grams」と呼ばれ、2 語から成る場合は「Bigrams」、3 語から成る場合は「Trigrams」とも呼ばれている。さらには、「Lexical bundles」(Biber and Conrad 1999, Biber et al. 2003, Biber et al. 2004)、「Clusters」(Scott and Tribble 2006)、「Chains」(Stubbs 2002, Stubbs and Barth 2003)、「Recurrent sequences」(De Cock 2004)、「Recurrent word combinations」(Altenberg 1998)など、研究者によって異なる呼称が与えられている。また、ここには連続的な MWUs しか含まない。

これらの 2 つの違いとして、前者の分析から得られる結果は機能語と内容語の両方を含む高頻度の MWUs であるのに対して、後者は内容語の連続から成る専門的な MWUs であることが多い (Biber 2009: 277)。

語彙研究の分野において、コーパス駆動的アプローチは、より信頼性が高くより徹底した言語記述を可能にし、以前から取り組まれてきた多くの問題を解決してきた。しかしこのアプローチの登場によって、予期していなかった新しいタイプの複雑性も露呈したのも事実である (Rundell 2002: 141-142)。たとえばこのアプローチでは、共起回数が一定以上の場合、2 文字以上で構成される全ての語連続が抽出される。したがって、分離複合語やハイフン付きの複合語のような広い範囲の句については抽出することができるが、複数の単語が合わさって一語となっているような複合語を抽出することができない (Granger and Paquot 2008: 32-33)。

### 3.2.3. 二つのアプローチの融合

伝統的アプローチや頻度アプローチの登場は、それぞれ別の観点からフレイジオロジーの分野に多くの価値を与えた。しかし、伝統アプローチと頻度アプローチの支持者には今もなお広い隔たりが存在する。実際には、両アプローチが歩み寄ることのできるものも多い。伝統アプローチの支持者の多くは、コーパスに基づく自動抽出や分析から多くの恩恵を受けることができるということを認識していないように思える。一方、頻度アプローチの支持者は伝統アプローチに基づく言語分析から得られる恩恵がどれほど多いかについて正しく理解していないようである (Granger and Paquot 2008: 41)。そこで、この二つのアプローチの対立が緩和されるためには、まず、専門用語を厳密に明確化する必要があるだろう (Granger and Paquot 2008: 41-45)。

最近では、多くの研究者は代替可能性 (Commutability) や透明性 (Transparency) のような伝統アプローチの基準に加えて、頻度に基づくアプローチで使用される共起頻度の基準も使用している。つまり、両アプローチは時に併用されている。主に伝統アプローチを採用している研究者も、頻度を定義に用いる基準として捉えている。反対に頻度に基づくアプローチの枠組みで分析を行う研究者もまた、伝統アプローチ的な区別を行うこともある (Nesselhauf 2005: 16-18)。また、分析の多くは、コーパス基盤的手法とコーパス駆動的手法を組み合わせたものであることが多い (Biber 2009: 276)。

## 3.3. MWUs の測定法

### 3.3.1. コロケーション分析

重要なコロケーションを判断する最も単純な方法は、コロケーション自体の頻度を比べることであるが、この方法ではコロケーションを構成する個々の語の単独頻度は考慮されない。コロケーションを測定する方法として、石川(2012: 125-128)は 2 つの方法を挙げている：

#### 1) 共起強度の測定

これは、コロケーションを構成する個々の語に着目して、それらの共起強度を計量するものである。共起強度は、中心語と共起語の単独の頻度を基に、共起強度を補正した尺度のことで、コロケーション内の語の結びつきの強さを示す。代表的な指標として、ダイス係数や相互情報量が挙げられる。

#### 2) 共起有意性の測定

測定した共起が偶然による共起確率を超えて、真に意味のあるものになっているのかどうかを確認する測定法である。仮説検定の考え方に沿って、観測された共起が全くの偶然によるという仮説から、データがどのくらい離れているのかを統計値によって判断する。代表的な指標として、対数尤度比がある。

コーパス言語学では、スコアの大きさが有意性の指標とみなされているため、文脈によっては、共起強度と共起有意性の測定は一体的に扱われることもある(石川 2012:128)。

### 3.3.2. N-grams 分析

N-grams は、コロケーションとは異なり、あらかじめ決められた語彙に対する共起関係を調べるものではなく、テキスト内の全連鎖を対象とする。N-grams の抽出には、*WordSmith Tools* のようなツールを用いる。このようなツールは、まとまりがあるかどうかという主観的判断を行わず、語の連鎖の頻度を機械的に数え上げる(石川 2008:60)。

## 3.4. 中間言語 とフレイジオロジー

フレイジオロジーの分野は主に、伝統的アプローチにせよ頻度アプローチにせよ母語話者の間や母語話者のデータに基づいて研究が行われてきた。しかし、中間言語においてもフレイジオロジーの研究は大きな意味を持つ。本節では、中間言語とフレイジオロジーの関係について、理論的背景と実際のデータを用いて分析した研究を概観する。

### 3.4.1. 理論的背景

Pawley and Syder (1983)が指摘するように、母語話者は多くの MWUs を記憶しているため、正確で流暢に書いたり話したりすることができる。たとえば、この点について Altenberg (1998)は *London-Lund Corpus of Spoken English* を用いて、3語以上からなる MWUs について分析を行った。その結果、完全節、節構成要素、不完全句の3つの観点から、母語話者の MWUs の使用を明らかにした。その結果は、以下の通りである。

#### 1) 完全節

完全節の中で、独立節について分析を行ったところ、機能の大半は応答であった(*yes I do, that's all right* など)。また、*I don't know* や *I'm not sure* などの陳述緩和付加句も見られた。これらは、法的評言節として機能している。これらの

MWUs に特徴的な点は、語用論的に特殊化されており、特定の発話の状況に限定されている事であった。また、多くの MWUs は可変的で合成表現が可能であった。そして、拡張されて出来た変異形の方が拡張されていない中核表現よりも多く見られた。たとえば *I see* が単独で生起される回数よりも、*I see* の合成表現の方が 2 倍の頻度であった。

次に従属節を分析したところ、従属節は評言節(発話形式や発話内容について話者が意見を表明する節)、婉曲的条件(丁寧性やメタ言語的意見を表明するもの)、そして同格標識(言い直しを行うもの)の 3 つ機能に分類することが出来た。なお、評言節は *as it were, as you know*、婉曲的条件は *if I may, if you like*、同格標識は *that is to say* などである。

## 2) 構成的要素

主語の前におかれた主題要素から構成されているフレームの位置には、相互作用的、対人関係的、テキスト的機能を持つ MWUs が多く現れた (*and you know, well I mean, well you see* など)。主部と動詞を含む主観部には、陳述緩和機能を持つ *I think* や存在を意味する *there BE*、そして伝達機能を持つ *and he/she said* などが見られた。また、サンプル内で繰り返し生起される単独的節構成要素には、曖昧性付加句(*and so on, or something like that*)、緩和表現(*in a way*)、強意詞や数量詞(*the whole thing, a bit more*)、接続表現(*first of all*)、時間的表現(*at the moment*)、空間表現(*in this country*)などが見られた。

## 3) 不完全句

不完全句に特徴的だったのは、*as...as* や *a...of* のような、コロケーション枠が多く見られた点であった。

以上から、Altenberg は、母語話者は MWUs の巨大貯蔵庫のようなものを持っていて、そこから慣習化された建設ブロックを自由を使用していると結論づけた。

これらの MWUs は、第一言語習得だけでなく、第二言語習得にとっても有効であると広く認知されるようになってきている (Cowie 1998: 1)。なぜなら、MWUs を使用する際には処理労力をあまり要さないと考えられるからである (Schmid 2003: 255)。実際、“good morning”の各単語は学習者によって個別に認識されておらず、固まりとして学習され、使用されていると考えられる。つまり、各単語は既知のものから成り立っているが、その表現の使用頻度が高いために一つのかたまりとして捉えられている。このような MWUs も多く存在する (Nation and Meara 2002: 36)。

言語学習には当然、高頻度のコロケーション、句、イディオムなどの MWUs が含まれるべきであろう (Véronique 2008: 626)。また、学習者がより流暢に話したり書いたりできるようになるために、もし MWUs が有効であるのであれば、教科書に取り入れるべきである (Granger 1998: 3)。しかし、母語話者に近い言語習熟度が、さまざまな困難度や内部構造の固定度を持つ MWUs を蓄積することによって本質的に決定するという考え方は、生成文法に根付く原子論的見方と対立すると言える。なぜなら生成文法の考え方によれば、言語の働きは一般的な適応可能性のある規則体系、主に最小単位で構成されている語彙項目、そして一組の基本的な意味解釈の原理から説明できるからである (Cowie



1998: 1-2)。生成文法の流れを組むように、今日もなお語彙教育は、もっぱら単語が基盤である。しかし、特に非母語話者の教師は、言語のフレイジオロジーの側面に注意を払うべきであり、フレイジオロジーに関する研究によって効果的と認められた練習やツールを取り入れていくべきである (Granger and Meunier 2008: 251)。

中間言語研究におけるフレイジオロジーを分析するのに役立つ可能性があるのは、伝統アプローチと頻度アプローチを組み合わせることである。まず頻度アプローチを使用し、非常に幅広い概念から分析を出発するのが望ましい。しかし、量的手法によって確認された全ての MWUs が有益であるとは限らない。そこで伝統的なアプローチの理論によって補完される必要が出てくる。伝統アプローチによって、計量的単位を言語的に定義されたカテゴリーに細分化することができる (Granger 1998: 3)。

しかし、第二言語習得の分野では MWUs の研究は目立って行われてこなかった。研究の中心は、第二言語習得における創造的処理や規則に支配された処理の重要性や、学習者言語の体系的特徴に重点が置かれてきたが、MWUs の側面は無視されてきた (Myles et al. 1998: 324)。その原因の一つは、母語の研究でも MWUs に関する明確な定義の基準が設けられていないため、中間言語研究における MWUs の分析はより困難を極めるからだと考えられる。たとえば、学習者の産出言語の MWUs をエラーかエラーでないかという区分をはっきりと行うことはできない (Nesselhauf 2005: 39)。

しかし学習者言語の特徴として、Kjellmer (1991: 124)は以下のように比喻を用いて述べている：

(learner's) building material is individual bricks rather than prefabricated sections

((学習者の)建築資材は、既成の部分というよりは個々のレンガから成り立っている。)

つまり、学習者言語は語の連鎖よりも個々の語として特徴付けることができる。また、中間言語における MWUs の特徴として、Peters (1983, 1985)や Weinert (1995)の研究を基にして、Myles et al. (1998: 325)は以下の点を挙げている：

- 1) 少なくとも 2 つ以上の形態素から成り立つ。
- 2) 音韻的にまとまっており、流暢に発音されている。
- 3) 学習者の話し言葉における生産的パターンとは関係ない。
- 4) 学習者の別のアウトプットと比較して、複雑性が高い
- 5) 同じ形を繰り返したり、いつも使ったりする。
- 6) 統語的、意味的、語用的に不適切、または特異である可能性がある。
- 7) 状況に依存する。
- 8) 使用は同じ学習者グループ内に見られる。

さらに、Myles et al. (1998: 327)は、学習者の MWUs の使用理由について、ストラテジーの観点からもまとめている。まず、学習者の MWUs の使用はコミュニケーションストラテジーであると考えられる。つまり、学習者は MWUs を目標言語の規則を補うために使用している。また、生産における流暢さや処理を早めるための生産的ストラテジーとし

て使用される場合もある。しかし、第一言語習得研究と第二言語習得研究の両方で、三つ目の機能について広く議論されている。それは、学習者が非分析的発話を模倣することで文法能力が発達するという機能である。議論の中心は、学習者が非分析的発話を徐々に分解し、新しい発話を生成する際に生産的にそれらの部分を使用し始めるのかどうか、または、学習者は話し言葉のインプットリストの中から単に機械的に学習した発話を引き出しているのかどうかについて扱われている (Myles et al 1998 : 327)。

語の連鎖を分析することで学習者の「外国語らしさ」や発達状況を特定することは可能であるかもしれない。

#### 3.4.2. 中間言語データに基づく MWUs の分析

コーパスを用いた中間言語における MWUs レベルの分析は、これまで書き言葉に関しては行われてきたものの、話し言葉分析ではあまり行われていない。話し言葉における MWUs 分析について、英語学習者に関して行われた研究である De Cock et al. (1998)、De Cock (2004, 2007)、そしてフランス語学習者について行われた研究である Myles et al. (1998, 1999)をまとめる。

##### 3.4.2.1. De Cock et al. (1998)

本論文は、フランス語母語話者である上級学習者の話し言葉コーパスに基づいた MWUs の研究である。学習者言語の外国語らしさについて、Kjellmer (1991:124)が「健在に既成のレンガではなく、自前のレンガを使っている」との仮説を立てているが、この仮説の妥当性を検証することを目的としている。母語話者コーパスと学習者コーパスに基づいて対照アプローチから学習者の MWUs の能力を比較分析している。

分析には、ベルギーのフランス語母語話者である英語学習者の口頭産出データ 25 名分の 62975 語から成るコーパスと、英語母語話者 25 名分の 80448 語のコーパスを用いた。まず、2~5 語から成る MWUs を抽出した。抽出の際、MWUs の長さと同頻度が反比例することを考慮にいれ、2 語連鎖は 10 回、3 語連鎖は 5 回、4 語連鎖は 4 回、5 語連鎖は 3 回を出現頻度の閾値として設定し、それ以上の頻度のものを分析対照とした。

その結果、英語学習者の MWUs の数が母語話者と同じか、それより若干高いことが分かった。さらに、4 語連鎖と 5 語連鎖については、学習者の MWUs の数が有意に高い。したがって、Kjellmer の仮説に反して、学習者は既成表現を良く使用していることが分かった。しかし、学習者の既成表現の種類は一部に限られていた。これは、学習者の既成表現が繰り返しやフィラーを含むものであることから裏付けられる。また、学習者の過少使用に顕著だったのは曖昧表現であり、母語話者は学習者の約 4 倍もこれらの表現を使用していることが分かった。

#### 3.4.2.2. De Cock (2004)

この研究は、英語母語話者と英語学習者の語連鎖について比較を行ったものである。学習者コーパスは、*LINDSEI (Louvain International Database of Spoken English Interlanguage)*の一部で、ベルギーのフランス語母語話者で英語学習者である 50 名分の口頭産出データを使用した。なお、これらの学習者は上級者である。参照コーパスとして、英語母語話者であるランカスター大学の学生 50 名分のデータを使用した。2 つのコーパスとも、インタビュー形式の会話に基づいている。語連鎖数の比較における、出現回数の制約として、2 語連鎖には 12 回以上、3 語連鎖には 6 回以上、4 語連鎖には 4 回以上、5 語連鎖と 6 語連鎖には 3 回以上現れた語連鎖を選択した。

まず、母語話者と学習者の語連鎖の数を比較した所、タイプについて学習者は 2 語連鎖、3 語連鎖、5 語連鎖をわずかながら多く使用していたが、統計的有意差は無かった。4 語連鎖については、学習者が有意に多く使用していた。トークンについては、学習者が 3 語連鎖、4 語連鎖、5 語連鎖を有意に過剰使用しており、2 語連鎖と 6 語連鎖についてもわずかに過剰使用していた。しかし、学習者のこれらの語連鎖には、繰り返しやためらい語などが高い割合で構成されていた。実際に、母語話者と比較すると学習者はこれらの特徴を含む語連鎖を 3 倍から 4 倍使用していた。したがって、繰り返しやためらい語などを除去した後、母語話者と学習者の語連鎖について分析を行った。その結果、タイプもトークンについても 2 語連鎖から 6 語連鎖までの分析した全ての語連鎖を過少使用する傾向があった。また、学習者によって使用された語連鎖は、母語話者の語連鎖と比較すると、インタラクションを促すような語連鎖や会話への参加を示すような語連鎖をあまり用いていないことが分かった。

#### 3.4.2.3. De Cock (2007).

この論文は、話し言葉英語における母語話者と上級学習者の節に関する MWUs について、記述を試みたものである。分析には、母語話者コーパスとして *LOCNEC* を、学習者コーパスとして *LINDSEI* を用いた。前者は 117417 語、後者は 90300 語である。MWUs の抽出として、2-gram、3-gram、4-gram、5-gram そして 6-gram について、それぞれ生起頻度が 12 回、6 回、4 回、3 回、3 回以上のものを母語話者コーパスと学習者コーパスから選択した。

学習者の MWUs の傾向として、繰り返しまたはためらい語を含む MWUs が多いことが分かった。また、機能の観点からは、母語話者コーパス、学習者コーパスとも、会話やターンの始めにインタラクションに必要な MWUs が多く見られた。しかし、質的な面では両者に差異が見られた。学習者は、特にインフォーマルなインタラクションに見られるような *you know, I mean, just, sort of, like, cos, yeah* を含む MWUs を過少使用していた。そして、好き嫌いなどの態度を示す MWUs についても違いが見られた。学習者は *like* という動詞に主に限定されていたが、母語話者は *like* の他にも *enjoy* や *love* などの単語の使用が多く見られ、母語話者の方が MWUs のバラエティが広い。

#### 3.4.2.4. Myles et al. (1998)

Myles et al.は、特に3つの非分析的 MWUs である *j'aime*, *j'adore*, *j'habite* に注目し、学習者は非分析的 MWUs を徐々に分解できるようになるか、新しい発話を生成する際に生産的にそれらの部分を使うことができるかどうかを研究した。そして、MWUs の分解に関する研究によって学習段階を推察できるかどうかについても考察を行った。

データは、*Progression in Foreign Language Research Project* で構築されたコーパスの中から抽出した。データサンプルは、2つの学校からそれぞれ男女4名ずつで合計16名のデータから構成されている。これらの被験者は、フランス語を学習し始めて1年後に語学教師によって能力別にグループ分けされ、その結果8名は能力が高く、5名が中程度の能力、そして3名が低度の能力であると判断された。

結果として、以下の点が明らかになった：

- 1) MWUs の分解が関係する場合、学習者には明らかな発達の連続体が存在することが分かった。早い段階で分解が始まる学習者と、動詞から代名詞主語を分節するのに長く時間がかかる学習者と、そしてまだその段階にもさしかかかっていない学習者があり、段階的な発達段階がある。
- 2) 無意識的に MWUs を分解できるかどうかは、非定型句において、人称の選択などの主語代名詞システムに独創的に現れることと関連する。
- 3) 発達の連続体において、学習者を3つのカテゴリーに分類することができる；
  - a) 研究の終盤までに MWUs を構成要素に分析的に判断することができ、主語代名詞を自由に使用することができる学習者
  - b) MWUs の分解の発達が遅く、完璧ではないが一応 MWUs の構成要素を分析的に判断ができる学習者
  - c) 二年間の間に全く発達が見られなかった学習者

また、初期段階において MWUs はコミュニケーションの手段として役立つが、学習が進むにつれて、MWUs を構成要素に分類し、人称の変化など MWUs を状況に合わせて正しい形に直せない場合、コミュニケーションの阻害となる。

#### 3.4.2.5. Myles et al. (1999)

本論文は、英語母語話者である初級フランス語学習者の発話における疑問形を含む MWUs の役割についてまとめたものである。学習者と母語話者の MWUs の使用が異なることを指摘し、学習者の MWUs の特徴として以下の6点を挙げている：

- 1) 学習者の別のアウトプットと比較して、MWUs の長さが長く、複雑である
- 2) 同一イントネーションカーブ内で途切れが無く、流暢でよどみない一貫した音韻体系をもつ
- 3) 統語的、意味的、語用的に誤って使用されている

- 4) コンテキストにあわせて MWUs 内を別の単語に代替することができず、常に同じ形を使用する
- 5) 学習者の別の発話と比較して、文法的に難しい形を使用する
- 6) 学習者に特徴的な MWUs は、限定的な状況やコンテキストで現れる。特に教室活動で使用される表現について、MWUs のバリエーションが広い

学習者は言語規則を適応して生成するのではなく、これらの特徴を持つ MWUs を記憶していると仮定し、学習者の学習段階が進むにつれどのように疑問形を含む MWUs が分解されるかを分析した。

データとして、*Progression in Foreign Language Research Project* で構築されたコーパスから無作為に 16 名の会話データを抽出した。このデータは長期データである。会話データ収集初回の学習者の年齢は全て 11 歳から 12 歳であり、録音は 3 年にわたって行われた。主なタスクはペアによるインフォメーションギャップタスクで、テキストや絵を見てペアの間で質疑応答した。

その結果、MWUs の分解は学習が進むにつれて進み、疑問形を含む MWUs は後の分析的・創造的な表現の生成の基礎となることが分かった。学習者は初期段階から *comment t'appelles-tu* のような疑問詞+倒置形といった複雑な構文を MWUs として記憶しており、学習初期段階では記憶した MWUs を三人称が主語となる場合のような不適切なコンテキスト下でそのまま使用していた (例: \**Comment t'appelles-tu un garçon ?*)。しかし、徐々に適切な形に応用して用いることができるようになった。しかし、この発達段階をスムーズに踏むことができる学習者がいる一方で、学習初期に学習した MWUs の記憶の保持ができず、動詞を欠く疑問文の段階に逆戻りしてしまう学習者もいた (例: \**je grand maison?*)。つまり、その発達方法やスピードの差は学習者の間によって認められた。

### 3.5. まとめ

本章では、2 語以上から構成されている MWUs について、その定義や歴史的背景を俯瞰し、中間言語に置ける MWUs の重要性を確認した。中間言語の話し言葉における MWUs の使用に関する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よる MWUs の使用の研究は確認できなかった。

次章では、本論で扱う研究について方法を説明する。

## 第四章 研究方法

### 4.1. 研究目的と研究背景

本研究は、日本語を母語に持つフランス語学習者を対象に、口頭産出における使用語彙を記述することを目的としている。本稿で扱うフランス語学習者の間には学年、学習の背景など様々な差が存在するが、これらの差が使用語彙の差にどのような影響をもたらすのかも調査の対象とする。

使用語彙の分析は、Word Units (WUs) レベルと Multi-Word Units (MWUs) レベルの 2 つの観点から行う。WUs レベルの分析と MWUs レベルの分析を行うことで、学習者の使用語彙を幅広く記述できると考える。

前章までで俯瞰した先行研究によって、このような学習者コーパスに基づく学習者の使用語彙研究は英語学習者に関するものが大半であることが分かった。しかし、コーパス分析で使用される手法は英語にのみ通用するものではなく、他の言語にも応用することは可能である。実際に英語学習者の分析によって、多くの有意義な示唆が得られている。手法をフランス語学習者に適用し、フランス語教育への示唆が得られることが期待できる。日本人フランス語学習者に特化したこのような研究は今のところ存在せず、先行研究の手法を応用し、日本人フランス語学習者に焦点を当てた研究は存在意義があると言える。

本研究では「語彙の豊かさ」、「特徴語」、「コロケーション」、「N-gram」という 4 つの観点についてコーパスに基づく分析を行い、包括的に学習者の使用語彙について記述を行う。Nation (2001: 31) は、コーパス研究の必要条件として以下の 3 つを挙げている：

- 1) データの拠所となるコーパス
- 2) コーパス研究によって明らかにできる良質な研究設問
- 3) コーパスからデータを系統立てる作業を簡素化できるコンピュータプログラム

本章ではまず、本研究で使用するコーパスデータの説明を行う。次に、4 つの研究のリサーチクエスチョンを述べ、最後にデータ処理方法であるコンピュータプログラムの説明を含む分析方法について説明を行う。

### 4.2. コーパスデータ

使用するコーパスは、全部で 4 つある。それぞれ、学習者によるタスクに基づくコーパス、母語話者によるタスクに基づくコーパス、学習者による自由会話コーパスである IPFC-JP コーパス、そして母語話者による自由会話コーパスである『フランス語(エックス)多言語話しことばフランス語コーパス』である。二種類のコーパスを使用したのは、より幅広い学習者の使用語彙を分析するためである。学習を始めて 1 年強の学習者にとって、自由会話をするのは難しいと考えられるため、このような学習者はタスクに基づいて会話を行った。それ以上のレベルを持つと考えられる大学学部 3 年生以上の学習者は自由会話を行った。なお、タスクに基づくコーパスデータは Ellis (1994) の分類におけ

る「言語使用データ」の「誘導型・臨床的抽出データ」であり、自由会話コーパスデータは「言語使用データ」の「非誘導型データ」であると言える。このように、学習者のレベルもデータの性質も異なるため、積極的に2つのデータを比較することはしない。

#### 4.2.1. タスクに基づくコーパス(学習者コーパス)

##### 4.2.1.1. 対象者

コーパスは東京外国語大学2年生39名(男10名女29名)の協力のもと、2007年4月26日から7月19日にデータを収集した<sup>37</sup>。最終タスク実行後に、フランス語学習について、タスクについてのアンケートを行い、26名の回答を得た。アンケートは以下の3点についての質問から構成されている：1)フランス語学習歴 2)フランス語圏への滞在歴 3)タスクの難易度。学習歴に関しては、以下のような結果となった。

表2 フランス語学習歴<sup>38</sup>

フランス語学習歴	人数
1. 大学入学時より	22 (2)
2. 大学入学以前に個人的に学んだことがある	1
3. 高校の第二外国語として学んだ	3 (1)
4. 大学入試でフランス語選択だった	0
5. フランス語母語話者である	0

大半の学生は、大学入学以降フランス語を学び始め、アンケート集計時点で計15か月間学んでいることになる。大学入学以前に個人的にフランス語を学習したことがある学生1名と高校でフランス語を第二外国語として学んだ学生が3名いたが、タスクを行った時点では、大学入学時よりフランス語を学びはじめた学生とは大差がないと考える。

フランス語圏への滞在期間に関する質問の結果は、以下のとおりである。

表3 フランス語圏滞在期間

フランス語圏滞在歴	人数
1. フランス語圏に行ったことがない	18
2. 旅行で行ったことがある	8
3. 6か月未満、フランス語圏に語学留学したことがある	0
4. 6か月以上2年未満、フランス語圏に語学留学したことがある	0
5. 2年以上、フランス語圏に住んだことがある	0

フランス語圏へ長期間滞在したことのある人は皆無であった。すべての学生が日本でフランス語を学ぶ環境におり、フランス語母語話者やフランス語におけるメディアとの接

<sup>37</sup> 実験に協力頂いた東京外国語大学川口裕司教授、学生の皆様に感謝する。

<sup>38</sup> カッコ内の人数は、2年生の履修が2回目の者の人数である。

触がかなり限定されているため、フランス語の能力に関しては大きな差がつかないと考える。

タスクの難易度に関しては、以下のようになった。

表 4 タスク難易度

タスクの難易度	人数
1. とても簡単	0
2. 簡単	1
3. 普通	10
4. 難しい	15
5. とても難しい	0

フランス語圏滞在歴より、フランス語能力に大差はないと仮定したが、タスク難易度に関する質問からも、学習者間でフランス語能力の差がないことが言える。ほとんどの学生がタスクについてやや難しい、もしくは普通であると考えている。タスクは学生のフランス語能力を適切に引き出せるものであったといえる。

このアンケートの結果より、学習者のフランス語能力はある程度均一であることがわかった。そして、性別や年齢の差を含めた個人差は、このコーパス構築においては重視しない。

学習者が行ったタスクはすべてヨーロッパ共通参照枠の A1 レベルに基づくもので、大半の学習者はそのタスクを自分のレベルと同等もしくはそれ以上のものとみなしているため、このコーパス構築に参加した学生は、フランス語初級学習者のグループ、つまりヨーロッパ共通参照枠の A1 レベルから A2 レベルであるとする。

#### 4.2.1.2. タスク

タスクは、筆者が5つの教材より選択した。いずれも、ヨーロッパ共通参照枠の A1 レベルに公式に基づいた教材である。

- 1) Beacco, M. et Beacco, J-C. (2007) *Alors ? Niveau A1*. Didier ;
- 2) Breton, G., Cerdan, M. et Dayez, Y. (2005) *Réussir le DELF Niveau A1 du cadre européen commun de référence*. Didier ;
- 3) Crepieux, G. et al. (2006) *Spirale*. Hachette-Pearson Education Japan ;
- 4) Grandet, Eliane (2001) *DELF A1/A2/A3/A4 : 450 activités*. CLE International ;
- 5) Parizet, M., Grandet, E., Corsain., M. (2005) *Activités pour le cadre européen commun de référence : Niveau A1*. CLE International.

学習者は、モノログ形式とインタビュー形式の2種類のタスクを行った。モノログ形式のタスクを行う際は日本語で指示を読んでフランス語で答え、インタビュー形式のタスクは、フランス語で指示を聞きフランス語で回答した。以下は学習者が行ったタスクのリストである。



表 5 タスクのリスト<sup>39</sup>

タスク番号	タスク内容
j001	あなたの好きな季節は何ですか？
j002	長期休暇中、どこで過ごすのが好きですか？
j003	あなたは、今日携帯を忘れてしまいました。携帯を忘れたことを伝え、それを貸してくれるように友達に頼むときには何といいますか？
j004	あなたの友達が、携帯を貸してくれるようあなたに頼んできました。携帯を貸してあげるときに、あなたは何と言って貸してあげますか？
j005	あなたの友達が、携帯を貸してくれるようあなたに頼んできましたが、あなたは携帯を貸したくありません。何と言って断りますか？
j006	今晚レストランで夕食を一緒に食べに行こうと、友人に誘ってください。
j007	あなたは、友人から映画に誘われました。誘いを受け入れて、会う約束をしてください。
j008	あなたの家から学校までの行き方と、それにかかる時間を説明してください。
j009	あなたは旅行者から成田空港で、上野駅までの行き方を尋ねられました。 以下の目的地への行き方と所要時間を参照して説明してください。 成田 → 日暮里 → 上野 京成線 山手線 (1時間10分) (5分)
j010	普段どんな映画をよく見ますか？フランス語でどんな映画か説明してください。
j011	普段どんな本を読むのが好きですか？フランス語で説明してください。
j012	今日の昼ごはんは何を食べましたか？
j013	普段、自宅では何を飲みますか？
*j014, *j015, *j016	あなたは観光で訪れていたパリで、アンケートの協力を求められました。 質問を聞いて、それにフランス語で答えてください。 1. quelle est votre nationalité? 2. ou êtes-vous née? 3. quelle est votre profession?
j017	あなたは、旅先で bio Natura というお店を探しています。道を歩いている人にそのお店の場所を聞く時、何と言いますか？
j018	旅先で入ったレストランのメニューの中に、聞きなれない料理がありました。あなたはお店の人に何と尋ねますか？
*j019, *j020, *j021,	あなたはフランス人の知り合いにフランス語で話しかけられました。 質問を聞いて、それについてフランス語で即答してください。 1. Qu'est-ce que vous aimez faire le week-end? 2. Qu'est-ce que tu aimes faire à la maison? 3. Qu'est-ce que vous voudriez faire ce week-end, rester a la maison ou sortir?
j022	あなたは今、フランスでカフェに来ています。 コーヒーを頼むことにしました。 あなたはウエーターに何と言って注文しますか？

<sup>39</sup> タスク番号の前にアスタリスク(\*)のあるものは、インタラクティブ形式のタスクであり、その他はモノローグ形式のタスクである。

j023	あなたは今、フランスでカフェに来ています。 友達の Fabien が何のドリンクを頼むのか、あなたは知りません。 友達に何のドリンクを注文するのかを聞く時、何と言いますか？															
j024	自己紹介してください。(苗字、名前、国籍、年齢、職業、住所、家族構成など)															
j025	以下の身分証明書をよく見てください。 Nom: Tanaka Prénom: Kazuko Nationalité: Japonaise Âge: 38ans Profession: Professeur Situation de famille: Mariée – 2 enfants Adresse au Japon: 3 Asahi-cho Fuchu Adresse en France: 67, rue de la Liberté 69000 Lyon この人物を紹介してください。															
j026	あなたの親友について話してください。															
j027	週末にあなたがしたいこと、そうでないことについて話してください。															
j028	都会と田舎どちらが好きですか？フランス語で答え、その理由も述べてください。															
j029	何をするのが嫌いですか？フランス語で答え、その理由も述べてください。															
j030	あなたが最近行った旅行について話してください。															
j031	あなたの先週末の過ごし方について話してください。															
j032	あなたはこの週末一緒に出かけられるかどうかを聞こうと、友人に電話をしています。ところが、留守番電話に接続されました。「いつ」、「どこで」、「どんなことをしたいのか」を留守番電話に吹き込み、友人を誘ってください。															
j033	あなたは、友人からサッカーの試合を見に行こうと誘われました。あなたはサッカーが好きではないので誘いを断りたいと思っています。理由を添えて、誘いを断ってください。															
j034	<table border="1"> <thead> <tr> <th>Gares de départ et d'arrivée</th> <th>Horaires</th> <th>Trains</th> </tr> </thead> <tbody> <tr> <td>PARIS</td> <td>14h40</td> <td>TGV</td> </tr> <tr> <td>BORDEAUX</td> <td>17h39</td> <td></td> </tr> </tbody> </table> <p>あなたは、Paris から Bordeaux までの列車の片道チケットを1枚予約します。 このチケットを予約してください。 (出発地、目的地、時間、片道チケット、枚数を情報に入れてください。)</p>	Gares de départ et d'arrivée	Horaires	Trains	PARIS	14h40	TGV	BORDEAUX	17h39							
Gares de départ et d'arrivée	Horaires	Trains														
PARIS	14h40	TGV														
BORDEAUX	17h39															
j035	<table border="1"> <thead> <tr> <th>Gares de départ et d'arrivée</th> <th>Horaires</th> <th>Trains</th> </tr> </thead> <tbody> <tr> <td>PARIS</td> <td>14h40</td> <td>TGV</td> </tr> <tr> <td>BORDEAUX</td> <td>17h39</td> <td></td> </tr> <tr> <td>BORDEAUX</td> <td>18h04</td> <td>TGV</td> </tr> <tr> <td>TOULOUSE</td> <td>20h12</td> <td></td> </tr> </tbody> </table> <p>表を参考にして、Paris から Toulouse までの道順と出発・到着時間を説明してください。</p>	Gares de départ et d'arrivée	Horaires	Trains	PARIS	14h40	TGV	BORDEAUX	17h39		BORDEAUX	18h04	TGV	TOULOUSE	20h12	
Gares de départ et d'arrivée	Horaires	Trains														
PARIS	14h40	TGV														
BORDEAUX	17h39															
BORDEAUX	18h04	TGV														
TOULOUSE	20h12															
j036	普段どんなテレビをよく見ますか？フランス語で説明してください。															
j037	あなたは普段、朝ごはんは何を食べますか？															
j038	レストランでは普段、何を飲みますか？															

*j039, *j040, *j041, *j042	あなたは観光で訪れていたパリで、アンケートの協力を求められました。 質問を聞いて、それにフランス語で答えてください。 1. Vous vivez dans cette ville depuis longtemps? 2. Vous pouvez me parler de l'endroit ou vous habitez? 3. Qu'est-ce que vous allez faire aujourd'hui? 4. Vous faites souvent du sport?
j043	あなたは、何曜日が一番好きですか？なぜですか？
j044	あなたは、食べ物で何が一番好きですか？なぜですか？
*j045, *j046	あなたはフランス人の知り合いにフランス語で話しかけられました。 質問を聞いて、それについてフランス語で即答してください。 1. Est-ce que vous avez une voiture? 2. Tu as un chien?
j047	あなたは今、パリのキオスクに来ています。 L'officiel des spectacles という、パリの週刊情報誌を買おうと思っています。 あなたは何と言って店員に頼みますか？
j048	あなたは今、パリのキオスクに来ています。 L'officiel des spectacles がいくらなのか、あなたは分かりません。 何と言って値段をたずねますか？
j049	あなたは来週、大事な試験を受けることになっています。その気持ちをフランス語で表現してください。
j050	あなたの親友が、一年間海外留学することになり、しばらく会うことが出来ません。その気持ちをフランス語で表現してください。
j051	あなたのフランス人の友達が、バカロレアに合格しました。あなたは何と言いますか？
j052	あなたのフランス人の友達が結婚することになりました。あなたは何と言いますか？

#### 4.2.1.3. 録音システム

学習者は、東京外国語大学が開発した e-learning システムを使用してタスクを行い、録音した<sup>40</sup>。コンピュータ画面上で指示を読み、ヘッドセットを使ってコンピュータ上で録音を行った。このシステムによって、6つのタスクを行うことが可能になる：

- 1) 指示を読み、口頭で答える
- 2) 指示を聞き、口頭で答える
- 3) 絵や図を見て、口頭で答える
- 4) 指示を読み、記述する
- 5) 音を聞き、記述する
- 6) 絵や図を見て、記述する

<sup>40</sup> 本研究に用いた録音システムを開発された、林俊成教授に感謝する。

([http://www.el.tufs.ac.jp/html\\_src/login.htm](http://www.el.tufs.ac.jp/html_src/login.htm) からアクセス可能)

今回のコーパス構築には1番目と2番目のタイプのタスクを行った。学習者は録音前に準備する時間が少し与えられている。しかし、辞書などを使用することは一切禁止されており、録音前にメモすることも禁じられている。

データは指名、学籍番号、回答、音声、回答日時などが含まれ、Excelの形式でダウンロードすることができる。

	A	B	C	D	E	M	N	O	P
1									
2		ID	章-節	チャプター名	説明文	生徒名	学籍番号	回答 音声	回答時間
3		439	1-1	教室用4月26日	以下の説明を読み、フランス語で答えてください。			音声	2007/4/26 13:30
4		444	1-1	教室用4月26日	以下の説明を読み、フランス語で答えてください。			音声	2007/4/26 13:32
5		474	1-1	教室用4月26日	以下の説明を読み、フランス語で答えてください。			音声	2007/4/26 13:39
6		464	1-1	教室用4月26日	以下の説明を読み、フランス語で答えてください。			音声	2007/4/26 13:36
7		488	1-1	教室用4月26日	以下の説明を読み、フランス語で答えてください。			音声	2007/4/26 13:46
8		453	1-1	教室用4月26日	以下の説明を読み、フランス語で答えてください。			音声	2007/4/26 13:33
9		443	1-1	教室用4月26日	以下の説明を読み、フランス語で答えてください。			音声	2007/4/26 13:31
10		481	1-1	教室用4月26日	以下の説明を読み、フランス語で答えてください。			音声	2007/4/26 13:42
11		483	1-1	教室用4月26日	以下の説明を読み、フランス語で答えてください。			音声	2007/4/26 13:44

図 6 Excel形式でダウンロードしたデータ

#### 4.2.1.4. 転写

転写は、学習者自身によるものと、筆者によるものの2段階で行った。学習者はタスクを行ったあと、転写するページを開き、自身の発話を聞きながらその発話を記述する。

**Answer**

録音後に送信ボタンを押してから次のページへ進むボタンを押してください

下記に録音した音声内容を書き留めて下さい

図 7 録音・転写ページ

しかし、学習者による転写には、綴りの間違い、文法のエラー、転写の欠如などの問題があった。このような問題は、分析時に大きな問題を引き起こすことが考えられるため後に筆者が録音と学習者による転写を比較し、必要な時にはそれらの間違いを訂正した。また語彙サイズを把握するにあたり、一般的に固有名詞、フライングやいい間違いなどは一般的に語彙の総数からは除外されている (Milton 2009 : 9)。そのため、本研究においても分析に公平性が出るよう、また、転写の訂正が一貫して行えるよう、いくつかの決まりを定めた：

- 1) 綴りや活用間違いは発音を優先し、発音が正しい場合訂正する(例 : \*Je m'apele-> Je m'appelle)
- 2) 発音上に違いがなく、転写時に性数一致を間違えている場合も同様に発音を優先し、訂正する(例 : \*bon restaurants -> bons restaurants)
- 3) 発話が不明瞭な場合や意味が通じない場合は、学習者による転写を優先する
- 4) 学習者が自身の発話の間違いに気づき、転写時に修正した場合は、録音時の発話の方を優先する(例 : [bɔ̃]と発音しているのにもかかわらず、belle と転写しており、後者が文法的に正しい場合でも、bon もしくは bons に書き換える)
- 5) ためらい語の転写を一貫させるため、全て HESIT に置き換える。
- 6) 固有名詞は全て NP に置き換える。
- 7) 日本語や英語を用いて、フランス語発音を行っていない場合は FW に置き換える。
- 8) 間投詞は全て INT に置き換える。
- 9) 語を途中で言いかけて言い直した場合は転写を行わない。

#### 4.2.2. タスクに基づくコーパス(母語話者コーパス)

##### 4.2.2.1. 対象者

コーパスはボルドー第三大学の学生 39 名(男 14 名女 25 名)の協力のもと、2012 年 4 月 26 日から 6 月 7 日にデータを収集した<sup>41</sup>。年齢は 18 歳から 28 歳までであり、ほとんどの被験者の学年は大学 1 年生であるが博士課程の学生も 1 名いる(図 8 と図 9 を参照)。

---

<sup>41</sup> 本録音に協力していただいたボルドー第三大学の皆様に感謝の意を表す。

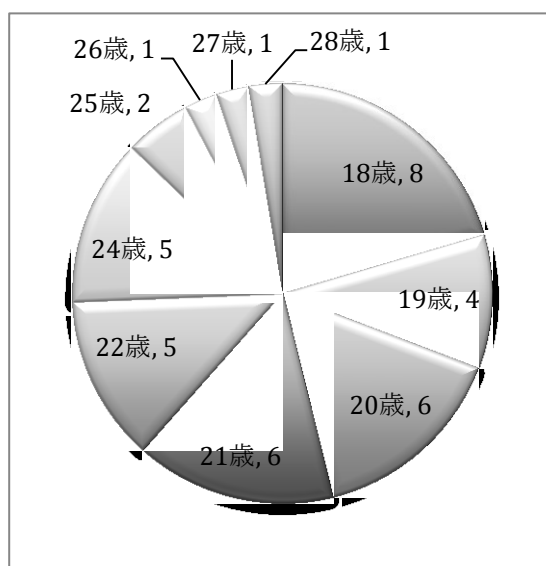


図 8 タスクに基づくコーパス(母語話者)年齢の分布

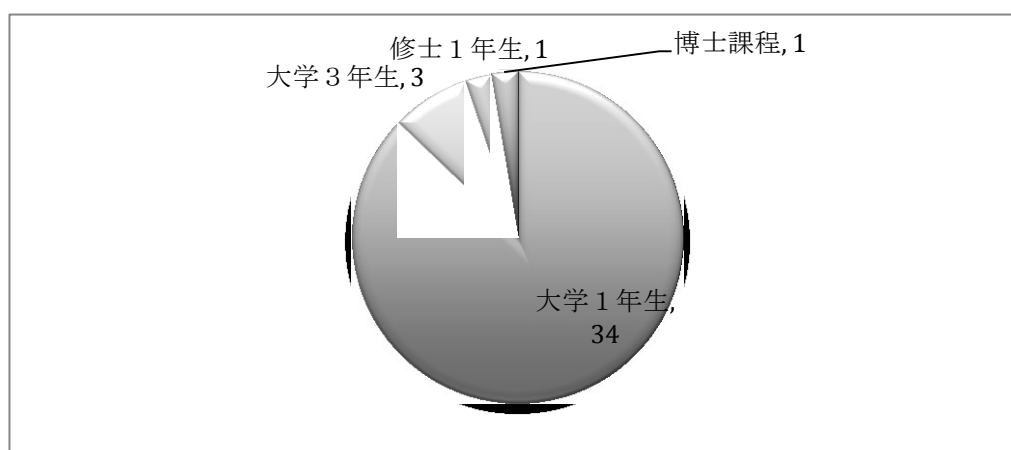


図 9 タスクに基づくコーパス(母語話者)学年の分布

また被験者が録音時点までで最も長く居住した地域は、おおよそボルドー周辺のアキテーヌ地方であるが、それ以外の地域の出身者も多い(図 10 参照)。

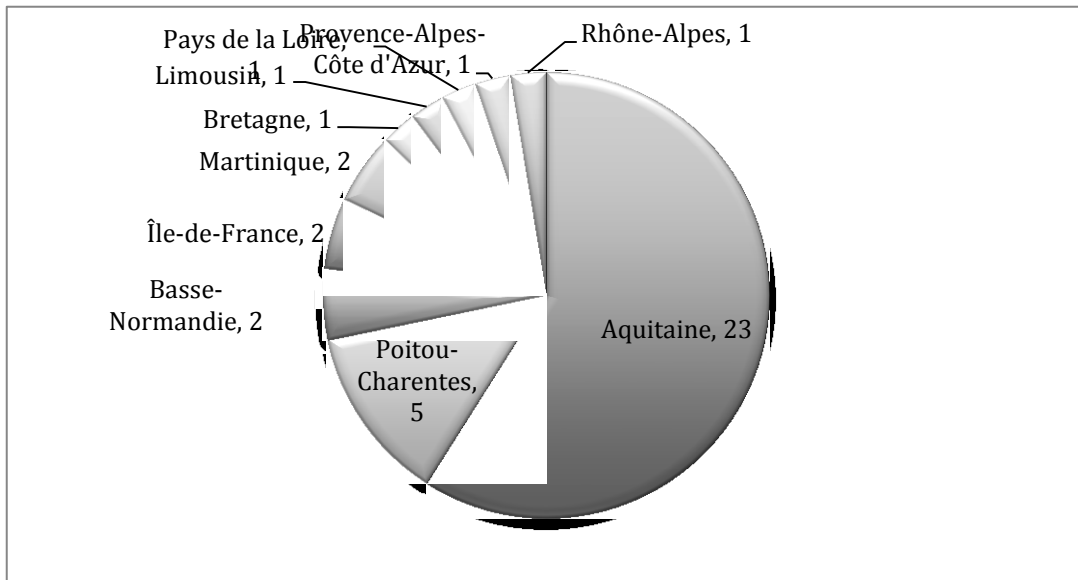


図 10 タスクに基づくコーパス(母語話者)出身地の分布

#### 4.2.2.2. タスク

学習者コーパスでは、モノログ形式とダイアログ形式のタスクを行ったが、フランス語母語話者コーパスでは、モノログ形式のタスクで行った。モノログ形式のタスクを行う際はフランス語で指示を読んでフランス語で回答した。以下は被験者が実際に行ったタスクのリストである。

表 6 タスクのリスト

N°	Questions
f001	Quelle est votre saison préférée? Pourquoi?
f002	Où aimez-vous passer vos vacances?
f003	Vous n'avez pas votre téléphone portable aujourd'hui. Expliquez à un(e) ami(e) que vous l'avez oublié et demandez-lui de vous prêter le sien.
f004	Un(e) ami(e) veut emprunter votre portable. Que diriez-vous quand vous me le prêtez ?
f005	Un(e) ami(e) veut emprunter votre portable, mais vous n'aimez pas le prêter. Que diriez-vous pour refuser sa demande?
f006	Vous voulez proposer à un(e) ami(e) de dîner ce soir au restaurant. Que lui diriez-vous?
f007	Un(e) ami(e) vous propose d'aller voir un film. Vous acceptez et vous lui donnez rendez-vous.
f008	Expliquez comment aller à l'université de chez vous et dites combien de temps il faut pour arriver. Indiquez les étapes.
f009	Narita → Nippori → Ueno La ligne Keisei La ligne Yamanote (1h10) (5min) On vous demande comment aller de Narita (aéroport) à Ueno. Expliquez comment y aller et combien de temps il faut.

f010	Quel genre de film regardez-vous?									
f011	Quel genre de livres lisez-vous?									
f012	Qu'est-ce que vous avez mangé ce matin?									
f013	Qu'est-ce que vous buvez chez vous?									
f014	Quelle est votre nationalité?									
f015	Où êtes-vous né(e)?									
f016	Quelle est votre profession?									
f017	Vous êtes dans la rue, et demandez à un passant où se trouve le magasin "bio Natura".									
f018	Au restaurant, vous voyez un plat que vous ne connaissez pas. Comment vous informeriez-vous auprès du serveur ?									
f019	Qu'est-ce que vous aimez faire le week-end ?									
f020	Qu'est-ce que vous aimez faire à la maison?									
f021	Qu'est-ce que vous voudriez faire ce week-end, rester à la maison ou sortir?									
f022	Vous commandez un café au café. Que dites-vous au garçon?									
f023	Vous êtes au café. Demandez à votre ami, Fabien, ce qu'il veut commander.									
f024	Présentez-vous. (nom, prénom, nationalité, âge, profession, adresse, situation de famille, etc)									
f025	Prénom: Kazuko Nom : Tanaka Nationalité: Japonaise Âge: 38ans Profession: Professeur Situation de famille: Mariée – 2 enfants Adresse au Japon: 3 Asahi-cho Fuchu Adresse en France: 67, rue de la Liberté 69000 Lyon Observez la fiche d'identité ci-dessous et présentez cette personne.									
f026	Parlez de votre meilleur(e) ami(e).									
f027	Indiquez ce que vous aimez ou ce que vous n'aimez pas faire le week-end.									
f028	Qu'est-ce que vous préférez, la ville ou la campagne ? Pourquoi?									
f029	Qu'est-ce que vous n'aimez pas faire?									
f030	Racontez votre dernier voyage.									
f031	Racontez votre dernière fin de semaine.									
f032	Vous téléphonez à un(e) ami(e) pour savoir s'il (elle) peut sortir ce week-end. Cependant, il (elle) ne répond pas au téléphone. Laissez un message en indiquant quelle activité vous voulez proposer, où, quand. Et invitez-le (la).									
f033	Un(e) ami(e) vous propose d'aller voir un match de foot. Vous ne voulez pas y aller, parce que vous n'aimez pas le foot. Refusez sa proposition et trouvez des excuses.									
f034	<table border="1"> <thead> <tr> <th>Gares de départ et d'arrivée</th> <th>Horaires</th> <th>Trains</th> </tr> </thead> <tbody> <tr> <td>PARIS</td> <td>14h40</td> <td>TGV</td> </tr> <tr> <td>BORDEAUX</td> <td>17h39</td> <td></td> </tr> </tbody> </table> Vous voulez réserver un billet de Paris à Bordeaux pour une personne.	Gares de départ et d'arrivée	Horaires	Trains	PARIS	14h40	TGV	BORDEAUX	17h39	
Gares de départ et d'arrivée	Horaires	Trains								
PARIS	14h40	TGV								
BORDEAUX	17h39									



	Gares de départ et d'arrivée	Horaires	Trains
	PARIS	14h40	TGV
	BORDEAUX	17h39	
	BORDEAUX	18h04	TGV
	TOULOUSE	20h12	
f035	A l'aide du tableau, expliquez comment il faut faire pour aller de Paris à Toulouse en indiquant l'heure de départ et l'heure d'arrivée.		
f036	Qu'est-ce que vous regardez à la télé en général? Et quelle est votre opinion sur les programmes que vous regardez?		
f037	Qu'est-ce que vous prenez d'habitude au petit déjeuner?		
f038	En général, qu'est-ce que vous buvez quand vous allez au restaurant?		
f039	Vous vivez dans la région de Bordeaux depuis longtemps?		
f040	Pouvez-vous me parler de l'endroit où vous habitez?		
f041	Qu'est-ce que vous allez faire aujourd'hui?		
f042	Vous faites souvent du sport?		
f043	Quel jour de la semaine préférez-vous ? Pourquoi ?		
f044	Quel est votre plat préféré ? Pourquoi ?		
f045	Est-ce que vous avez une voiture?		
f046	Avez-vous un chien?		
f047	Vous êtes au kiosque à Paris. Vous voulez acheter "l'officiel des spectacles". Que dites-vous ?		
f048	Vous êtes au kiosque à Paris. Vous voulez acheter "l'officiel des spectacles", mais vous ne savez pas le prix. Demandez le prix.		
f049	Vous allez passer un examen très important. Exprimez votre sentiment.		
f050	Votre meilleur(e) ami(e) va à l'étranger pour un an et vous ne pourrez pas le voir. Exprimez votre sentiment.		
f051	Votre ami(e) a réussi à son bac. Qu'est-ce que vous lui diriez?		
f052	Votre ami(e) va se marier. Qu'est-ce que vous lui diriez?		

#### 4.2.2.3. 録音システム

被験者は質問用紙を手渡され、IC レコーダーを使用して録音を行った。録音の直前に質問用紙が手渡されたため、学習者コーパスの場合と異なり、母語話者はほとんど事前に準備する時間は与えられていない。もちろん、辞書などを使用することは一切禁止されており、録音前にメモをとることも禁じられている。

#### 4.2.2.4. 転写

転写は筆者が行い、フランス語母語話者が確認を行った。転写に際し、以下の点を一貫させた：

- 1) 事前準備時間が与えられなかったため、回答の前に多くのためらい語が見られた。回答前のためらい語はいっさい転写を行わなかった。

- 2) ためらい語<sup>42</sup>の転写を一貫させるため、全て HESIT に置き換えた。
- 3) 固有名詞は全て NP に置き換えた。
- 4) 間投詞は全て INT に置き換えた。
- 5) 語を途中で言いかけて言い直した場合は転写を行わない。
- 6) 主に英語由来の外国語起源の単語は、そのまま転写を行うが、日本語については FW を用いた。

#### 4.2.3. 自由会話コーパス(IPFC-JP コーパス)

##### 4.2.3.1. 対象者

IPFC-JP コーパスは、日本語を母語とするフランス語学習者のコーパスである。このコーパスは現在までに 100 名近い学習者のデータを含んでおり、学習者の CEFR レベルも様々である。このコーパスは、単語リストの反復、単語リストの音読、テキストの音読、方向付けられたインタビュー、そして自由会話が含まれている<sup>43</sup>。しかし、本研究では自由会話のみをデータの対象とする。自由会話はペアで行われた。転写が完了しているフランス語学習者 38 名を対象とする。インフォーマントの多くは録音当時に学部生であったが、修士課程の学生と博士課程の学生も含まれる。

##### 4.2.3.2. 会話の主題

会話の主題として以下の 8 つが被験者に与えられる。

- 1) 観た映画
- 2) 読んだ本
- 3) 訪れた場所(国、町、美術館、など)
- 4) 日本のこと(料理、習慣、生活様式、祭)
- 5) 日本と他の国の文化的な違い
- 6) 大切な思い出  
(子供の頃、家族、学校、旅行、エピソード、はじめてのフランス語の授業)
- 7) 現在の話題
- 8) 新聞などの三面記事

もし、ペア通しで同じ主題を選んでしまった場合は、どちらかが別のテーマを選択する。会話を活発に行うために、日本語を使うことは許可されているが、辞書を使用することは禁止されている。会話時間はペアによって異なっていた。また、このテーマが起点となって、このテーマと異なる話題へシフトするケースも多く見られた。

<sup>42</sup> bah beh ben boh eh euh など

<sup>43</sup> 詳しくは第一章 1.4.6.4.を参照のこと

#### 4.2.3.3. 録音システム

録音には IC レコーダーを使用した。できるだけ第三者のいない騒音の少ない場所で録音を行うよう指示が与えられ、会話は主に教室内で録音された。

#### 4.2.3.4. 転写

転写は東京外国語大学の大学院生と、転写の専門業者が行った。転写規則に従って転写を行ったものの以下の点について個人差が見られた。また、これらの転写方法の違いがもたらすものは本研究の対象外と考えた。よって、本研究に適応するため以下の点を一貫させた：

- 1) ためらい語<sup>44</sup>の転写を一貫させるため、全て HESIT に置き換えた。
- 2) 相づち<sup>45</sup>は全て BC に置き換えた。
- 3) 固有名詞は全て NP に置き換えた。
- 4) 間投詞は全て INT に置き換えた。
- 5) 語を途中で言いかけて言い直した場合は転写を行わない。
- 6) 外国語は FW に置き換えた。

#### 4.2.4. 自由会話コーパス(フランス語(エックス)多言語話しことばフランス語コーパス)

##### 4.2.4.1. 対象者

本コーパスは、東京外国語大学大学院の 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もので、フランスのエックス・アン・プロバンスにあるエックス・マルセイユ大学との共同研究の一環で 2005 年度より録音が断続的に行われている。音声的、統語意味的、語用論的分析を行うことを目的に作成されたコーパスである。本コーパスには、大学生や大学教員、書籍店や旅行代理店の店員等の会話が含まれているが、本研究では大学生の友人間の会話のみを扱う。2009 年に録音された大学生の会話コーパスデータから、無作為抽出した 20 名分のデータを参照コーパスとして使用する。

##### 4.2.4.2. 会話の主題

学習者による自由会話コーパスとは異なり、録音時に会話のテーマは特に与えられず、自由なテーマで会話が行われた。話題の大半が大学生活についてであったが、会話の中でテーマは一つに固定されず、様々なテーマで会話が行われていた。

---

<sup>44</sup> bah beh ben boh eh euh など

<sup>45</sup> mmm hum hm など

#### 4.2.4.3. 録音システム

会話の録音には IC レコーダーが使用された。被験者は、第三者のいない静かな場所で録音を行うように指示が与えられた。会話の多くは教室内で録音されたが、屋外で録音した場合もある。

#### 4.2.4.4. 転写

転写はエックス・マルセイユ大学が行った。転写はエックス・マルセイユ大学の規則に則って行われたが、本研究に適応するため以下の点を一貫させた：

- 1) ためらい語<sup>46</sup>の転写を一貫させるため、全て HESIT に置き換えた。
- 2) 相づち<sup>47</sup>は全て BC に置き換えた。
- 3) 固有名詞は全て NP に置き換えた。
- 4) 間投詞は全て INT に置き換えた。
- 5) 語を途中で言いかけて言い直した場合は転写を行わない。
- 6) 主に英語由来の外国語起源の単語はそのまま転写を行うが、日本語の場合は FW に置き換えた。

### 4.3. 使用語彙分析の構造

本論文では、Word Units (WUs)の研究と Multi-Word Units (MWUs)研究の2つを行う。つまり、語レベルである WUs と語より大きな単位である MWUs の2つの観点から、日本人フランス語学習者の使用語彙についての記述を行う。いずれの場合も、学習者コーパスと母語話者コーパスとの比較を行いながら、学習者の中間言語分析を行う。

WUs レベルの研究として、まず語彙の豊かさの分析を行う。*VocabProfil* を使用して、語彙の洗練性、語彙の多様性、語彙密度という3つの指標をもとに、学習者と母語話者の語彙の豊かさを統計的に比較する。次に、別の角度からによる WUs レベルの分析として、特徴語分析を行う。ここでの特徴語とは、過剰使用語と過少使用後のことを指す。次に、MWUs レベルの研究として、コロケーション分析を行う。これは MUs 分析で抽出された特徴語を中心語として、どのような環境でこれらの特徴語が使用されるのかを分析するものである。最後にコーパス特徴的 N-grams を抽出して、分析を行う。

したがって、本論文における学習者の使用語彙研究の構造は以下のように示すことができるだろう。

---

<sup>46</sup> bah beh ben boh eh euh など

<sup>47</sup> mmm hum hm など

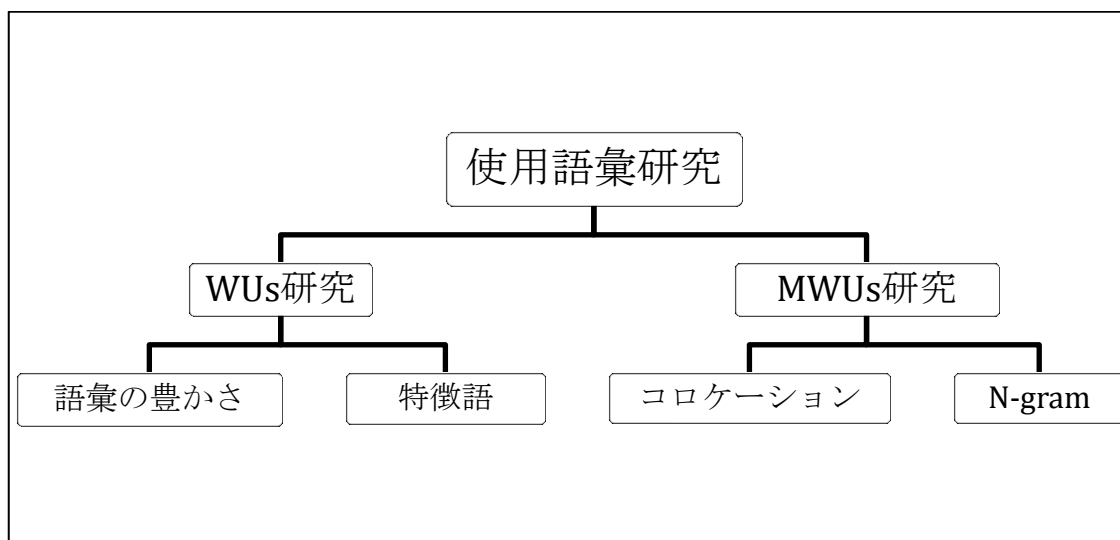


図 11 本論文における使用語彙研究の構成

そして、これらの4つの観点について「タスクに基づくコーパス」と「自由会話コーパス」に基づいて分析を行う。

既に述べた通り、タスクに基づくコーパスデータはフランス語専攻の二年生から得られたものであり、自由会話コーパスにおける録音の参加者は大学三年生から大学院生までと異なる。つまり、データ収集方法のみならず学習者の言語運用能力が異なる可能性があると言える。したがって、コーパス単位による使用語彙の違いには積極的に着目しない。しかし、データ収集方法や言語運用能力の違いがあるのにも関わらず、一貫して観察される学習者の特徴については扱うこととする。

#### 4.4. リサーチクエスチョン

##### 4.4.1. 語彙の豊かさ

先行研究から以下の前提に基づいて、分析を進める。

- 1) 高頻度語は、低頻度語と比較して習得が容易である
- 2) 高頻度語を産出することよりも低頻度語を産出することは困難である。

そして、語彙の豊かさの指標を使用して次のリサーチクエスチョンを検証していく。

- 1) 母語話者は学習者よりも低頻度語の語彙の豊かさの指標が高いか。
- 2) 言語能力の高い学習者は、母語話者の語彙の豊かさに近い指標を持つか(自由会話コーパス)。

##### 4.4.2. 特徴語

過剰使用語と過少使用語について統計的手法を用いて抽出し、質的な分析を行うことで、以下のリサーチクエスチョンを明らかにしていく：

- 1) 学習者の過剰使用語にはどのような特徴があるのか。
- 2) 学習者の過少使用語にはどのような特徴があるのか。

- 3) 学習者の特徴語に、外国語運用能力の差が関係するのか(自由会話コーパス)。
- 4) タスクに基づくコーパスや自由会話に基づくコーパスなどのデータの差に関わらない一貫した特徴語が存在するのか。

過剰使用語と過少使用語の選定は、*WordSmith* を使用して対数尤度比( $G^2$ )を算出することによって求める。なお、どの程度の基準をもって過剰使用語、過少使用語とするかは判断によって異なるため、 $G^2 \geq 50$  の場合を過剰使用語とし、 $G^2 \leq -50$  の場合を過少使用語とした。この値は恣意的なものであるが、これを設定することで抽出された語がある一定以上の特徴を持つ語であるということが保証される。これらの特徴語について、頻度層との関係を分析する。また、特徴語を品詞別に分類し、特徴語の傾向を探る。

#### 4.4.3. コロケーション分析

コロケーション分析として、過剰使用語と過少使用語について前後のコンテキストに当てはめて分析を行う。コロケーション分析のリサーチクエスションは以下の通りである：

- 1) 学習者の過剰使用語を含むコロケーションにはどのような特徴があるか。
- 2) 学習者の過少使用語を含むコロケーションにはどのような特徴があるか。
- 3) 学習者のコロケーションの差に、外国語運用能力の差が関係するのか(自由会話コーパス)。
- 4) タスクに基づくコーパスや自由会話に基づくコーパスなどのデータの差に関わらない一貫したコロケーションが存在するのか。

#### 4.4.4. N-grams 分析

N-grams 分析では *WordSmith tools* を使用して、2-5 語連鎖を両コーパスから自動的に取り出す。また、母語話者と学習者の N-grams を比較し、過剰使用 N-grams、過少使用 N-grams といった特徴的 N-grams を抽出する。そして、以下のリサーチクエスションについて検証を行う

- 1) 学習者は母語話者と比較して、一部の N-grams を繰り返し使用するか。
- 2) 特徴的 N-grams はどのような頻度層の語で構成されているか。
- 3) 特徴的 N-grams と特徴語にはどのような関係があるのか。
- 4) 特徴的 N-grams にはどのような傾向があるのか。
- 5) タスクに基づくコーパスや自由会話に基づくコーパスなどのデータの差に関わらない一貫した N-gram の特徴が存在するのか。

これらのリサーチクエスションについて、タスクに基づくコーパスおよび自由会話コーパスを用いて答えを明らかにしていく。

## 4.5. 分析方法

### 4.5.1. 使用語彙研究の分析方法

上記のとおり、タスクに基づくコーパスも自由会話コーパスも、一貫した分析方法を用いる。語彙の豊かさの分析、特徴語分析、コロケーション分析、N-gram 分析の方法について、順に説明する。

#### 4.5.1.1. 語彙の豊かさの分析

学習者と母語話者との比較と学習者間の比較を行うにあたり、内容語と機能語が関係する「語彙密度」、頻度が関係する「語彙の洗練性」、そしてタイプとトークンが関係する「語彙の多様性」の3つの観点を変数に使用する。

「語彙密度」と「語彙の洗練性」に関しては、高頻度内容語、高頻度機能語、中頻度語、低頻度語、リスト外単語の5つの変数を採用した。頻度層はそれぞれ *VocabProfil* に依拠しており、高頻度内容語は *VocabProfil* の K1 (上位 1000 語) に含まれる内容語、高頻度機能語は K1 に含まれる機能語、中頻度語は K2 (1001~2000 位)、低頻度語は K3 (2001~3000 位)、リスト外単語は Off-list (*VocabProfil* に含まれない単語) に相当する。なお、ためらい語である HESIT、固有名詞である NP、外国語起源の単語である FW、間投詞である INT は、*VocabProfil* では Off-list に分類されることが多く、本研究の趣旨と異なる結果が予想されるため、語彙の豊かさの分析にはこれらの語を除くこととする。

これらの頻度層における「語彙の多様性」を算出するため、ギロー値を採用した。以下の方法で各インフォーマントの変数を算出した。

$$\text{高頻度内容語ギロー値} = \text{高頻度内容語タイプ数} / \sqrt{\text{高頻度内容語トークン数}}$$

$$\text{高頻度機能語ギロー値} = \text{高頻度機能語タイプ数} / \sqrt{\text{高頻度機能語トークン数}}$$

$$\text{K2 ギロー値} = \text{K2 タイプ数} / \sqrt{\text{K2 トークン数}}$$

$$\text{K3 ギロー値} = \text{K3 タイプ数} / \sqrt{\text{K3 トークン数}}$$

$$\text{Off-list ギロー値} = \text{Off-list タイプ数} / \sqrt{\text{Off-list トークン数}}$$

分析にはまず、各頻度層におけるトークン数、タイプ数、比率、ギロー値について記述統計を行う。比率とギロー値については、学習者グループと母語話者グループの有意差を調べるため、t 検定を行う。次に、それぞれのコーパスについて、ギロー値と頻度層を変数にとり、相関係数の比較を行う。さらに、主成分分析という多変量解析を利用して語彙の豊かさの総合指標を求める。そして、主成分得点をもとにして被験者間の順位づけを行い、学習者と母語話者の比較と、学習者間の比較を行う。

#### 4.5.1.2. 特徴語の分析

特徴語の分析は、語彙の豊かさの分析の質的研究の側面を持つ。まず、*WordSmith* (ver. 6.0) の *Keyword List* の機能を用いて、学習者の過剰使用語と過少使用語を算出し、これを特徴語とする。これらの特徴語を明らかにすることで、学習者コーパスに見られる使用語彙の特徴をより詳細に分析することが可能となる。なお統計手法として、対数尤度比カイ二乗統計量(log-likelihood ratio または  $G^2$ とも呼ばれる)を選択し、これを *Keyness* の値とする。対数尤度比は、実測値と期待値の差に注目するものである。特徴語を算出の際には、母語話者コーパスを参照コーパスと位置づける。正の値を示したものを過剰使用語とし、負の値を示したものを過少使用語とする。閾値の設定については、分析時に説明する。

次に、過剰使用語と過少使用語について俯瞰するため、特徴語について学習者コーパスと母語話者コーパスの実測値や *Keyness* をまとめる。また、学習者コーパスと母語話者コーパスの規模が異なるため、10万語あたりの頻度に調整したものも示すこととする。

さらに、以上で得られたリストを、『語彙の豊かさ』の分析で使用した *VocabProfil* の頻度層と当てはめる。このことによって、過剰使用語と過少使用語がどの頻度層で現れるのかを視覚化することができる。また、『語彙の豊かさ』では明らかにできなかった質的な側面が明らかにできると考える。

最後に、*VocabProfil* の頻度層ごとにどのような品詞が含まれているのかを分析する。このことによって、過剰使用語や過少使用語といった特徴語の傾向を異なる視点から把握することができると期待できる。

#### 4.5.1.3. コロケーション分析

コロケーション分析では、『*Word Units* 研究』であらかじめ明らかになった特徴語について、それらの共起語を分析する。つまり、学習者の過剰使用語と過少使用語を含む *MWUs* について、母語話者の使用とどのように使用コンテキストが異なるのかを分析する。過剰使用語や過少使用語に結びつきの強い語や *MWUs* は何かについて分析を行う。また、コロケーション分析では連続した語のみに限定せず、非連続的なコロケーションについても興味の対象とする。

まず、全ての特徴語について、*WordSmith tools* の *concordance* 機能を使用し、学習者コーパスと母語話者コーパスのそれぞれに対して特徴語の前後 3 語についてリスト化し、特徴語の共起語について比較を行う。このことによって、特徴語を中心とした最大 7 語連鎖までを観察することが可能となる。その際、調整頻度 10 万語あたり 20 語以上生じたものを分析対象とする。

タスクに基づく分析については、どのようなタスクのもとで以上の *MWUs* が現れるのかを観察する。自由会話コーパスについては、学習者と母語話者間だけでなく、学習者間のレベル別での特徴があるのかについても分析する。



#### 4.5.1.4. N-gram 分析

N-gram 分析では、連続的な MWUs を *WordSmith tools* を使用して機械的・自動的に抽出する。まず、学習者は繰り返し使用される MWUs を持っているのかどうかを調べるため、N-grams のトークンとタイプについて 10 万 N-grams あたりの調整頻度に直し、母語話者コーパスと比較する。なお、調整頻度は次のように計算を行うこととする：

- 1) 学習者コーパスにおける N-gram の総数を *WordSmith* で求める：
- 2) 母語話者コーパスにおける N-gram の総数を *WordSmith* で求める：
- 3) 10 万 N-grams あたりの調整頻度を求めるため、各コーパスのそれぞれの N-grams の生起頻度を総数で割って 10 万を掛ける。

2 つのコーパスの比較にあたって、調整頻度の閾値を 100、50 以上 100 未満、20 以上 50 未満、20 未満と設定し、各頻度層に含まれる N-grams 数の割合を比較する。

次に、頻度最上位の N-grams の比較を行うため、各コーパスの 10 位までの調整頻度を比較する。

さらに、質的な分析を行うため、*WordSmith* を利用して特徴的 N-grams を抽出する。そして *VocabProfil* の頻度情報を基に、特徴的 N-gram を構成する語の頻度層を分析する。次に、これらの特徴的 N-grams と特徴語の関係について分析する。最後に、抽出された特徴的 N-grams を目録化し、特徴的 N-grams の傾向を探る。2-gram から 5-gram までに重複する特徴的 N-gram が出てきた場合は、語連鎖数が高いものを基に語連鎖数が低いものを下位分類する。

次章より、学習者の話し言葉に置ける使用語彙の特徴を、コーパスに基づいて分析を行う。

## 第五章 語彙の豊かさの分析

本章では、語彙の豊かさという観点から、学習者の話し言葉における使用語彙について明らかにしていく。まず、K1 機能語(頻度 1~1000 位までの機能語)、K1 内容語(頻度 1~1000 位までの内容語)、K2(頻度 1001 位~2000 位)、K3(頻度 2001 位~3000 位)、そして Off-list (リスト外単語: 頻度 3000 位~) について、トークン数とタイプ数を求める。次に、各頻度層のトークン数とタイプ数の比率の平均について学習者と母語話者を比較する。そして、各頻度層のギロー値の平均についても、学習者と母語話者を比較する。ギロー値と頻度層の関係について相関関係があるか調べた後、主成分分析を行って被験者の使用語彙を得点化し、学習者グループ、母語話者グループの傾向を探る。語彙の豊かさの分析は、ためらい語である HESIT、固有名詞である NP、外国語起源の単語である FW、間投詞である INT を除いて行う。

### 5.1. タスクに基づくコーパス

#### 5.1.1. 記述統計量

学習者コーパスのトークン総数は 27044 語で、タイプ総数は 1451 語であった。また、母語話者コーパスのトークン総数は 48982 語で、タイプ総数は 2451 語であった。各頻度層のトークン数とタイプ数は以下のようになった(表 7)。

表 7 学習者コーパスと母語話者コーパスのトークン数とタイプ数

	K1 機能語	K1 内容語	K2	K3	OFF	合計
学習者 トークン	15644	9013	1216	302	869	27044
学習者 タイプ	123	776	194	69	289	1451
母語話者 トークン	28622	15705	2056	484	2115	48982
母語話者 タイプ	150	1165	364	115	658	2452

母語話者と学習者は同じタスクに基づいて会話を行っているが、全頻度層のトークン数とタイプ数において、総じて学習者よりも母語話者の使用語数が多いことが分かる。特に Off-list の単語数に顕著な違いが見られる。母語話者コーパスのトークン総数に関しては学習者コーパスの約 1.81 倍、トークン総数については約 1.69 倍となった。

なお、各コーパスの各頻度層のタイプ数とトークン数の平均は以下のようになった(表 8)。

表 8 各コーパスのトークン数とタイプ数の平均

		K1 機能語	K1 内容語	K2	K3	OFF	合計
学習者	トークン	401.13	231.10	31.18	7.74	22.28	693.44
	タイプ	56.44	130.56	22.64	5.92	17.08	232.64
母語話者	トークン	733.90	402.69	52.72	12.41	54.23	1 255.95
	タイプ	72.92	200.23	35.79	9.56	40.21	358.72

次にトークンについて、t 検定を用いて学習者グループと母語話者グループの各頻度層の割合を比較する。比較の結果、学習者は K1 内容語が占める比率が有意に高く ( $t=2.95$ ,  $p<.001$ )、Off-list の比率は有意に低かった ( $t=-5.87$ ,  $p<.001$ )。しかし、その他の頻度層の比率では学習者と母語話者の間に有意差は見られなかった(表 9 参照)。これは、学習者が母語話者よりも高頻度語を多用し、母語話者が学習者よりも低頻度語を使用しているということをおも程度示す結果となったが、頻度 1000 語から 3000 語レベルにおいて有意差が見られないということより、疑問の残る結果となった。

表 9 トークンの頻度層ごとの比率の記述統計

頻度層	グループ	人数	平均値(%)	標準偏差	t	自由度	有意確率
K1 機能語	学習者	39	57.84	1.22			
	母語話者	39	58.08	2.02	-.65	62.37	.52
K1 内容語	学習者	39	33.30	1.26			
	母語話者	39	32.29	1.72	2.95	69.55	.00
K2	学習者	39	4.54	.72			
	母語話者	39	4.35	.76	1.12	76	.27
K3	学習者	39	1.11	.32			
	母語話者	39	1.00	.35	1.47	76	.15
Off-list	学習者	39	3.21	.59			
	母語話者	39	4.28	.96	-5.87	63.24	.00

学習者コーパスと母語話者コーパスのタイプの各頻度層の比率に関する記述統計は以下の通りになった(表 10)。

表 10 タイプの頻度層ごとの比率の記述統計

頻度層	グループ	人数	平均値(%)	標準偏差	t	自由度	有意確率
K1 機能語	学習者	39	24.47	2.08	6.29	76	.00
	母語話者	39	20.94	2.81			
K1 内容語	学習者	39	55.99	2.28	.42	76	.68
	母語話者	39	55.77	2.26			
K2	学習者	39	9.75	1.41	-.41	76	.69
	母語話者	39	9.88	1.32			
K3	学習者	39	2.53	.77	-.56	76	.58
	母語話者	39	2.63	.76			
Off-list	学習者	39	7.27	1.39	-7.47	58.27	.00
	母語話者	39	10.79	2.59			

その結果、K1 機能語の比率に関して、学習者が母語話者よりもその割合が有意に高く ( $t=6.39$ ,  $p<.001$ )、Off-list の比率は母語話者よりも学習者の方が有意に低かった ( $t=-7.47$ ,  $p<.001$ )。しかし、トークンの頻度別の比率で見られた K1 内容語の有意差はここでは見られず、またトークンに引き続き K2、K3 についても有意差が見られなかった。

ここまでタイプとトークンの頻度別の比率について有意差の検証を行ってきたが、有意差が存在したのは、以下の4点についてである：

- 1) K1 内容語のトークンの比率
- 2) K1 機能語のタイプの比率
- 3) Off-list のトークンの比率
- 4) Off-list のタイプの比率

このうち、1)と 2)については、学習者の比率が有意に高く、3)と 4)については母語話者の比率が有意に高かった。したがって、高頻度語である K1 が占める割合が学習者の方が高く、頻度 3000 位以上の単語である Off-list が占める割合が母語話者の方が高いという結果になり、リサーチクエスチョン 1) である「母語話者は学習者よりも低頻度語の語彙の豊かさの指標が高いか」という点について、肯定的な説明ができる。

しかし、トークン、タイプとも頻度 1000 位から 3000 位までの比率について、学習者と母語話者の有意差がいつい見られなかったことから、実際にグループの差を十分に説明できるとは考えにくい。なぜなら、フランス語学習歴が約 1 年である本被験者と母語話者との間に、中頻度語である K2 や低頻度語である K3 について、有意差が出ないとは考えにくいからである。そこで、各頻度層におけるギロー値を算出して学習者と母語話者との比較を行う。

学習者コーパスと母語話者コーパスのギロー値の記述統計は以下のようになった(表 11)。

表 11 ギロー値記述統計

頻度層	グループ	人数	平均値	標準偏差	t	自由度	有意確率
K1 機能語	学習者	39	2.84	.24	.71	76	.48
	母語話者	39	2.80	.29			
K1 内容語	学習者	39	8.59	.67	-8.59	76	.00
	母語話者	39	10.06	.84			
K2	学習者	39	4.05	.47	-5.85	63.33	.00
	母語話者	39	4.89	.76			
K3	学習者	39	2.10	.43	-5.28	76	.00
	母語話者	39	2.70	.56			
Off-list	学習者	39	3.60	.64	-8.19	58.58	.00
	母語話者	39	5.36	1.18			

その結果、K1 機能語には 2 グループ間で有意差が見られなかったが、その他 K1 内容語、K2、K3、Off-list について学習者が母語話者よりも有意に指標が低いことが明らかになった。したがって、ギロー値を用いた測定法がもっとも直感を反映した結果を表しているといえる。K1 機能語については、機能語という性質からギロー値の値が下がったと考えられる。つまり、機能語はいわゆる閉じられたクラスであり、数が限定されている。しかし文を生成するには機能語は必要不可欠であるため、発話をすればするだけ機能語のトークン数は上がるが、タイプ数は限定している。このような背景がこの結果をもたらしたものだと考えられる。

### 5.1.2. 相関関係

語彙の豊かさの指標間の関係を見るため、ギロー値と頻度層を変数としてピアソンの積率相関係数を求める。まず、学習者コーパスにおける結果は、表 12 の通りである。

表 12 学習者コーパスの各指標の相関行列

	K1 機能語	K1 内容語	K2	K3	Off-list
K1 機能語	1				
K1 内容語	.01	1			
K2	-.11	.37	1		
K3	-.38	.31	.11	1	
Off-list	.34	<b>.46</b>	.34	<b>.48</b>	1

学習者コーパスで相関関係が確認されたのは K1 内容語と Off-list の間と、K3 と Off-list の間だけであった。したがって、Off-list のギロー値が高い学習者は、頻度 1000 位以内に含まれる内容語と、2000 位から 3000 位のギロー値の値も高い。しかし、他の頻度

層のギロー値については、相関関係が見られなかった。母語話者の頻度層間の相関行列は以下ようになった。

表 13 母語話者コーパスの各指標の相関行

	K1 機能語	K1 内容語	K2	K3	Off-list
K1 機能語	1				
K1 内容語	-0.39	1			
K2	<b>-0.63</b>	<b>.63</b>	1		
K3	-0.26	<b>.48</b>	<b>.49</b>	1	
Off-list	<b>-0.68</b>	<b>.64</b>	<b>.77</b>	<b>.64</b>	1

母語話者コーパスでは、K1 機能語が他の頻度層に対して負の相関を示していることが分かる。特に K1 機能語と K2 そして Off-list の間で中程度の負の相関が見られた。これより、K1 機能語のギロー値の値が低ければ低いほど、他の頻度層のギロー値の値が高くなることを示している。

一方、K1 機能語以外の頻度層については、K2 と Off-list の間には高い正の相関が見られ、K1 内容語、K2、K3、Off-list については、それぞれ中程度の正の相関が見られた。これは、これら 1 つのギロー値が高ければ高いほど、その他のギロー値も高いということを示している。したがって、K1 機能語のギロー値のみが異なる性質を持つ指標であり、その他の指標は均質性が高いと考えられる。K1 機能語のギロー値が低く、その他の頻度層のギロー値が高いかどうかで、ある程度の学習者の言語能力を予測することは可能であろう。

### 5.1.3. 主成分分析

主成分分析を行い、語彙の豊かさの総合指数を求め、5 つの成分を抽出した (表 14)。

表 14 主成分分析結果

成分	固有値合計	寄与率%	累積 %
第 1 主成分	3.27	65.48	65.48
第 2 主成分	.84	16.89	82.37
第 3 主成分	.47	9.34	91.71
第 4 主成分	.25	4.98	96.69
第 5 主成分	.17	3.32	100.000

主成分分析では通常、固有値が 1.0 以上の主成分を解釈する。今回の分析では、固有値が 1.0 以上のものが第 1 主成分のみであるため、第 1 主成分のみを分析対象とする。

続いて主成分負荷量に注目する。これは、頻度層と第 1 主成分との相関を示している。第 1 主成分では、K1 機能語が負の値を示すのに対して、その他の頻度層に関しては、Off-list>K2>K1

内容語>K3 の順で高い正の値を返していることが分かる (表 15)。このことより、K1 機能語の第1主成分における特異性が追認できた。

表 15 主成分負荷量

	第 1 主成分
K1 機能語	-.51
K1 内容語	.86
K2	.87
K3	.81
Off-list	.93

次に、学習者と母語話者の第 1 主成分得点をまとめたものが表 16 である。なお、学習者の第 1 主成分得点の平均は-0.66 であり、母語話者の平均は 0.66 であった。つまり、第 1 主成分得点は、学習者よりも母語話者の方が高いことが分かる。

第 1 主成分得点を図にしたものが以下である (図 12)。図からも見て分かる通り、例外はあるものの総じて母語話者グループが上位に位置しているのに対して、学習者グループが下位を占めていることが分かる。

表 16 主成分得点

第1主成分得点		
順位	学習者	母語話者
1	0.44	2.58
2	0.43	1.96
3	0.42	1.94
4	0.10	1.90
5	-0.00	1.86
6	-0.11	1.64
7	-0.21	1.56
8	-0.27	1.46
9	-0.43	1.43
10	-0.44	1.32
11	-0.46	1.32
12	-0.47	1.22
13	-0.48	1.20
14	-0.50	1.19
15	-0.50	1.18
16	-0.54	1.15
17	-0.57	1.02
18	-0.59	0.86
19	-0.59	0.77
20	-0.75	0.73
21	-0.76	0.65
22	-0.79	0.63
23	-0.81	0.45
24	-0.90	0.44
25	-0.92	0.35
26	-0.93	0.19
27	-0.94	0.18
28	-0.99	0.16
29	-1.00	0.05
30	-1.02	-0.00
31	-1.02	-0.10
32	-1.06	-0.24
33	-1.12	-0.35
34	-1.14	-0.41
35	-1.15	-0.63
36	-1.37	-0.71
37	-1.40	-0.78
38	-1.45	-1.15
39	-1.61	-1.16



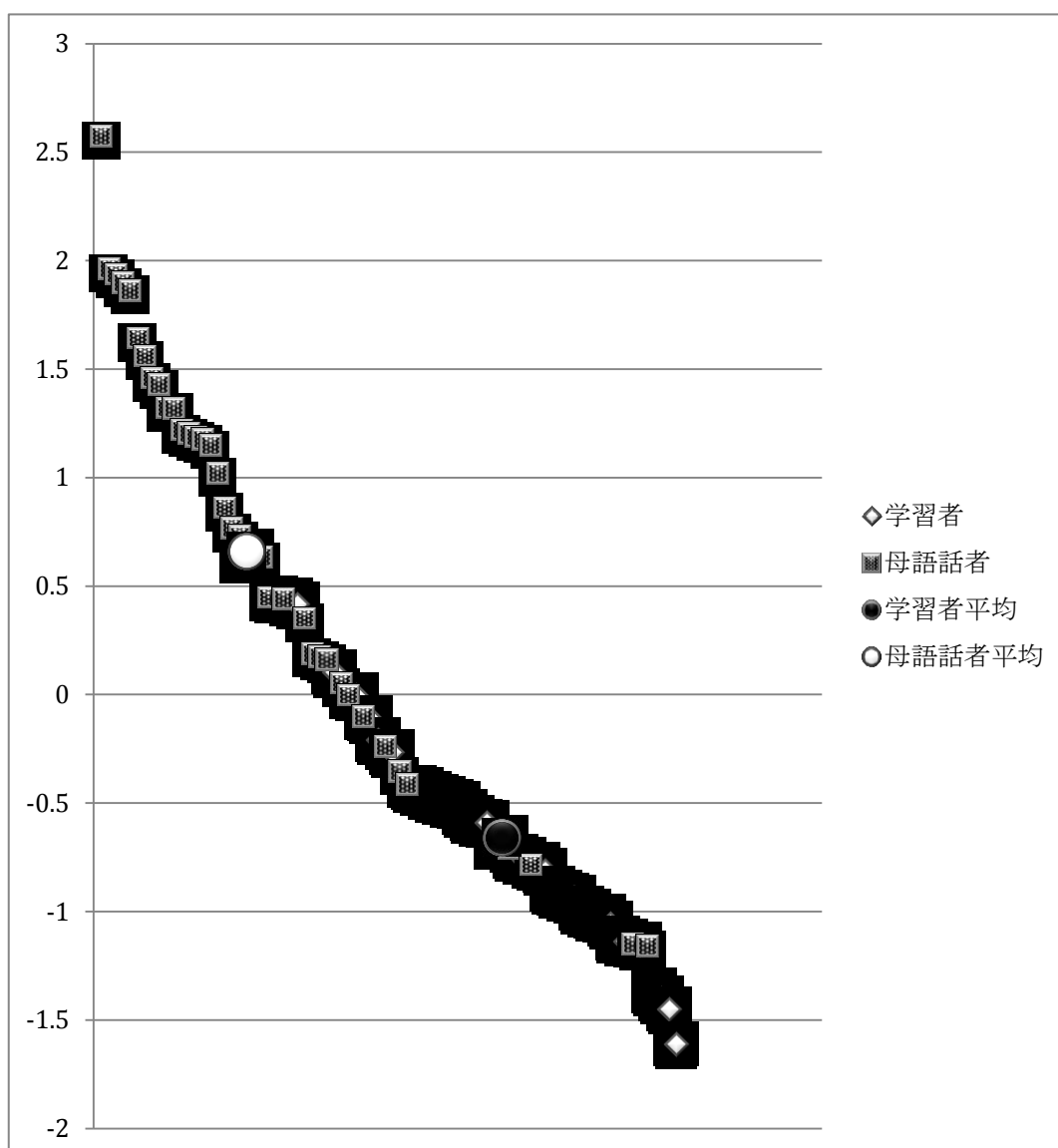


図 12 第 1 主成分得点

以上の結果を総合すると、いくつかの外れ値が見られるものの「語彙密度」、「語彙の洗練性」、「語彙の多様性」の3つの観点をもとにした語彙の豊かさの総合指標は、学習者の言語能力を予測するのに有効であると結論づけることができる。

## 5.2. 自由会話コーパス

### 5.2.1. 記述統計量

学習者コーパスのトークン総数は 30560 語で、タイプ総数は 2042 語であった。また、母語話者コーパスのトークン総数は 61545 語で、タイプ総数は 4458 語であった。各頻度層のトークン数とタイプ数は以下のようになった(表 17)。

表 17 学習者コーパスと母語話者コーパスのトークン数とタイプ数

		K1 機能語	K1 内容語	K2	K3	OFF	合計
学習者	トークン	18896	9135	1003	251	1275	30560
	タイプ	146	1036	194	104	431	2042
母語話者	トークン	37447	17826	1926	588	3758	61545
	タイプ	187	1807	751	289	1428	4458

母語話者コーパスは学習者コーパスの 2.01 倍の大きさである。各コーパスの各頻度層のタイプ数とトークン数の平均は以下ようになった(表 18)。

表 18 各コーパスのトークン数とタイプ数の平均

		K1 機能語	K1 内容語	K2	K3	OFF	合計
学習者	トークン	497.26	240.39	26.39	6.61	33.55	804.21
	タイプ	61.87	100.00	16.47	4.95	17.42	200.71
母語話者	トークン	1872.35	891.30	96.30	29.40	187.90	3077.25
	タイプ	104.85	305.70	65.65	21.70	100.45	598.35

次に、トークンについて母語話者と学習者の各頻度層の比率を比較した。t 検定を行ったところ、表 19 のような結果となった。

表 19 トークンの頻度層ごとの比率の記述統計

頻度層	グループ	人数	平均値(%)	標準偏差	t	自由度	有意確率
K1 機能語	学習者	38	62.39	4.22	2.20	55.30	.03
	母語話者	20	60.61	1.94			
K1 内容語	学習者	38	29.82	3.91	.73	55.70	.47
	母語話者	20	29.27	1.88			
K2	学習者	38	3.19	1.05	.23	56	.81
	母語話者	20	3.12	.85			
K3	学習者	38	.90	1.03	-1.10	56	.92
	母語話者	20	.93	.41			
Off-list	学習者	38	3.69	1.93	-4.28	56	.00
	母語話者	20	6.07	2.17			

学習者コーパスと母語話者コーパスの間に有意差が見られたのは、K1 機能語と Off-list のみであった。その他の頻度層では、2 グループ間に有意差が見られない。また t 値を見てみると、K1 機能語から K2 までは正の値を返しているが、K3 と Off-list については負の値を示している。つまり、K1 機能語に関してはこの頻度層を占める割合が学習者コーパスの方が有意に高いが、Off-list では母語話者の方が有意に高い。

学習者コーパスと母語話者コーパスのタイプの各頻度層の比率に関する記述統計は以下の通りになった(表 20)。

表 20 タイプの頻度層ごとの比率の記述統計

頻度層	グループ	人数	平均値(%)	標準偏差	t	自由度	有意確率
K1 機能語	学習者	38	30.30	8.56			
	母語話者	20	20.11	6.13	67.03	56	.00
K1 内容語	学習者	38	47.38	6.08			
	母語話者	20	51.00	2.35	-3.07	52.73	.03
K2	学習者	38	7.39	2.41			
	母語話者	20	10.27	3.00	-3.97	56	.00
K3	学習者	38	2.25	1.80			
	母語話者	20	3.34	1.45	-2.33	56	.02
Off-list	学習者	38	7.48	3.01			
	母語話者	20	15.28	4.57	-7.80	56	.00

その結果、全ての頻度層について 2 グループ間に有意差が見られた。t 値を見てみると、K1 機能語以外は全て負の値を示している。つまり、学習者コーパス に占める K1 機能語の割合は母語話者よりも有意に高い。一方、母語話者コーパスにおけるそれらの頻度層の割合が学習者コーパスよりも高いことを意味している。

ここまでタイプとトークンの頻度別の比率について有意差の検証を行ってきたが、有意差が存在したのは、以下の 7 点についてであった：

- 1) K1 機能語のトークンの比率
- 2) K1 機能語のタイプの比率
- 3) K1 内用語のタイプの比率
- 4) K2 のタイプの比率
- 5) K3 のタイプの比率
- 6) Off-list のトークンの比率
- 7) Off-list のタイプの比率

トークンについては、頻度 1000 位内に含まれる K1 機能語が占める割合が学習者の方が高く、頻度 3000 位以上の単語である Off-list が占める割合が母語話者の方が高いという結果になった。また、タイプについては K1 機能語が占める割合が学習者の方が高く、それ以外は母語話者の方が高かった。したがって、各頻度層の特にタイプの比率を調べることで、リサーチクエスチョン 1) である「母語話者は学習者よりも低頻度語の語彙の豊かさの指標が高いか」という点について、肯定的な結果を証明することができた。

次に、各頻度層におけるギロー値を算出して学習者と母語話者との比較を行う。

学習者コーパスと母語話者コーパスのギロー値の記述統計は以下のようになった(表 21)。

表 21 ギロー値記述統計

頻度層	グループ	人数	平均値	標準偏差	t	自由度	有意確率
K1 機能語	学習者	38	3.18	.53	3.04	56	.00
	母語話者	20	2.74	.51			
K1 内容語	学習者	38	6.48	1.62	-8.88	56	.00
	母語話者	20	10.38	1.59			
K2	学習者	38	3.08	1.12	-6.92	25.17	.00
	母語話者	20	6.47	2.03			
K3	学習者	38	1.71	1.00	-6.50	56	.00
	母語話者	20	3.79	1.42			
Off-list	学習者	38	2.98	1.18	-6.88	23.37	.00
	母語話者	20	7.12	2.55			

その結果、全ての頻度層で有意差が見られた。K1 機能語については、学習者のギロー値が母語話者よりも有意に高く、その他の頻度層では母語話者のギロー値が学習者よりも有意に高かった。つまり、頻度が最も高く、閉じられたクラスである機能語をのぞいて、母語話者が学習者よりも語彙が豊かであると言う結果を示した。したがって、ギロー値を用いた測定法も、母語話者は学習者よりも語彙の豊かさの指標が高いと言う結果を示している。また、比率の比較のみでは語彙の洗練性と語彙密度の観点だけにしか焦点が当てられておらず、語彙の豊かさを総合的に測定できているとは言えない。ギロー値の比較は、語彙の洗練性と語彙密度に加えて、語彙の多様性も測定できるため、本研究で使用した指標は語彙の豊かさを測る指標として有効であると言える。

### 5.2.2. 相関関係

語彙の豊かさの各指標間の関係を見るため、ギロー値を使用してピアソンの積率相関係数を求める。まず、学習者コーパスの相関行列は以下のようになった(表 22)：

表 22 各指標の相関行列

	K1 機能語	K1 内容語	K2	K3	Off-list
K1 機能語	1				
K1 内容語	-0.34	1			
K2	<b>-0.53</b>	<b>.79</b>	1		
K3	-0.37	<b>.55</b>	<b>.70</b>	1	
Off-list	<b>-0.42</b>	<b>.67</b>	<b>.83</b>	<b>.67</b>	1

表より K1 機能語は他の頻度層に対して全て負の相関を示していることが分かる。これより、K1 機能語のギロー値の値が高ければ高いほど、他の頻度層のギロー値の値が低くなることを示している。特に K1 機能語と K2、そして K1 機能語と Off-list の間で中程度の負の相関が見られた。

一方、K1 内容語、K2、K3、Off-list については、全て中程度から高い正の相関が見られた。これは、これら 1 つのギロー値が高ければ高いほど、その他のギロー値も高いということである。母語話者コーパスにおける相関関係を調べたところ、以下のようになった(表 23) :

表 23 各指標の相関行列

	K1 機能語	K1 内容語	K2	K3	Off-list
K1 機能語	1				
K1 内容語	<b>-.86</b>	1			
K2	<b>-.85</b>	<b>.85</b>	1		
K3	<b>-.80</b>	<b>.74</b>	<b>.87</b>	1	
Off-list	<b>-.84</b>	<b>.91</b>	<b>.79</b>	<b>.71</b>	1

母語話者コーパスでは、全ての頻度層間のギロー値について高い相関が見られた。K1 機能語は高い負の相関を示しており、K1 機能語の特異性が伺える。また、全体的に学習者コーパスの結果よりも、相関が強いことも分かる。つまり、母語話者の特徴として、K1 機能語のギロー値が低ければ、その他の頻度層のギロー値は高い傾向にあるということがいえる。

### 5.2.3. 主成分分析

主成分分析を行い、語彙の豊かさの総合指数を求め、5 つの成分を抽出した(表 24)。

表 24 主成分分析結果

成分	固有値合計	寄与率%	累積%
第 1 主成分	4.13	82.54	82.54
第 2 主成分	.50	9.08	92.35
第 3 主成分	.21	4.20	96.55
第 4 主成分	.11	2.13	98.68
第 5 主成分	.07	1.32	100.000

主成分分析では通常、固有値が 1.0 以上の主成分を解釈するため、第1主成分のみを分析対象とする。続いて主成分の因子負荷量の結果は以下ようになった。これは、頻度層と第 1 主成分との相関を示している。第 1 主成分では、K1 機能語が高い負の値を示し、その他も K3<K1 内容語<Off-list<K2 という順番で高い正の値を示している(表 25)。

表 25 主成分負荷量

	第 1 主成分
K1 機能語	-.77
K1 内容語	.93
K2	.97
K3	.92
Off-list	.95

次に、学習者と母語話者の第 1 主成分得点をまとめたものが表 26 である。なお、学習者の第 1 主成分得点の平均は-0.53 であり、母語話者の平均は 1.00 であった。第 1 主成分得点は学習者よりも母語話者の方がかなり高く、学習者グループの中で最も主成分得点が高かった学習者の主成分得点でも母語話者の平均に満たしていない。

第 1 主成分得点を図にしたものが以下である (図 13)。図からも見て分かる通り、例外はあるものの総じて母語話者グループが上位に位置しているのに対して、学習者グループが下位を占めていることが分かる。

表 26 主成分得点

第1主成分得点		
順位	学習者	母語話者
1	0.85	2.32
2	0.52	2.10
3	0.44	2.03
4	0.33	2.00
5	0.24	1.85
6	0.18	1.81
7	0.04	1.79
8	-0.01	1.42
9	-0.05	1.29
10	-0.11	1.08
11	-0.16	0.95
12	-0.30	0.68
13	-0.32	0.41
14	-0.32	0.40
15	-0.32	0.39
16	-0.44	0.29
17	-0.47	0.01
18	-0.66	-0.02
19	-0.72	-0.10
20	-0.72	-0.63
21	-0.74	
22	-0.86	
23	-0.87	
24	-0.88	
25	-0.93	
26	-0.95	
27	-0.96	
28	-0.97	
29	-0.99	
30	-1.01	
31	-1.01	
32	-1.03	
33	-1.05	
34	-1.07	
35	-1.12	
36	-1.15	
37	-1.20	
38	-1.24	

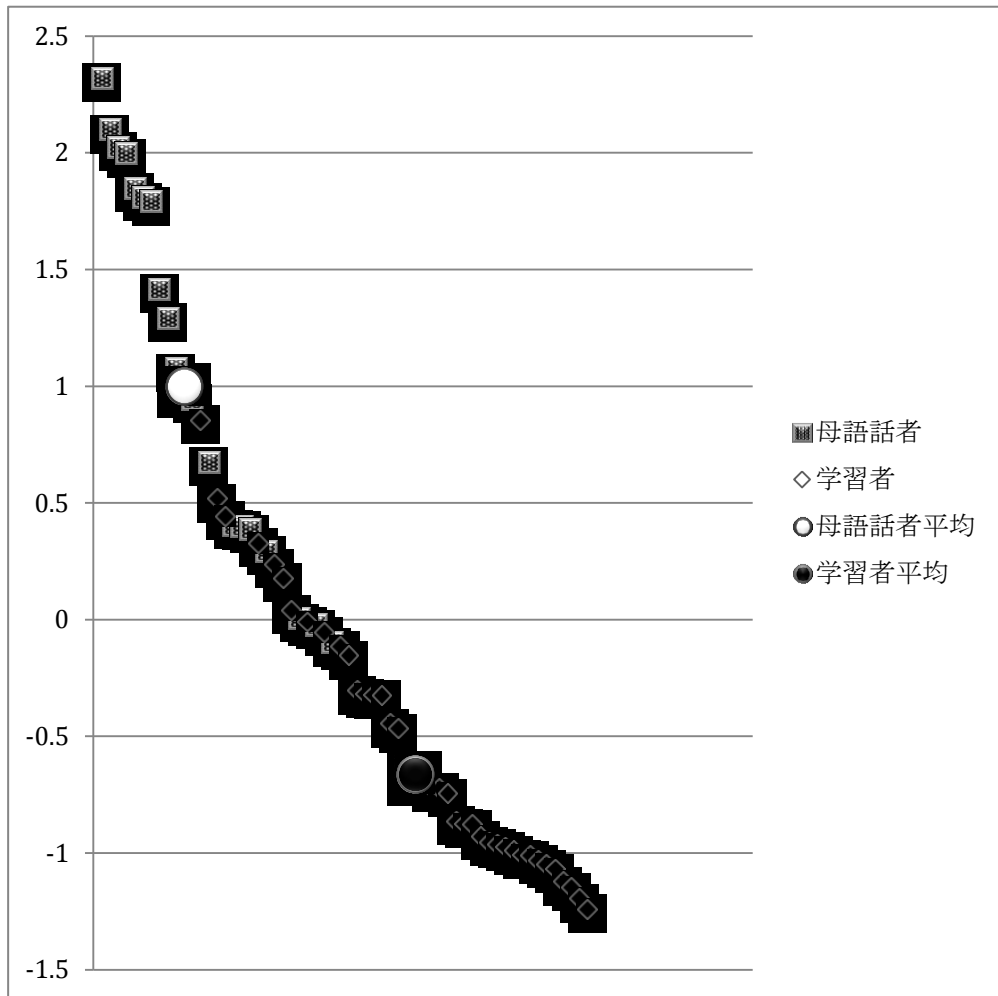


図 13 第 1 主成分得点

学習者間の差は大きく、いくつかのグループが存在していることが分かる。学習者の第 1 主成分得点のみを示しているのが以下の図である。



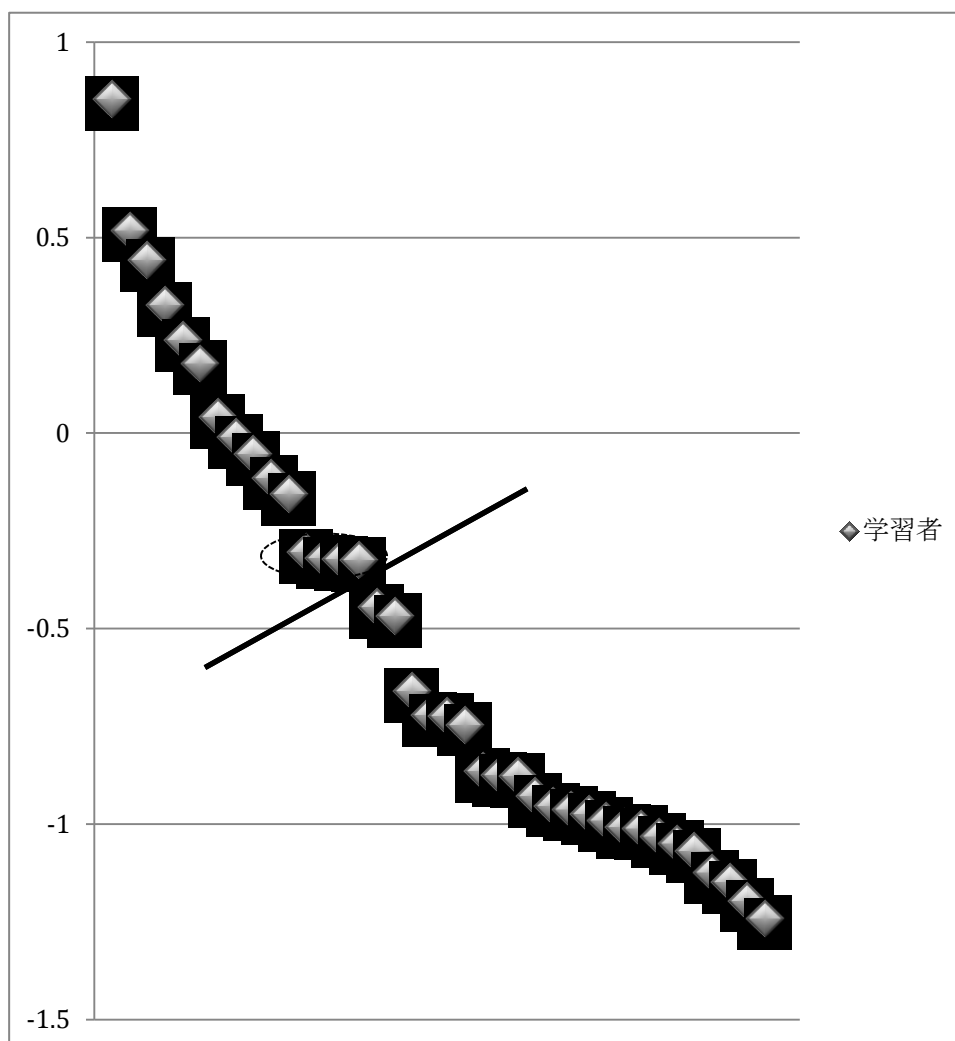


図 14 学習者の第 1 主成分得点

主成分得点が高い学習者のグループの間にも差が見られるが、得点の低い学習者は特に -1 前後で固まっていることが分かる。上位グループの中で得点が固まっているところが見られるため、そこを基準として 2 グループに分けた(図内点線を参照)。その結果、2 グループ間に質的な違いが見られることが分かった。まず、2 グループ間の大きな違いは 1 年以上の長期留学経験の有無である。特に上位 5 位までの 4 人は録音当時にフランス語圏で生活している学習者であった。上位グループの 8 割は長期留学の経験があるが、下位グループの中で長期留学経験者は 1 人しか見られなかった。次に、学年の差である。学年の差はフランス語学習歴と完全に一致するものではないが、ある程度の目安となる。修士課程以上の学生と学部生の人数を比較したところ、上位グループの 1 人を除いて修士課程以上の学習者であったのに対して、下位グループは 2 人以外を除いて学部生の学習者であった。

表 27 学習者のグループの特性

	人数	留学経験	修士以上
上位グループ	15	12 (80.00%)	14 (93.33%)
下位グループ	23	1 (4.35%)	2 (8.70%)

つまり、留学経験がある学習者で特にフランス語圏で生活している学習者の主成分得点が高く、また学習歴が長い学習者の得点も高いことが分かる。学習者よりも母語話者の主成分得点が高いことと、以上の結果を総合すると、リサーチクエスチョン 2)の「言語能力の高い学習者は、母語話者の語彙の豊かさに近い指標を持つか」について、肯定することができる。

### 5.3. まとめ

タスクに基づくコーパスでも、自由会話コーパスでも、ギロー値を用いた語彙の多様性の指標、内容語と機能語を区分する語彙密度、そして *VocabProfil* が与える頻度層情報を使用する語彙の洗練性を組み合わせた語彙の豊かさの指標が、学習者の言語能力を予測する際に有効であることが分かった。なぜなら、語彙の豊かさの指標は全体的に母語話者の方が学習者よりも高く、フランス語圏への留学経験がある学習者はない学習者よりも使用が高く、母語話者に近い指標を持っていたからである。

次章では、語彙レベルでの使用語彙分析を特徴語という別の観点から行う。

## 第六章 特徴語分析

本章では、学習者コーパスと母語話者コーパスを比較し、過剰使用語と過少使用語がどのような語であるのかについて分析する。つまり、特徴語という観点から、学習者の話し言葉における使用語彙を明らかにする。まず、過剰使用語と過少使用語を抽出する。次に、過剰使用語と過少使用語について *VocabProfil* における頻度層との関係を分析する。最後に、特徴語の性質をより詳しく分析するため、特徴語を品詞別にまとめる。

語彙の豊かさの分析では、タスクに基づくコーパスも自由会話コーパスもためらい語、相づち、固有名詞、間投詞、外国語も全て分析の対象外としたが、特徴語分析ではこれらの語を全て分析対象とする。各コーパスのトークン数とタイプ数は以下の通りである(表 28)。

表 28 トークン数とタイプ数

	タスクに基づくコーパス		自由会話コーパス	
	母語話者	学習者	母語話者	学習者
トークン	51823	28436	65591	34518
トークン平均	1328.80	729.13	3279.55	908.37
タイプ	2474	1488	4467	2064
タイプ平均	371.05	242.64	602.75	205.11

自由会話コーパスについては、語彙の豊かさの分析で分類することができた「初中級学習者」グループと「上級学習者」グループ間の使用語彙の差についても分析を行う。初中級学習者コーパスと上級学習者コーパスのトークン数とタイプ数については、以下の通りである。

表 29 トークン数とタイプ数の平均

	初中級学習者	上級学習者
トークン	9991	24527
トークン平均	434.39	1635.13
タイプ	910	1758
タイプ平均	128.30	322.87

### 6.1. タスクに基づくコーパス

過剰使用語と過少使用語を抽出するため、閾値を決定した。過剰使用語は  $G^2 \geq 50$  とし、過少使用は  $G^2 \leq -50$  とした。なお、過剰使用語に見られる NP は固有名詞を指し、過少使用語に見られる HESIT は、ためらい語 *uh, eh, bah, bin*などを指す。

表 30 過剰使用語と過少使用語

	過剰使用語	過少使用語
$100 \leq G^2$	VOUDRAIS	HESIT
	ELLE	ON
	NP	DONC
	NE	
	JE	
$50 \leq G^2 < 100$	PRENEZ	PEU
	JAPONAISE	BIEN
	VOUS	C'
	PEUX	VRAIMENT
	ABORD	TOUT
	FOOTBALL	
	UNIVERSITE	

次の節から、過剰使用語と過少使用語について、詳しい数値とともに、*VocabProfil* の頻度層との対応や品詞分類を行い、分析を進めていく。

#### 6.1.1. 過剰使用語

過剰使用語について、学習者と母語話者による実測値、10万語あたりの調整頻度、グループ内での使用割合を示すレンジ、そして *Keyness* は、以下の通りである(表 31)。

表 31 過剰使用語の頻度と *keyness*

順位		学習者			母語話者			<i>Keyness</i>
		実測値	調整頻度	レンジ (%)	実測値	調整頻度	レンジ (%)	
1	VOUDRAIS	212	745.53	100.00	59	113.85	79.49	208.30
2	ELLE	388	1364.47	100.00	264	509.43	94.87	157.37
3	NP	1546	5436.77	100.00	1872	3612.30	100.00	145.20
4	NE	181	636.52	100.00	74	142.79	71.79	133.62
5	JE	1589	5587.99	100.00	1995	3849.64	100.00	126.29
6	PRENEZ	63	221.55	74.36	7	13.51	12.82	91.43
7	JAPONAISE	93	327.05	97.44	40	77.19	56.41	65.45
8	VOUS	297	1044.45	100.00	276	532.58	100.00	64.71
9	PEUX	139	488.82	89.74	89	171.74	82.05	61.48
10	ABORD	46	161.77	69.27	7	13.51	15.38	60.25
11	FOOTBALL	35	123.08	56.41	2	3.86	5.13	58.85
12	UNIVERSITÉ	89	312.98	94.87	46	88.76	76.92	51.82

特にレンジの値に注目すると、興味深い点が2点あることが分かる。まず、学習者のレンジが100%の語があったが、母語話者による使用割合が7割台のものがある。これ

は、*voudrais* と *ne* であるが、全ての学習者が必ずこれらの単語を一度は使用していたのに対して、これらの単語を使用していなかった母語話者がいたことを意味している。また、*prenez, abord, football* は、学習者のレンジに対して母語話者のレンジが明らかに低い。これは、学習者のほとんどがこれらの単語を使用していたが、母語話者の中でこれらの単語を使用した人が少なかったということの意味している。

次に *VocabProfil* における頻度層に従い、各過剰使用語が K1 機能語、K1 内容語、K2、K3、Off-list のどれに属するのかを分析したところ、表 32 のようになった。なお、NP(固有名詞)はこの分析の対象外とした。

表 32 過剰使用語と VP の頻度層との関係

順位	VP	過剰使用語
1	K1 内容語	VOUDRAIS
2	K1 機能語	ELLE
3	NA	NP
4	K1 機能語	NE
5	K1 機能語	JE
6	K1 内容語	PRENEZ
7	K1 内容語	JAPONAISE
8	K1 機能語	VOUS
9	K1 内容語	PEUX
10	K1 内容語	ABORD
11	K1 内容語	FOOTBALL
12	K1 内容語	UNIVERSITE

以上の表より、学習者による過剰使用語はすべて K1 機能語ないし K1 内容語に属することが分かる。つまり、学習者による過剰使用語は、フランス語の頻度層第 1 位から 1000 位のものであると言える。

また、*VocabProfil* による頻度層と品詞の関係について、まとめたものが表 33 である。

表 33 *VocabProfil* の頻度層と品詞

VP	品詞	過剰使用語
	代名詞	ELLE
	副詞	NE
	代名詞	JE
K1 機能語	代名詞	VOUS
	動詞	VOUDRAIS
	動詞	PRENEZ
	形容詞	JAPONAISE
	動詞	PEUX
	名詞	ABORD
	名詞	FOOTBALL
K1 内容語	名詞	UNIVERSITE

K1 機能語についてはもともと品詞が限定されているが、その中でも人称代名詞主語が多いことが分かる(*elle, je, vous*)。このことは、語彙的要因による過剰使用であるのか統語的要因であるのかというような様々な要因が考えられる。この点については、共起関係や N-grams 分析などの MWUs レベルで分析する必要があるため、後の MWUs 分析で詳しく扱うこととする。

K1 内容語では、動詞が 3 語、名詞が 3 語、形容詞が 1 語となった。したがって、過剰使用語において、名詞と動詞が占める割合が高い。また、固有名詞を含めれば、名詞の占める割合はさらに高くなる。つまり、学習者の使用語彙は名詞に大きく依存していると言える。

#### 6.1.2. 過少使用語

過少使用語と判断された語における、学習者と母語話者による実測値、10 万語あたりの調整頻度、レンジ、Keyness は、以下の通りである(表 34)。

表 34 過少使用語の頻度と keyness

順位		学習者			母語話者			Keyness
		実測値	調整頻度	レンジ	実測値	調整頻度	レンジ	
1	HESIT	57	200.45	28.21	1199	2313.64	100.00	-711.58
2	ON	94	330.57	82.05	521	1005.35	100.00	-125.73
3	DONC	41	144.18	53.85	320	617.49	84.62	-109.91
4	PEU	8	28.13	17.95	163	314.53	92.31	-94.73
5	BIEN	56	196.93	76.92	313	603.98	92.31	-76.13
6	C'	237	833.45	97.44	795	1534.07	100.00	-76.04
7	VRAIMENT	6	21.10	12.82	105	202.61	69.23	-57.69
8	TOUT	14	49.23	30.77	135	260.50	76.92	-54.38

これより、ためらい語である HESIT の keyness が極めて高いことが分かる。学習者は母語話者よりもためらい語の使用が少ないのは、学習者が言葉に詰まった際の対処として、ためらい語を使用せず無言のポーズをとった可能性がある。しかし、このような無言のポーズはためらい語の機能を果たす一方で、話者のターンの終了や、発言権の移行も意味する。つまり、ポーズは誤った解釈を引き起こす可能性があり、ポーズを多用する学習者は間を持たせている間に、自身のターンを失う可能性がある (Gilquin 2008:141)。そこで、学習者は外国語で容認可能な表現を探している間を持たすための技術が必要である (Gilquin 2008: 121)。これまで、ためらい語は語彙教育で重要項目として考えられてこなかったが、談話面では極めて重要なものであると言える (McCarthy and Carter 1997: 28)。しかし、コーパス言語学の推進力と、その結果引き起こされた言語の真正性への興味の高まりのもと、ためらい語は教科書の中に採用されるようになってきた。ためらい語は目立たないことから、ためらい語に単に触れるだけでは十分に学習者の注意を喚起することができない。したがって、ためらい語を引き立て、適切なアクティビティを通して学習者の注意を引きつける必要がある (Gilquin 2008: 142-143)。ためらい語の過少使用のその他の可能性として、日本語のためらい語を使用していることも考えられる。日本語は FW と転写したこと、また学習者が語を発しなかった際にはコーパスには全く反映されないことより、これは憶測から脱することはできない。

次に、VocabProfil における頻度層に従い、過少使用語について K1 機能語、K1 内容語、K2、K3、Off-list のどれに属するのかを分析したところ、以下ようになった(表 35)。なお、HESIT(ためらい語)はこの分析の対象外とした。

表 35 過少使用語と VP の頻度層との関係

順位	VP	
1	NA	HESIT
2	K1 機能語	ON
3	K1 機能語	DONC
4	K1 内容語	PEU
5	K1 内容語	BIEN
6	K1 機能語	C'
7	K1 内容語	VRAIMENT
8	K1 機能語	TOUT

以上の表より、過少使用語においても全て K1 機能語と K1 内容語であることが分かる。つまり、学習者による過少使用語は、頻度層 1 位から 1000 位のものである。これは、フランス語において高頻度語である単語であるのにも関わらず、母語話者と比較して学習者はそれらの語彙を過小使用しているということの意味している。Ringbom(1998: 48)はこの点について、どの学習者言語でも母語話者と比べて、高頻度語を使用する頻度が一貫して低いということを指摘しており、この見解を支持する結果となった。もちろん、語によって大きなばらつきがあり、学習者グループの性質によって過剰使用もあれば過少使用もある点についても言及されており、本分析で過剰使用に高頻度語が多かった点についても、これは矛盾する現象ではない。

次に、*VocabProfil* による頻度層と品詞の関係についてまとめたものが表 36 である。

表 36 *VocabProfil* の頻度層と品詞

VP	品詞	
	代名詞	ON
	接続詞、副詞	DONC
	代名詞	C'
K1 機能語	形容詞、副詞、代名詞、名詞	TOUT
	副詞	PEU
	副詞、名詞、形容詞	BIEN
K1 内容語	副詞	VRAIMENT

*VocabProfil* の頻度層からは、過少使用語は高頻度語ばかりであるということが分かったが、品詞ごとに分類してみると明らかに過剰使用語の様相とは異なることが分かる。まず、品詞の種類の違いである。過剰使用語には名詞や形容詞が多く見られたが、過少使用語には、副詞が多いのが分かる。もう一点として、複数の品詞を持つ単語が多いことが分



かる。つまり、多義語が多い。また、過剰使用語に多くの人称代名詞主語が含まれているが、過少使用語にも *on* という人称代名詞主語が含まれている。

以上の点は、もちろん前後の文脈を考慮する必要があるため、後の MWUs 分析で追従することとする。

## 6.2. 自由会話コーパス

タスクに基づくコーパスの分析方法と同じく、過剰使用語と過少使用語を算出し、閾値を決定した。過剰使用語は  $G^2 \geq 50$  とし、過少使用は  $G^2 \leq -50$  とした。

表 37 過剰使用語と過少使用語

	過剰使用語	過少使用語
$100 \leq G^2$	INT	ENFIN
	OUI	HEIN
	NP	QUI
	FW	QUOI
	TRES	LES
	VISITE	ÇA
	BEAUCOUP	DES
	BC	
$50 \leq G^2 < 100$	ALLEE	FAIT
	FRANÇAIS	ILS
	JAPONAIS	OUAIS
	ALLE	QU'
	PENDANT	VOIS
	EST	LA
	C'	BON
	MUSIQUE	TOUT
	J'	AVOIR
	DIFFICILE	LEUR
	JE	MEME
	COMMENT	SAIS

次の節から、過剰使用語と過少使用語について、詳しい数値とともに、*VocabProfil* の頻度層との対応や品詞分類を行い、分析を進めていく。

### 6.2.1. 過剰使用語

過剰使用語のリストと、学習者と母語話者による実測値、10 万語あたりの調整頻度、レンジ、Keyness は、以下の通りである(表 38)。

表 38 過剰使用語の頻度と keyness

順位		学習者			母語話者			Keyness
		実測値	調整頻度	レンジ	実測値	調整頻度	レンジ	
1	INT	900	2607.34	100.00	402	612.89	100.00	656.16
2	OUI	1 034	2995.54	100.00	582	887.32	90.00	592.26
3	NP	1 150	3331.60	100.00	819	1248.65	100.00	477.91
4	FW	270	782.20	81.58	16	24.39	20.00	466.46
5	TRÈS	250	724.26	92.11	110	167.71	85.00	182.96
6	VISITÉ	74	214.38	44.74	1	1.52	5.00	147.92
7	BEAUCOUP	170	492.50	86.84	81	123.49	80.00	115.13
8	BC	673	1949.71	68.42	757	1154.12	75.00	97.34
9	ALLÉE	82	237.56	55.26	22	33.54	45.00	86.00
10	FRANÇAIS	103	298.40	68.42	39	59.46	35.00	85.51
11	JAPONAIS	60	173.82	50.00	10	15.25	25.00	78.87
12	ALLÉ	49	141.95	44.74	6	9.15	20.00	71.56
13	PENDANT	75	217.28	60.53	25	38.11	50.00	68.46
14	EST	1 334	3864.65	100.00	1 930	2942.48	100.00	59.49
15	C'	1 243	3601.02	100.00	1 789	2727.51	100.00	57.30
16	MUSIQUE	35	101.40	18.42	3	4.57	5.00	56.10
17	J'	605	1752.71	97.37	767	1169.37	95.00	54.95
18	DIFFICILE	46	133.26	57.89	10	15.25	30.00	53.90
19	JE	890	2578.36	100.00	1 227	1870.68	100.00	53.16
20	COMMENT	86	249.15	65.79	47	71.66	70.00	50.19

過剰使用語のレンジに着目することによって、話者が一般的に使用している語であるのか、ある特定の話者が使用しているのかが分かる。*INT*、*oui*、*NP*、*très*、*beaucoup*、*est*、*c'*、*j'*、*je*などは、学習者と母語話者ともに高いレンジの値が見られる。つまり、これらはほとんどの話者が使用する単語であるが、学習者によるこれらの単語の使用頻度が母語話者による頻度よりも高いことが分かる。一方、*FW*、*visité*、*français*、*japonais*は、学習者のレンジと比較して母語話者のレンジがかなり低い。この中でも特に*FW*のレンジの差が大きいが、*FW*が過剰使用語として抽出されたのは、転写方法が学習者と母語話者で若干異なっていたことに起因すると考えられる。なぜなら、母語話者は日本語が使用された場合のみ*FW*と転写したが、学習者では日本語が使用された場合だけでなく、主に英語発音のものも*FW*として転写を行ったからである。学習者が録音した際には、言葉にまつまっ際には日本語を使用しても良い旨が伝わっており、会話の中断を防ぐために日本語が用いられていた。同様の理由で、フランス語の単語が分からないが英語の単語を知っていた場合、英語が用いられる例も見られた。

VocabProfilにおける頻度層に従い、各過剰使用語がK1機能語、K1内容語、K2、K3、Off-listのどれに属するのかを分析したところ、表39のようになった。なお、*INT*、*NP*、*FW*、*BC*(間投詞、固有名詞、外国語、相づち)はこの分析の対象外とした。

表 39 過剰使用語と VP の頻度層との関係

順位	VP	過剰使用語
1	NA	INT
2	K1 機能語	OUI
3	NA	NP
4	NA	FW
5	K1 機能語	TRÈS
6	K2	VISITÉ
7	K1 内容語	BEAUCOUP
8	NA	BC
9	K1 内容語	ALLÉE
10	K1 内容語	FRANÇAIS
11	K1 内容語	JAPONAIS
12	K1 内容語	ALLÉ
13	K1 機能語	PENDANT
14	K1 機能語	EST
15	K1 機能語	C'
16	K1 内容語	MUSIQUE
17	K1 機能語	J'
18	K1 内容語	DIFFICILE
19	K1 機能語	JE
20	K1 内容語	COMMENT

以上の表より、学習者による過剰使用語は *visité* が K2 であった以外は、すべて K1 機能語ないし K1 内容語に属することが分かる。つまり、学習者による過剰使用語は、おおよそフランス語の頻度層第 1 位から 1000 位のものであると言える。

また、*VocabProfil* による頻度層と品詞の関係について、まとめたものが表 40 である。

表 40 *VocabProfil* の頻度層と品詞

VP	品詞	過剰使用語
K1 機能語	副詞	OUI
	副詞	TRÈS
	前置詞	PENDANT
	動詞、助動詞、名詞	EST
	代名詞	C'
	代名詞	J'
	代名詞	JE
K1 内容語	副詞	BEAUCOUP
	動詞	ALLÉE
	形容詞	FRANÇAIS
	形容詞	JAPONAIS
	動詞	ALLÉ
	名詞	MUSIQUE
	形容詞	DIFFICILE
K2	副詞	COMMENT
	動詞	VISITÉ

K1 機能語についてはもともと品詞が限定されているが、その中でも代名詞主語が多い (*c', je, j*)。K1 内容語では、国や国籍を表す形容詞 (*français, japonais*) や動詞 *ALLER* (行く) の過去分詞である *allée* と *allé* が過剰使用語として見られた。また、K2 で動詞 *VISITER* (訪れる) の過去分詞である (*visité*) が見られていることより、学習者は会話の主題の中から「訪れた場所(国、町、美術館、など)」を多く選んだことより、これらの動詞が過剰使用語として抽出された可能性がある。

#### 6.2.2. 過少使用語

過少使用語と判断された語における、学習者と母語話者による実測値、10 万語あたりの調整頻度、Keyness は、以下の通りである(表 41)。

表 41 過少使用語の頻度と keyness

順位		学習者			母語話者			Keyness
		実測値	調整頻度	レンジ	実測値	調整頻度	レンジ	
1	ENFIN	12	34.76	13.16	336	512.27	100.00	-205.81
2	HEIN	1	2.90	2.63	242	368.95	75.00	-194.09
3	QUI	95	275.22	57.89	615	937.63	100.00	-164.53
4	QUOI	40	115.88	42.11	387	590.02	100.00	-147.39
5	LES	201	582.30	89.47	892	1359.94	100.00	-140.46
6	ÇA	269	779.30	76.32	1 021	1556.62	100.00	-116.67
7	DES	143	414.28	65.79	651	992.51	100.00	-106.98
8	ILS	93	269.42	42.11	478	728.76	100.00	-95.24
9	FAIT	112	324.47	63.16	535	815.66	100.00	-95.23
10	OUAIS	268	776.41	31.58	923	1407.21	90.00	-82.10
11	QU'	103	298.40	81.58	477	727.23	100.00	-80.58
12	VOIS	9	26.07	15.79	152	231.74	70.00	-78.40
13	LÀ	72	208.59	44.74	379	577.82	100.00	-78.08
14	BON	65	188.31	73.68	332	506.17	100.00	-65.44
15	TOUT	62	179.62	52.63	315	480.25	100.00	-61.59
16	AVOIR	3	8.69	5.26	97	147.89	85.00	-61.51
17	LEUR	2	5.79	5.26	89	135.69	80.00	-60.33
18	MÊME	48	139.06	36.84	253	385.72	95.00	-52.15
19	SAIS	77	223.07	65.79	335	510.74	90.00	-50.54

母語話者と学習者の間に、頻度差があるのはもちろんのこと、レンジの差も注目に値する。過少使用語のレンジにおける学習者と母語話者のレンジの差は、過剰使用語よりも大きい。また過剰使用語と比較して、全般的に過少使用語のレンジは学習者の値がほとんど低く、母語話者の値は高くなっている。つまり、これらの単語をほとんどの母語話者が使用しているのに対して、学習者については一部の学習者のみがこれらの単語を使用している。この点について、語彙の豊かさの分析で行ったグループ分けに基づき、主成分得点の高かった上位 15 名までを「上級学習者」、その他を「初中級学習者」とし、この 2 グループ間で過少使用語のレンジがどのように変わるのかを比較した。

表 42 初級学習者と上級学習者による過少使用語のレンジ

順位		初中級学習者		上級学習者	
		実測値	レンジ	実測値	レンジ
1	ENFIN	0	0	12	33.33
2	HEIN	0	0	1	6.67
3	QUI	19	39.13	76	86.67
4	QUOI	12	30.43	28	60.00
5	LES	44	82.61	157	100.00
6	ÇA	35	60.87	234	100.00
7	DES	33	47.83	110	93.33
8	ILS	7	21.74	86	73.33
9	FAIT	26	52.17	86	80.00
10	OUAIS	38	13.04	230	60.00
11	QU'	54	86.96	49	73.33
12	VOIS	3	8.7	6	26.67
13	LÀ	17	26.09	55	73.33
14	BON	28	69.57	37	80.00
15	TOUT	11	30.43	51	86.67
16	AVOIR	0	0	3	13.33
17	LEUR	0	0	2	13.33
18	MÊME	7	13.04	253	95.00
19	SAIS	23	56.52	54	80.00

以上の表から、*qu'*を除いて上級学習者によるこれらの語のレンジが初中級学習者を全て大幅に上回っていることが分かる。これは、上級学習者達の多くがこれらの語を少なくとも一回は使用したのに対して、初中級学習者はこれらの語を使用した学習者の割合が少ないということを意味している。

以下の図は、初中級学習者、上級学習者、母語話者のレンジを比較したものである：

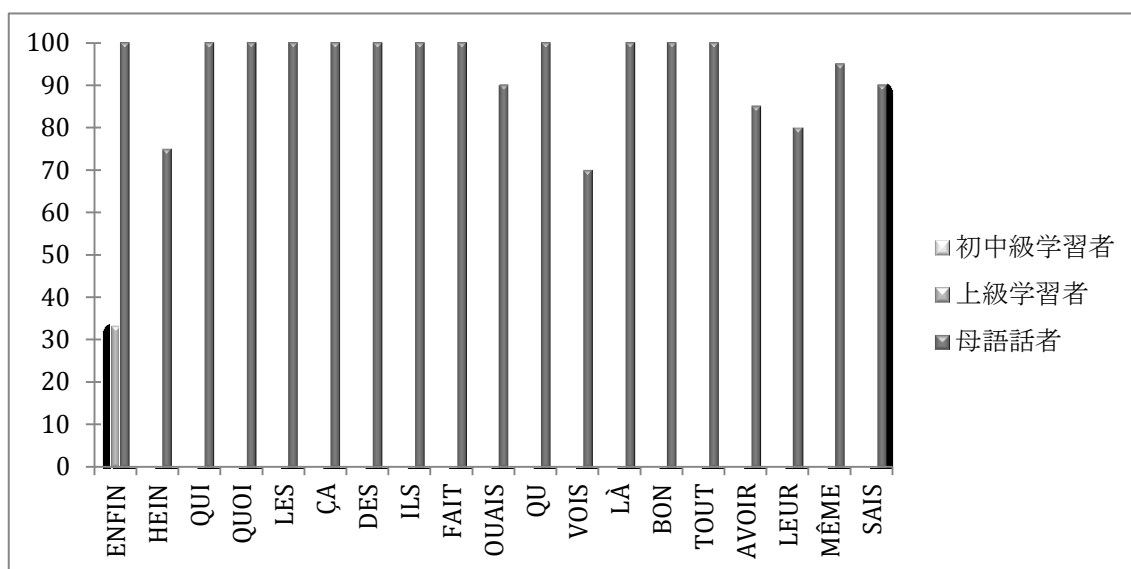


図 15 過少使用語のレンジの比較

*qu'*を除いて、全ての過少使用語のレンジは初中級学習者<上級学習者<母語話者の順番になっていることが分かる。このことから、母語話者と学習者の間で、頻度差のみが過少使用語の主な要因であるのではなく、これらの語を使用している人の割合も過少使用語の要因の一つであり、初級学習者よりも上級学習者の方が母語話者の割合に近いことが分かった。

つぎに、*VocabProfil*における頻度層に従い、過少使用語について K1 機能語、K1 内容語、K2、K3、Off-list のどれに属するのかを分析したところ、以下のようになった(表 43)。

表 43 過少使用語と VP の頻度層との関係

順位	VP	
1	K1 内容語	ENFIN
2	K1 内容語	HEIN
3	K1 機能語	QUI
4	K1 機能語	QUOI
5	K1 機能語	LES
6	K1 機能語	ÇA
7	K1 機能語	DES
8	K1 内容語	FAIT
9	K1 機能語	ILS
10	OFF-list	OUAIS
11	K1 機能語	QU'
12	K1 内容語	VOIS
13	K1 機能語	LÀ
14	K1 内容語	BON
15	K1 機能語	TOUT
16	K1 機能語	AVOIR
17	K1 内容語	LEUR
18	K1 内容語	MÊME
19	K1 内容語	SAIS

以上の表より、過少使用語においても、*ouais* を除いて全て K1 機能語と K1 内容語であることが分かる。これは、フランス語において高頻度語であるのにも関わらず、母語話者と比較して学習者はそれらの語彙を過小使用しているということを意味している。また、リスト外単語である *ouais* も過剰使用語として抽出された *oui* と同じ意味を持つが、くだけた言い方である。このため、書き言葉を基にした *VocabProfil* では *ouais* はリスト外単語として分類された。この 2 語がそれぞれ過剰使用語と過少使用語として抽出されたことは、注目すべき点である。つまり、学習者は形式的で規範的な *oui* を多く使用するのに対して、母語話者はより非形式的な *ouais* を使用する傾向にある。

次に、*VocabProfil* による頻度層と品詞の関係についてまとめたものが表 44 である。

表 44 *VocabProfil* の頻度層と品詞

VP	品詞	
K1 機能語	代名詞	QUI
	代名詞、間投詞	QUOI
	冠詞、代名詞	LES
	代名詞、間投詞	ÇA
	冠詞、縮約形	DES
	代名詞	ILS
	接続詞、代名詞、副詞	QU'
	副詞	LÀ
	形容詞、代名詞、副詞	TOUT
	動詞、助動詞	AVOIR
K1 内容語	副詞	ENFIN
	間投詞	HEIN
	名詞、動詞	FAIT
	動詞	VOIS
	形容詞	BON
	代名詞、形容詞	LEUR
	形容詞、代名詞、副詞	MÊME
	動詞	SAIS
Off-list	副詞	OUAIS

過剰使用語と比較すると、過少使用語は異なる意味や機能を持つ単語が多いことが分かる。つまり、多義語が多い。また、過剰使用語として間投詞が抽出されたが、INT と転写されたものは一音節から構成されており、主に長母音母音に特徴づけられる間投詞であった。これらの間投詞は、日本語にもフランス語にも存在するものである。一方、過少使用語では、*quoi*、*ça*、*hein* といった、フランス語に特有な間投詞で構成されている。

### 6.3. まとめ

タスクに基づくコーパスも自由会話コーパスも、過剰使用語と過少使用語のほとんどは高頻度語であることが分かった。また、過剰使用語には代名詞が多く含まれており、過少使用語は多義語が多いという点も共通していた。また、自由会話コーパスに基づく分析で、学習者のレベル別に過少使用語のレンジを比較したところ、初中級学習者<上級学習者<母語話者の順となり、レベルが上がるにつれて、過少使用語を使用する人の割合が増えることが分かった。



次章では、本章で明らかになった過剰使用語と過少使用語がどのような文脈で使用されているのかについて MWUs レベルから特徴語の分析を行う。

## 第七章 コロケーション分析

本章では、前章で明らかになった学習者による過剰使用語と過少使用語について、これらの特徴語をより広いコンテキストに当てはめて分析していく。共起環境を概観することによって、これらの特徴語が WUs レベルにおいて特徴語であるのか、それとも他の語との共起から引き起こされる MWUs レベルでの特徴語なのかを明らかにする。

### 7.1. タスクに基づくコーパス

#### 7.1.1. 過剰使用語との共起語

前節で抽出した過剰使用語について、調整頻度 20 回以上共起した語を前後 3 語ずつまとめた。これらの MWUs について、個々に分析を行っていく。なお、表における共起語および、特徴語の隣にある数字は調整頻度を示し、カッコ内の数字は生起回数を示している。

まず、表をもとに学習者コーパスと母語話者コーパスにそれぞれ特徴的なコロケーションおよびコリゲーションを選択し、比較する。そして、それらの MWUs がタスクに大きく依存するものである場合は、どのようなタスクのもとで MWUs が現れるのかを分析する。

#### 7.1.1.1. *voudrais*

学習者と母語話者による *voudrais* の共起語は以下の通りである：

表 45 学習者による *voudrais* の共起語<sup>48</sup>

L3	L2	L1	KW		R1	R2	R3						
CE	49.23 (14)	NP	59.78 (17)	JE	710.37 (202)	VOUDRAIS	745.53 (212)	ALLER	133.63 (38)	À	116.05 (33)	LA	66.82 (19)
À	35.17 (10)	WEEKEND	56.27 (16)	NE	31.65 (9)			RÉSERVER	59.78 (17)	UN	80.88 (23)	DES	42.2 (12)
		MAIS	42.2 (12)					RESTER	59.78 (17)	LE	59.78 (17)	NP	38.68 (11)
		JE	35.17 (10)					UN	52.75 (15)	AU	56.27 (16)	DE	24.62 (7)
		QUE	28.13 (8)					ACHETER	49.23 (14)	DES	31.65 (9)	QUI	24.62 (7)
		ET	21.1 (6)					SORTIR	42.2 (12)	AVEC	31.65 (9)	S'	24.62 (7)
								FAIRE	38.68 (11)	OFFICIEL	28.13 (8)	BILLET	21.1 (6)
								PRENDRE	38.68 (11)	NP	28.13 (8)	FRANÇAIS	21.1 (6)
								L'	31.65 (9)	BILLET	24.62 (7)	RESTAURANT	21.1 (6)
								PAS	31.65 (9)	CAFÉ	24.62 (7)		

<sup>48</sup> L は中心語の左側、R は右側を指し、数字は中心語からどれだけ離れているのかを指す。

表 46 母語話者による *voudrais* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
—	—	BONJOUR	44.38 (23)	JE	113.85 (59)	VOUDRAIS	113.85 (59)	UN	40.52 (21)	UN	23.16 (12)	BILLET	23.16 (12)
								L'	21.23 (11)	OFFICIEL	21.23 (11)	DES	21.23 (11)

以上の2つの表から、2つのグループとも *voudrais* という動詞の主語は全て *je* であることがわかる。しかし、その他の点で違いが見られた。学習者と母語話者の *voudrais* の使用方法の違いについてまとめたものが、以下の表である。

表 47 学習者および母語話者に特徴的な MWUs (*voudrais*)

MWUs	NS	NNS	種類	NS	NNS
<i>je voudrais</i> +動詞句	21	160			
<i>je voudrais</i> +名詞句	33	28			
時間+ <i>je voudrais</i>	1	37	weekend	1	16
			soir	0	6
			その他	0	15
<i>parce que je voudrais</i>	0	6			
呼びかけ+ <i>je voudrais</i>	31	8	excusez-moi	6	6
			bonjour	23	2
			その他	2	0
<i>je ne voudrais pas</i>	0	9			

まず2グループの差で顕著な点は、学習者は *je voudrais* の右側に動詞句を多く用いているが、母語話者は名詞句の方を多く使用しているということである。そこで *je voudrais*+動詞句と *je voudrais*+名詞句の頻度についてピアソンのカイ二乗検定を行った結果、以下の通りとなった。

表 48 *je voudrais*+動詞句と *je voudrais*+名詞句の使用頻度と有意差検定

	母語話者頻度	学習者頻度	$\chi^2$ 統計量、有意確率
<i>je voudrais</i> +動詞句	21	160	$\chi^2=47.53$ ; $p<.001$
<i>je voudrais</i> +名詞句	33	28	

つまり、*je voudrais*+動詞句について、学習者は母語話者よりも多く使用しているということを意味している。または、*je voudrais*+名詞句が有意に少ないとも言える。これは、学習者はこの語を『～したい』という意味で用いるのに対して、母語話者は『～が欲しい(～をください)』という意味として使用する傾向があることに起因している。このことは、学習者コーパス内にある *ce weekend* のような時間を表す語との共起頻度が高いことから裏付けられる。なぜなら、時間を表す語が先行する場合は、『～が欲しい(～をください)』よりも『～したい』という意味を持つ文や語句が後続することが自然であると言えるからである。また、以上から *parce que*(なぜなら)との共起も同様の説明ができるだろう。実際に、*parce que je voudrais* という MWUs の後に後続するのは全て動詞句であった。

je voudrais rester à la maison **parce que je voudrais étudier** le français et je voudrais aider ma mère  
(JT10 ; j002)<sup>49</sup>

また、『～したい』という意味で使用された *je voudrais* の中には、*je voudrais rester à la maison* の一部として使用され、また以下のタスクの答えとして産出されたものがある。

表 49 *je voudrais rester à la maison* の生起回数<sup>50</sup>

タスク番号	タスク	学習者 (生起回数)	母語話者 (生起回数)
j002	長期休暇中、どこで過ごすのが好きですか？	10.61 (2)	0 (0)
*j021.	あなたはフランス人の知り合いにフランス語で話しかけられました。 質問を聞いて、それについてフランス語で即答してください。 Qu'est-ce que vous voudriez faire ce week-end. rester a la maison ou sortir?	58.35 (11)	0 (0)
	計	68.96 (13)	0 (0)

タスクの答えの内容として、この生起回数の差が見られた可能性も考えられるため、*rester à la maison* のみに限定して比較した結果、以下ようになった：

表 50 *rester à la maison* の生起回数

タスク番号	タスク	学習者 (生起回数)	母語話者 (生起回数)
j002	長期休暇中、どこで過ごすのが好きですか？	13.22 (3)	0 (0)
j021	あなたはフランス人の知り合いにフランス語で話しかけられました。 質問を聞いて、それについてフランス語で即答してください。 Qu'est-ce que vous voudriez faire ce week-end. rester a la maison ou sortir?	74.94 (17)	24.04 (11)
	計	88.16 (20)	24.04 (11)

*rester à la maison* という MWUs についても生起回数は学習者の方が多いが、内容の差だけではないことが分かる。つまり、母語話者は *rester à la maison* という MWUs に対して *je voudrais* という MWUs を使用しないということである。母語話者が *rester à la maison* という MWUs とともに使用した MWUs は以下の表 51 のとおりである。なお、この MWUs の間に副詞やためらい語が含まれたものも同じ MWUs としてカウントした。

<sup>49</sup> J=日本人学習者、T=タスクに基づくコーパス、数字(前者)=被験者番号、j=日本人学習者用タスク、数字(後者)=タスク番号

<sup>50</sup> 数字は、「10万 MWUs あたりの調整頻度(生起回数)」を示す。以下同様。

表 51 母語話者による *rester à la maison* との共起 MWUs

MWUs	生起回数
je vais	4
je préfère	2
je + aime + (bien)	3
(j'aime bien, j'aimerais, j'aimerais bien)	

一方、母語話者は呼びかけ表現(*bonjour* こんにちは)とともに *je voudrais* を多く用いている。表には見られないが、*excusez-moi*(すみません)との共起も 6 回見られた。このことは、母語話者が『～が欲しい(～をください)』という意味でこの句を多用しているということを意味している。これらは、タスク f022、f034、f047 で見られた。f022 はカフェでコーヒーを頼むタスク、f034 は駅で切符を買うタスク、f047 はキオスクで雑誌を買うタスクである。*voudrais* の生起頻度に対する、呼びかけ表現+*je voudrais* の頻度について、カイ二乗によって有意差検定を行ったところ、有意差が見られた ( $\chi^2=84.80$ ,  $p<.001$ )。

**bonjour je voudrais** un billet HESIT de NP jusqu'à NP pour une personne HESIT le train de quatorze heures quarante (FT25 ; f034)<sup>51</sup>

**bonjour excusez moi** HESIT **je voudrais** l'officiel des spectacles s'il vous plaît (FT39 ; f047)

**excusez-moi bonjour je voudrais** un café s'il vous plaît (FT03 ; f022)

また、*je voudrais* の否定文にあたる *je ne voudrais pas* が学習者にしか用いられていない点についても、学習者は『～したい』の否定文である『～したくない』という意味でこの MWUs を使用していることが考えられる。実際にこの否定文は一例をのぞいて j027 で用いられていた。

j027	週末にあなたがしたいこと、そうでないことについて話してください。
------	----------------------------------

je voudrais aller voir un film mais **je ne voudrais pas** étudier (JT01 ; j027)

je voudrais jouer du piano **je ne voudrais pas** travailler (JT08 ; j027)

一方、母語話者はこのタスクで『したくないこと』を表現する際に、主に *j(e n') aime pas* という別の MWUs を使用している。

donc le weekend **j'aime pas** travailler ni étudier mais c'est un peu obligatoire (FT14 ; f027)

**j'aime pas** voir des gens que je n' aime pas le weekend (FT25 ; f027)

<sup>51</sup> F=フランス語母語話者、T=タスクに基づくコーパス、数字(前者)=被験者番号、f=フランス語母語話者用タスク、数字(後者)=タスク番号

しかし、この結果は一概に MWUs の選択の差であるとは言い切れない。なぜなら、母語話者が行ったタスクの指示に以下の MWUs が含まれているからである。

f027	Indiquez ce que vous aimez ou ce que <u>vous n'aimez pas</u> faire le week-end.
------	---

指示文が母語話者の発話を左右した可能性は否定できないが、*je ne voudrais pas* という MWUs を母語話者は一例も使用していないという点は、無視することはできない。

#### 7.1.1.1.2. *elle*

*elle* は学習者だけでなく、母語話者にも多く用いられた人称代名詞であるが、高い Keyness が見られた。学習者によって最も多く使用された *elle* の共起語は表 52 の通りである。*elle* のコンコーダンスを見てみると、右側に続く動詞の上位 3 つは同じであることが分かる。そこで、*elle* が使用される際に、それらの動詞が使用される割合について学習者と母語話者に有意差があるかどうかを調べたところ、以下の通りとなった。

表 52 *elle* と共に用いられる動詞

	母語話者頻度	学習者頻度	$\chi^2$ 統計量、有意確率
<i>elle est</i>	113	153	$\chi^2=0.74$ ; $p>.05$
<i>elle a</i>	46	92	$\chi^2=3.72$ ; $p>.05$
<i>elle habite</i>	26	52	$\chi^2=1.88$ ; $p>.05$

*elle* と最も共起された 3 つの動詞の使用頻度については有意差が見られなかった。したがって、母語話者も学習者も *elle* を使用する際にはこれらの動詞を共に使用しているということである。これらは、学習者と母語話者によって以下のタスクで主に用いられていた。

j025	<p>以下の身分証明書をよく見てください。</p> <p>Nom: Tanaka          Prénom: Kazuko          Nationalité: Japonaise          Âge: 38ans          Profession: Professeur          Situation de famille: Mariée – 2 enfants          Adresse au Japon: 3 Asahi-cho Fuchu          Adresse en France: 67, rue de la Liberté 69000 Lyon          この人物を紹介してください。</p>
------	--

*elle est* という MWU は、国籍、職業、社会的立場(既婚)を表す際に、*elle a* という MWU は年齢と子供の数を説明する際に、*elle habite* という MWU は住所を説明する際に主に用いられていた。そのうち、最も用いられていた MWUs とその生起頻度は以下の

通りである。この MWUs 内にためらい語や別の語句が含まれていても、同じ機能を持つ場合は頻度にカウントする。

表 53 タスク 025 で用いられていた文例

	母語話者	学習者
<i>elle est</i>		
<i>elle est japonaise</i>	11	34
<i>elle est professeur de japonais</i>	25	34
<i>elle est mariée</i>	26	29
<i>elle a</i>		
<i>elle a trente huit ans</i>	20	40
<i>elle a deux enfants</i>	11	32
<i>elle habite</i>		
<i>elle habite à</i>	23	35

いずれの場合も同タスクにおいて *elle* が主語に置かれる頻度は母語話者よりも学習者の方が高いことが分かる。つまり、母語話者はこのタスクを行う際に、必ずしも主語 *elle* を使用していないということである。母語話者によって主語 *elle* が使用されなかったパターンについてまとめたものが以下の表である。この MWUs の中にためらい語やほかの語や句が挿入されていても、機能が変わらなかった場合は頻度を含めた。

表 54 *elle* を使用しない文例

	母語話者	学習者
国籍		
<i>c'est une japonaise</i>	5	0
名前 <i>de nationalité japonaise</i>	3	0
先行詞 <i>qui est japonaise</i>	2	0
職業		
<i>c'est un(e) professeur de japonais</i>	3	0
既婚		
名詞+ <i>mariée</i> (形容詞として)	4	0
<i>et mariée</i>	2	0
年齢		
句 <sup>52</sup> <i>de trente huit ans</i>	11	0
子供の数		
<i>(et) a deux enfants</i>	17	7
句 <sup>53</sup> <i>avec deux enfants</i>	5	0
住所		
<i>son (l') adresse (c') est</i>	28	38

さらに、このタスクにおいて 2 グループ間の大きな違いがあらわれたのは、*elle s'appelle* という MWU の頻度差であった。学習者はこのタスクについて *elle s'appelle* という MWUs を使用している例が 32 例見られたのに対して、フランス語母語話者はこのタスクにおいて、*elle s'appelle NP NP* で始める例が 1 例も見られなかった。フランス語母語話者はこのタスクを行う際に、以下のような MWUs を使用する傾向にある。なお、この MWUs 内にためらい語や間投詞が入る例も、生起回数に含めてある：

<sup>52</sup> *c'est un professeur de trente huit ans* のように、前の説明文に付随しているもの

<sup>53</sup> 同上

表 55 フランス語母語話者による第三者の紹介に使用する MWUs

第三者の紹介	生起回数
je + PRESENTER + 名前 (je vous présente, je vais présenter, je veux présenter etc..)	9
voici + 名前	8
cette personne s'appelle + 名前	5
c'est + 名前	4

上記の表の通り、第三者に対する紹介において、*elle s'appelle* という MWUs を使用することは一般的ではない。また、S'APPELER という動詞を使用するとしても、主語には人称代名詞主語を置かず、*cette personne* という主語を選択していた。実際に、同タスクにおいて、*elle s'appelle* という MWU を使用するのは不適切であると言える。なぜなら、*elle* は代名詞であり前方照応機能を持つため、発話単位の先頭で *elle* を用いる際には話者の間でこの人物が既に共通認識となっていなくてはならないからである。



表 56 学習者による *elle* の共起語

L3	L2	L1	KW		R1	R2	R3						
ELLE	137.15 (39)	NP	200.45 (57)	NP	320.02 (91)	ELLE	1364.47 (388)	EST	538.05 (153)	APPELLE	172.32 (49)	NP	291.88 (83)
EST	137.15 (39)	EST	161.77 (46)	ANS	182.87 (52)	A	323.53 (92)	A	323.53 (92)	À	154.73 (44)	ELLE	137.15 (39)
TRENTE	126.6 (36)	HUIT	133.63 (38)	ET	154.73 (44)	HABITE	182.87 (52)	TRENTE	182.87 (52)	TRENTE	140.67 (40)	HUIT	133.63 (38)
APPELLE	119.57 (34)	DE	94.95 (27)	JAPONAISE	144.18 (41)	S'	179.35 (51)	JAPONAISE	179.35 (51)	JAPONAISE	130.12 (37)	ENFANTS	116.05 (33)
PROFESSEUR	94.95 (27)	MARIÉE	87.92 (25)	JAPONAIS	105.5 (30)			DEUX	123.08 (35)	DEUX	123.08 (35)	ET	116.05 (33)
NP	70.33 (20)	DEUX	49.23 (14)	ENFANTS	42.20 (12)			PROFESSEUR	98.47 (28)	PROFESSEUR	98.47 (28)	DE	87.92 (25)
A	56.27 (16)	EN	42.2 (12)	QU'	31.65 (9)			MARIÉE	80.88 (23)	MARIÉE	80.88 (23)	SOIXANTE	35.17 (10)
ET	35.17 (10)	DU	31.65 (9)					UNE	45.72 (13)	UNE	45.72 (13)	PROFESSEUR	31.65 (9)
À	28.13 (8)	AU	31.65 (9)					TRÈS	35.17 (10)	TRÈS	35.17 (10)	DU	28.13 (8)
ENFANTS	28.13 (8)	APPELLE	28.13 (8)					LE	28.13 (8)	LE	28.13 (8)	NEUF	21.1 (6)
S'	28.13 (8)	PARCE	24.62 (7)					DIX	24.62 (7)	DIX	24.62 (7)	ANS	21.1 (6)
DE	24.62 (7)	À	21.10 (6)					NP	21.1 (6)	NP	21.1 (6)		
DIX	21.10 (6)							ÉTUDIANTE	21.1 (6)	ÉTUDIANTE	21.1 (6)		
								EST	21.10 (6)	EST	21.10 (6)		

表 57 母語話者による *elle* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
ELLE	71.40 (37)	EST	79.12 (41)	NP	61.75 (32)	ELLE	509.43 (264)	EST	218.05 (113)	MARIÉE	50.17 (26)	HUIT	36.66 (19)
EST	42.45 (22)	HUIT	30.87 (16)	HESIT	48.24 (25)			A	86.83 (45)	PROFESSEUR	48.24 (25)	ET	32.80 (17)
TRENTE	30.87 (16)	NP	30.87 (16)	ET	44.38 (23)			HABITE	50.17 (26)	AU	42.45 (22)	NP	25.09 (13)
ET	27.02 (14)	EN	28.94 (15)	PROFESSEUR	34.73 (18)			VIT	23.16 (12)	TRENTE	36.66 (19)	ENFANTS	23.16 (12)
				ANS	34.73 (18)					DEUX	23.16 (12)		
				JAPONAISE	30.87 (16)								
				QU	23.16 (12)								

### 7.1.1.3. NP

NP は固有名詞と言うこともあり、前後に使用される語も様々である。そのため、それぞれのコーパスの傾向も見えにくい。そこで、NP を中心とした前後 3 語の 7 語連鎖で頻度が高かったものを比較した。

学習者が最も多く使用した 7 語連鎖は、*elle s'appelle NP NP elle est* であった。学習者コーパスの中では 64 例見られたのに対して、母語話者コーパスには一例も見られなかった。これは、既に前節で扱った *elle s'appelle* という MWUs による影響であると考えられる。一方、母語話者コーパスで最も多く見られた MWUs は *je m'appelle NP NP je suis* というものであり、48 例見られた。これは、学習者コーパスにおいても NP を中心に置く 7 語連鎖の 2 番目に多く見られ、56 例あった。この 2 つの 7 語連鎖である *elle s'appelle NP NP elle est* と *je m'appelle NP NP je suis* とを比べると、統語構造は全く同じであり、主語が異なるだけである。この MWUs の後ろには、国籍を示す形容詞もしくは *de nationalité*+形容詞などの前置詞句が続く。これらは、タスク 024 と 025 で使用されていた。

j024	自己紹介してください。(苗字、名前、国籍、年齢、職業、住所、家族構成など)
f024	Présentez-vous. (nom, prénom, nationalité, âge, profession, adresse, situation de famille, etc)
j025	以下の身分証明書をよく見てください。 Nom: Tanaka Prénom: Kazuko Nationalité: Japonaise Âge: 38ans Profession: Professeur Situation de famille: Mariée – 2 enfants Adresse au Japon: 3 Asahi-cho Fuchu Adresse en France: 67, rue de la Liberté 69000 Lyon この人物を紹介してください。

f025	Prénom: Kazuko Nom : Tanaka Nationalité: Japonaise Âge: 38ans Profession: Professeur Situation de famille: Mariée – 2 enfants Adresse au Japon: 3 Asahi-cho Fuchu Adresse en France: 67, rue de la Liberté 69000 Lyon  Observez la fiche d'identité ci-dessous et présentez cette personne.
------	--

両タスクの指示文とも名前の後に国籍が並んでいるため、以上の 7 語連鎖のような構造となったのは、指示文の影響であると言える。

一方、フランス語母語話者コーパスで 2 番目に多く見られた NP を中心に置く 7 語連鎖は *de NP à NP il faut prendre*(NP から NP まで～に乗らなければならない)という MWUs であり、38 例見られた。この MWUs は学習者コーパスには一例も見られない。これはタスク 009 で用いられた MWUs である。

j009	あなたは旅行者から成田空港で、上野駅までの行き方を尋ねられました。 以下の目的地への行き方と所要時間を参照して説明してください。 成田 → 日暮里 → 上野 京成線          山手線 (1時間10分)      (5分)
f009	Narita → Nippori → Ueno La ligne Keisei La ligne Yamanote (1h10) (5min) On vous demande comment aller de Narita (aéroport) à Ueno. Expliquez comment y aller et combien de temps il faut.

このタスクにおいて、学習者によっても *de NP à NP* という MWUs は用いられており、23 例見られた。母語話者コーパスに見られた MWUs である *de NP à NP il faut prendre* は、発話単位の頭の位置に区間を示す前置詞句が置かれている。前置詞句を発話単位の先頭に持ってくることによって、話題の導入を行っていると考えられる。

一方、学習者コーパスの中にはこのような例は 1 例しか見られなかった。

**pour aller de NP à NP** il faut aller de NP à NP (JT29; j009)

表 58 学習者による NP の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
DE	422.00 (120)	NP	720.92 (205)	À	1339.85 (381)	NP	5436.77 (1546)	À	625.97 (178)	EST	344.63 (98)	À	355.18 (101)
JE	379.80 (108)	APPELLE	274.30 (78)	DE	900.27 (256)			ET	432.55 (123)	JE	203.97 (58)	HEURES	267.27 (76)
NP	365.73 (104)	HABITE	270.78 (77)	NP	749.05 (213)			ELLE	320.02 (91)	ELLE	200.45 (57)	EST	232.10 (66)
ELLE	291.88 (83)	LA	218.03 (62)	APPELLE	337.60 (96)			JE	270.78 (77)	DIX	200.45 (57)	JAPONAISE	186.38 (53)
SUIS	211.00 (60)	S'	211.00 (60)	LE	295.40 (84)			DE	221.55 (63)	À	147.70 (42)	DE	168.80 (48)
VOUS	207.48 (59)	DE	179.35 (51)	EST	196.93 (56)			IL	179.35 (51)	QUATORZE	144.18 (41)	SEPT	147.70 (42)
LE	193.42 (55)	ALLER	172.32 (49)	LA	168.80 (48)			POUR	161.77 (46)	SUIS	123.08 (35)	LE	147.70 (42)
J	182.87 (52)	À	133.63 (38)	AU	165.28 (47)			QUI	116.05 (33)	HABITE	94.95 (27)	SUIS	123.08 (35)
ET	151.22 (43)	ALLÉE	133.63 (38)	EN	147.70 (42)			J'	109.02 (31)	VOUS	94.95 (27)	UNE	105.50 (30)
À	144.18 (41)	M'	126.60 (36)	POUR	137.15 (39)			C'	105.50 (30)	ET	91.43 (26)	JE	101.98 (29)
S'	144.18 (41)	EST	126.60 (36)	LIGNE	116.05 (33)			AVEC	98.47 (28)	PART	87.92 (25)	HUIT	91.43 (26)
M'	130.12 (37)	PRENEZ	112.53 (32)	DU	98.47 (28)			EST	91.43 (26)	FAIT	84.40 (24)	ET	84.40 (24)
LA	116.05 (33)	JUSQU	105.50 (30)	TROIS	80.88 (23)			LIGNE	91.43 (26)	J'	77.37 (22)	ELLE	70.33 (20)
RUE	112.53 (32)	ADRESSE	98.47 (28)	ET	45.72 (13)			JUSQU	80.88 (23)	VINGT	77.37 (22)	IL	66.82 (19)
EST	87.92 (25)	ARRIVE	91.43 (26)	GARE	24.62 (7)			VOUS	77.37 (22)	DE	66.82 (19)	VOUS	56.27 (16)
SON	84.40 (24)	C'	80.88 (23)	PRENDRE	24.62 (7)			SON	66.82 (19)	ADRESSE	63.30 (18)	AI	56.27 (16)
UN	80.88 (23)	LIGNE	80.88 (23)	SUIS	21.10 (6)			ÇA	63.30 (18)	AI	63.30 (18)	CINQ	52.75 (15)
PRENEZ	73.85 (21)	LE	77.37 (22)	D'	21.10 (6)			EN	63.30 (18)	VOUDRAIS	59.78 (17)	TRÈS	45.72 (13)
QUI	73.85 (21)	VIENS	73.85 (21)					PARCE	56.27 (16)	FAUT	52.75 (15)	A	45.72 (13)
HABITE	66.82 (19)	TROIS	70.33 (20)					ON	49.23 (14)	A	52.75 (15)	CE	45.72 (13)
ON	66.82 (19)	PRENDS	70.33 (20)					LE	45.72 (13)	Y	49.23 (14)	UN	45.72 (13)

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
POUR	66.82 (19)	ET	63.30 (18)					HESIT	45.72 (13)	TROIS	49.23 (14)	POUR	45.72 (13)
HEURES	52.75 (15)	GARE	56.27 (16)					TU	38.68 (11)	QUE	49.23 (14)	JAPONAIS	42.20 (12)
ALLER	49.23 (14)	SUIS	52.75 (15)					TROIS	31.65 (9)	EN	42.20 (12)	DIX	42.20 (12)
IL	45.72 (13)	TRAIN	49.23 (14)					NOUS	28.13 (8)	SOIXANTE	42.20 (12)	EN	42.20 (12)
AU	42.20 (12)	PREND	49.23 (14)					MAIS	28.13 (8)	SON	38.68 (11)	ADRESSE	38.68 (11)
BONJOUR	42.20 (12)	ARRIVER	49.23 (14)					QU'	28.13 (8)	PRENEZ	35.17 (10)	LA	38.68 (11)
D'	42.20 (12)	JE	42.20 (12)					APRÈS	24.62 (7)	IL	31.65 (9)	QUI	38.68 (11)
DU	42.20 (12)	ARRIVEZ	42.20 (12)					AU	24.62 (7)	TU	31.65 (9)	PRENDRE	35.17 (10)
C'	38.68 (11)	PART	35.17 (10)							DÉPART	31.65 (9)	HABITE	35.17 (10)
LES	38.68 (11)	QUARANTE	35.17 (10)							PRENDS	28.13 (8)	SON	35.17 (10)
LIGNE	38.68 (11)	VOUS	28.13 (8)							MON	28.13 (8)	C'	35.17 (10)
VOUDRAIS	38.68 (11)	QUI	28.13 (8)							CE	28.13 (8)	TRENTE	31.65 (9)
DEUX	35.17 (10)	VOUDRAIS	28.13 (8)							CINQ	28.13 (8)	QUE	28.13 (8)
L'	35.17 (10)	AI	28.13 (8)							SEPT	24.62 (7)	L'	28.13 (8)
TROIS	35.17 (10)	OÙ	28.13 (8)							LA	24.62 (7)	ÇA	28.13 (8)
TU	35.17 (10)	PRENDRE	24.62 (7)							LE	24.62 (7)	ALLER	28.13 (8)
AI	28.13 (8)	ELLE	21.10 (6)							NOUS	24.62 (7)	SOIXANTE	24.62 (7)
POUVEZ	28.13 (8)	FILM	21.10 (6)							L'	24.62 (7)	JUSQU'	24.62 (7)
SA	28.13 (8)	CHANGE	21.10 (6)							CHANGE	21.10 (6)	VINGT	21.10 (6)
AIME	24.62 (7)									PREND	21.10 (6)	NOUS	21.10 (6)
ALLEZ	24.62 (7)									PUIS	21.10 (6)	LES	21.10 (6)
MON	24.62 (7)									PARTIR	21.10 (6)	MINUTES	21.10 (6)

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
AVEC	21.10 (6)												
MINUTES	21.10 (6)												
PART	21.10 (6)												
VAIS	21.10 (6)												

表 59 母語話者による NP の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
NP	293.31 (152)	NP	441.89 (229)	À	960.96 (498)	NP	3612.3 (1872)	À	397.51 (206)	EST	165.95 (86)	À	133.15 (69)
DE	181.39 (94)	LA	171.74 (89)	NP	468.9 (243)			ET	245.06 (127)	JE	162.09 (84)	HEURES	119.64 (62)
JE	165.95 (86)	DE	156.3 (81)	DE	384 (199)			HESIT	227.7 (118)	À	148.58 (77)	SUIS	109.99 (57)
POUR	133.15 (69)	JUSQU	144.72 (75)	EN	237.35 (123)			JE	142.79 (74)	FAUT	104.2 (54)	EST	98.41 (51)
SUIS	133.15 (69)	ALLER	125.43 (65)	LIGNE	148.58 (77)			IL	135.08 (70)	HESIT	102.27 (53)	PRENDRE	96.48 (50)
À	113.85 (59)	LE	96.48 (50)	AU	127.36 (66)			C'	117.71 (61)	IL	102.27 (53)	JE	88.76 (46)
PRENDRE	104.2 (54)	HESIT	96.48 (50)	LA	115.78 (60)			DE	92.62 (48)	ET	90.69 (47)	IL	79.12 (41)
LE	96.48 (50)	APPELLE	81.05 (42)	LE	106.13 (55)			JUSQU'	82.97 (43)	QUATORZE	79.12 (41)	HESIT	71.4 (37)
LA	92.62 (48)	ET	79.12 (41)	APPELLE	106.13 (55)			EN	73.33 (38)	EN	69.47 (36)	LA	71.4 (37)
RUE	77.19 (40)	À	77.19 (40)	TROIS	69.47 (36)			ELLE	61.75 (32)	DIX	69.47 (36)	UNE	63.68 (33)
J'	75.26 (39)	TROIS	75.26 (39)	EST	55.96 (29)			S'	61.75 (32)	SUIS	67.54 (35)	VOUS	61.75 (32)
M'	69.47 (36)	M'	67.54 (35)	MAGASIN	48.24 (25)			POUR	61.75 (32)	C'	65.61 (34)	C'	59.82 (31)
C'	67.54 (35)	HABITE	61.75 (32)	HESIT	38.59 (20)			QUI	57.89 (30)	VINGT	52.1 (27)	EN	57.89 (30)
AU	63.68 (33)	C'	57.89 (30)	UN	38.59 (20)			DEPUIS	57.89 (30)	PREND	48.24 (25)	ET	55.96 (29)
EN	59.82 (31)	ARRIVE	55.96 (29)	DU	36.66 (19)			ON	55.96 (29)	J'	42.45 (22)	DE	55.96 (29)

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
EST	57.89 (30)	EST	48.24 (25)	POUR	34.73 (18)			DONC	55.96 (29)	LE	42.45 (22)	LE	54.03 (28)
ON	55.96 (29)	NEUF	46.31 (24)	D'	30.87 (16)			ÇA	40.52 (21)	LA	38.59 (20)	SEPT	42.45 (22)
UN	55.96 (29)	ADRESSE	44.38 (23)	L'	27.02 (14)			SOIXANTE	40.52 (21)	UNE	38.59 (20)	J'	38.59 (20)
HESIT	48.24 (25)	NÉE	42.45 (22)	RUE	25.09 (13)			AU	40.52 (21)	PRENDS	36.66 (19)	FRANÇAISE	36.66 (19)
QUI	44.38 (23)	AU	40.52 (21)	TRAM	25.09 (13)			J	38.59 (20)	HABITE	36.66 (19)	ON	34.73 (18)
L'	38.59 (20)	S'	38.59 (20)	SUR	23.16 (12)			TROIS	38.59 (20)	NEUF	34.73 (18)	MILLE	32.8 (17)
ET	36.66 (19)	L'	36.66 (19)	MILLE	23.16 (12)			PENDANT	36.66 (19)	DE	34.73 (18)	HUIT	28.94 (15)
PRENDS	36.66 (19)	ÉTAIT	36.66 (19)					TU	36.66 (19)	VEUX	32.8 (17)	AU	28.94 (15)
SON	36.66 (19)	BILLET	30.87 (16)					MAIS	32.8 (17)	QUI	32.8 (17)	PREND	27.02 (14)
HABITE	34.73 (18)	TRAIN	30.87 (16)					EST	30.87 (16)	ON	32.8 (17)	AI	27.02 (14)
DIX	32.8 (17)	VIS	30.87 (16)					DANS	25.09 (13)	ELLE	30.87 (16)	JAPONAISE	27.02 (14)
HEURES	32.8 (17)	RUE	27.02 (14)					D'	23.16 (12)	ÇA	27.02 (14)	QUOI	25.09 (13)
SOIXANTE	30.87 (16)	PRENDRE	25.09 (13)							SON	27.02 (14)	UN	25.09 (13)
ALLER	28.94 (15)	NÉ	25.09 (13)							DONC	25.09 (13)	ADRESSE	25.09 (13)
DONC	28.94 (15)	RÉGION	25.09 (13)							AU	25.09 (13)	SOIXANTE	25.09 (13)
MES	28.94 (15)	POUR	23.16 (12)							ÉTAIT	25.09 (13)	SON	23.16 (12)
ELLE	25.09 (13)	UN	23.16 (12)							UN	23.16 (12)	CE	23.16 (12)
FAUT	23.16 (12)	PREND	21.23 (11)							LES	21.23 (11)	HEURE	23.16 (12)
NEUF	21.23 (11)	ENSUITE	21.23 (11)									FAUT	21.23 (11)
S'	21.23 (11)	ÉTUDIANTE	21.23 (11)									L'	21.23 (11)
TROUVE	21.23 (11)											TROIS	21.23 (11)
												JUSQU'	21.23 (11)

#### 7.1.1.4. *ne*

否定辞の *ne* は学習者によって過剰使用されていた。また、*ne* と共起された調整頻度 20 回以上の語の数や種類も学習者の方が母語話者よりも多いことが表 60 より分かる。実際に、話し言葉フランス語では否定辞 *ne* の省略が頻繁に見られる。そこで、否定辞 *ne* と最も多く使用され、*ne* が省略されていても否定のマーカ―として機能する副詞 *pas* について、調査を行う。また、*ne* の後ろに母音が続いた場合 *n'* となることより、*n'* についても同様の調査をする。まず、両コーパス中の副詞 *pas* を中心語として、左側の位置に *ne* が置かれる場合の頻度について調べ、*pas* と *ne* との共起強度について対数尤度比を使用して比較した。

表 60 *n(e)*と *pas* の共起強度

	母語話者コーパス		学習者コーパス	
	生起頻度	共起強度	生起頻度	共起強度
<i>ne + pas</i>	48	321.98	166	1482.14
<i>n' + pas</i>	138	1103.64	150	1358.34

共起頻度の値がいずれも学習者コーパスの方が高いことより、学習者の方が副詞 *pas* とともに否定辞 *n(e)* を使用していることが分かる。次に、*pas* が否定辞 *ne* を伴う否定文と、否定辞を伴わない否定文の生起頻度に有意差があるかどうかを調べた。その結果、否定辞ありと否定辞なしの生起頻度について、有意差があることが分かった。

表 61 否定辞 *ne* の保持と省略

	母語話者	学習者	$\chi^2$ 統計量、有意確率
否定辞あり	186	316	$\chi^2=434.41, p<.001$
否定辞なし	509	11	

以上より、学習者は否定文において母語話者よりも否定辞 *n(e)* を保持し、母語話者はこれを脱落させる傾向にあることがわかる。

表 62 学習者による *ne* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
DESOLÉE	38.68 (11)	MAIS	119.57 (34)	JE	583.77 (166)	NE	636.52 (181)	PEUX	137.15 (39)	PAS	531.02 (151)	ALLER	63.30 (18)
MOI	35.17 (10)	NON	63.30 (18)					VEUX	101.98 (29)	AI	21.10 (6)	PAS	52.75 (15)
SUIS	28.13 (8)	ET	35.17 (10)					COMPRENDS	45.72 (13)			LE	38.68 (11)
		DONC	24.62 (7)					VOUDRAIS	31.65 (9)			FAIRE	35.17 (10)
		MOI	21.10 (6)					SAIS	31.65 (9)			DE	31.65 (9)
								PEUT	28.13 (8)			CE	21.10 (6)
								FAIS	28.13 (8)			ÉTUDIER	21.10 (6)
								VOIS	24.62 (7)			TE	21.10 (6)
								L'	21.10 (6)				



表 63 母語話者による *ne* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
—	—	—	—	JE	102.27 (53)	NE	142.79 (74)	FAIS	21.23 (11)	PAS	84.90 (44)	—	—

#### 7.1.1.5. *je*

*je* の過剰使用については、学習者が話す発話の全体量が母語話者よりも少ないことより、人称代名詞主語である *je* の過剰使用が引き起こされたということが考えられる。専攻研究で見たように、日本人英語学習者についても一人称代名詞主語の過剰使用は、スピーチコーパス分析から指摘されている点である (Ishikawa 2007)。使用頻度の差だけでなく、使用方法について差があるかどうかを調べるため、両コーパスで見られたそれぞれ上位 10 個の *je* に後続する動詞を比較した。

表 64 *je* に後続する動詞上位 10 位

順位	母語話者		学習者	
	コロケート	調整頻度 (頻度)	コロケート	調整頻度 (頻度)
1	<i>suis</i>	910.79 (472)	<i>suis</i>	1192.15 (339)
2	<i>vais</i>	302.95 (157)	<i>voudrais</i>	710.37 (202)
3	<i>prends</i>	169.81 (88)	<i>vais</i>	277.82 (79)
4	<i>regarde</i>	133.15 (69)	<i>bois</i>	196.93 (56)
5	<i>peux</i>	131.22 (68)	<i>peux</i>	165.28 (47)
6	<i>voudrais</i>	113.85 (59)	<i>veux</i>	161.77 (46)
7	<i>bois</i>	111.92 (58)	<i>prends</i>	151.22 (43)
8	<i>vis</i>	104.20 (54)	<i>m'appelle</i>	133.63 (38)
9	<i>sais</i>	100.34 (52)	<i>préfère</i>	126.60 (36)
10	<i>préfère</i>	98.41 (51)	<i>dois</i>	109.02 (31)

その結果、*je* に後続する上位 10 の動詞のうち 7 つは同じものであることが分かった。次に、人称代名詞主語の *je* と 7 つ動詞との共起頻度について *je* の生起頻度に対して有意差があるかどうか調べた。

表 65 *je* に後続する共通の動詞

コロケート	母語話者コーパス	学習者コーパス	$\chi^2$ 統計量、有意確率	母語話者=学習者
<i>suis</i>	910.79 (472)	1192.15 (339)	$\chi^2=2.67$ ; $p>.05$	母語話者=学習者
<i>vais</i>	302.95 (157)	277.82 (79)	$\chi^2=12.01$ ; $p<.001$	母語話者>学習者
<i>prends</i>	169.81 (88)	151.22 (43)	$\chi^2=7.26$ ; $p<.01$	母語話者>学習者
<i>peux</i>	131.22 (68)	165.28 (47)	$\chi^2=0.57$ ; $p>.05$	母語話者=学習者
<i>voudrais</i>	113.85 (59)	710.37 (202)	$\chi^2=124.92$ ; $p<.001$	母語話者<学習者
<i>bois</i>	111.92 (58)	196.93 (56)	$\chi^2=1.11$ ; $p>.05$	母語話者=学習者
<i>préfère</i>	98.41 (51)	126.60 (36)	$\chi^2=0.31$ ; $p>.05$	母語話者=学習者

その結果、*suis*、*peux*、*bois*、*préfère* については 5%水準以下で有意差が見られなかった。つまり、学習者も母語話者も人称代名詞 *je* を使用する際には、同じような割合でこれらの動詞を用いているということが分かる。有意差が見られた 3 つの動詞については、学習者の方が有意に多く使用していた動詞は *voudrais* のみで、*vais* と *prends* については母語話者の方が学習者よりも多く使用していた。

*voudrais* については既出の通り、学習者による使用方法が母語話者と異なることより、過剰使用が引き起こされたと考えられる。しかしそれ以外の動詞については、人称代名詞主語の *je* との共起頻度は母語話者の方が高いか、共起頻度に有意差がなかった。

また、どちらかのコーパスの上位 10 位に含まれていたが、もう一方のコーパスには含まれなかった動詞について *je* との共起頻度の比較を行ったところ、以下のような結果となった。

表 66 *je* に後続する共通しない動詞

	母語話者	学習者	$\chi^2$ 統計量、有意確率	
<i>regarde</i>	133.15 (69)	49.23 (14)	$\chi^2=25.92 ; p<.001$	母語話者>学習者
<i>vis</i>	104.20 (54)	24.62 (7)	$\chi^2=27.10 ; p<.001$	母語話者>学習者
<i>sais</i>	100.34 (52)	10.55 (3)	$\chi^2=34.17 ; p<.001$	母語話者>学習者
<i>veux</i>	23.16 (12)	161.77 (46)	$\chi^2=29.28 ; p<.001$	母語話者<学習者
<i>m'appelle</i>	65.54 (35)	133.63 (38)	$\chi^2=1.81 ; p>.05$	母語話者=学習者
<i>dois</i>	28.94 (15)	109.02 (31)	$\chi^2=10.07 ; p<.01$	母語話者<学習者

これらについて、母語話者によって有意に多く *je* と共に用いられた動詞は *regarde*、*vis*、*sais* であり、*m'appelle* については有意差が見られなかった。学習者によって有意に多く使用されていた動詞は、*veux* と *dois* であった。

以上を総合すると、*je* と共起された頻度上位 10 位の動詞について学習者の方が有意に多く使用していたものは、*voudrais*、*veux* そして *dois* のみであった。

*je dois*+動詞という MWU は *il faut*+動詞、もしくは *il faut que je*+動詞に置き換えて使用することもできる。母語話者は *je dois* ではなく *il faut* を使用したことによって、人称代名詞主語 *je* の頻度が低くなったことも考えられる。これより、*je dois* の頻度と *je dois* の意味として使用されている *il faut* の頻度とを比較したところ、有意差が見られた。学習者は DEVOIR を使用するのに対して、母語話者は FALLOIR を好むというという結果は、先行研究の Debrock et al. (1999) の結果を支持するものであった。

表 67 *je dois* と *il faut que je* の頻度

	母語話者	学習者	$\chi^2$ 統計量、有意確率
<i>je dois</i>	15	31	$\chi^2=11.36 ; p<.001$
<i>il faut que je</i> +動詞	7	0	

以上から、母語話者コーパスと学習者コーパスとで *je* の生起頻度は学習者の方が有意に高いが、*je* と共起される動詞については、動詞の種類は似た傾向にある上、頻度も有意差があまり見られないことが分かった。一方、頻度に有意差が見られた動詞については、学習者に特有の使用方法が見られたため、*je* の使用頻度が高くなった一因となった可能性は高い。しかし、この少数の動詞の影響のみで *je* の使用頻度差がこれほど大きくなったとは考えにくい。実際、学習者コーパスの総語数と比較して、母語話者コーパスの総語数は 2 倍近くある。学習者の一文の長さは母語話者と比べて短いことより、主語を省略できないフランス語において学習者の *je* の生起頻度が相対的に高くなったことが一番の原因であると考えられる。

表 68 学習者による *je* の共起語

L3	L2	L1	KW	R1	R2	R3							
APPELLE	105.50 (30)	PARCE	225.07 (64)	QUE	309.47 (88)	JE	5580.95 (1587)	SUIS	1192.15 (339)	DU	263.75 (75)	PAS	935.43 (266)
NP	101.98 (29)	NP	203.97 (58)	MAIS	298.92 (85)			VOUDRAIS	710.37 (202)	AIME	253.20 (72)	À	407.93 (116)
À	101.98 (29)	EN	98.47 (28)	ET	281.33 (80)			NE	583.77 (166)	ALLER	249.68 (71)	NP	379.8 (108)
JE	94.95 (27)	CE	91.43 (26)	NON	281.33 (80)			N'	467.72 (133)	AI	228.58 (65)	J'	175.83 (50)
LE	94.95 (27)	À	91.43 (26)	NP	270.78 (77)			VAIS	277.82 (79)	DE	189.90 (54)	L'	175.83 (50)
SUIS	94.95 (27)	EXCUSEZ	87.92 (25)	MOI	116.05 (33)			BOIS	196.93 (56)	LE	189.90 (54)	LE	144.18 (41)
DE	70.33 (20)	LE	84.4 (24)	ANS	84.4 (24)			M'	165.28 (47)	ALLÉE	172.32 (49)	AU	144.18 (41)
LA	70.33 (20)	SUIS	63.30 (18)	WEEKEND	84.4 (24)			PEUX	165.28 (47)	JAPONAISE	161.77 (46)	UN	140.67 (40)
EST	49.23 (14)	MOI	63.30 (18)	DONC	66.82 (19)			VEUX	161.77 (46)	À	158.25 (45)	MAIS	123.08 (35)
DIX	42.20 (12)	LA	59.78 (17)	JE	52.75 (15)			PRENDS	151.22 (43)	ÉTUDIANTE	158.25 (45)	LA	116.05 (33)
EN	42.20 (12)	D'	59.78 (17)	PUIS	45.72 (13)			PRÉFÈRE	126.6 (36)	PEUX	140.67 (40)	DES	94.95 (27)
ALLER	38.68 (11)	DE	56.27 (16)	HESIT	35.17 (10)			DOIS	109.02 (31)	APPELLE	137.15 (39)	THÉ	84.4 (24)
LIBRE	35.17 (10)	DÉSOLÉE	56.27 (16)	OUI	35.17 (10)			VIENS	94.95 (27)	DÉSOLÉE	112.53 (32)	BEAUCOUP	80.88 (23)
PAS	35.17 (10)	AU	52.75 (15)	QUAND	31.65 (9)			VOIS	87.92 (25)	UN	105.50 (30)	DE	73.85 (21)
TRÈS	35.17 (10)	JE	49.23 (14)	GÉNÉRAL	31.65 (9)			MANGE	87.92 (25)	VEUX	105.50 (30)	DU	70.33 (20)
DES	31.65 (9)	ET	49.23 (14)	RESTAURANT	28.13 (8)			SOUVENT	77.37 (22)	FAIRE	101.98 (29)	CAFÉ	63.30 (18)
ET	31.65 (9)	EST	38.68 (11)	ABORD	28.13 (8)			ME	77.37 (22)	L'	101.98 (29)	PAIN	56.27 (16)
L'	31.65 (9)	NEUF	38.68 (11)	MAISON	24.62 (7)			DÉTESTE	73.85 (21)	LA	91.43 (26)	ET	52.75 (15)
AI	28.13 (8)	ANS	38.68 (11)	BONJOUR	21.10 (6)			FAIS	49.23 (14)	RESTER	80.88 (23)	OFFICIEL	49.23 (14)
AVEC	28.13 (8)	LES	35.17 (10)	SOIR	21.10 (6)			REGARDE	49.23 (14)	SOUVENT	77.37 (22)	POUR	49.23 (14)
C'	28.13 (8)	VINGT	35.17 (10)	DIMANCHE	21.10 (6)			L'	45.72 (13)	RÉSERVER	63.30 (18)	CHEZ	45.72 (13)
EXCUSEZ	28.13 (8)	DU	31.65 (9)					VOUS	35.17 (10)	SORTIR	59.78 (17)	AVEC	42.20 (12)

L3	L2	L1	KW	R1	R2	R3			
DU	24.62 (7)	WEEKEND	28.13 (8)	LIS	31.65 (9)	JAPONAIS	59.78 (17)	AI	38.68 (11)
FAIRE	24.62 (7)	MAISON	24.62 (7)	CHERCHE	31.65 (9)	ÉTUDIANT	59.78 (17)	BILLET	35.17 (10)
MAISON	24.62 (7)	QUE	24.62 (7)	PENSE	31.65 (9)	ACHETER	59.78 (17)	CAMPAGNE	31.65 (9)
MOI	24.62 (7)	MON	21.10 (6)	CHANGE	28.13 (8)	DES	56.27 (16)	TRAIN	31.65 (9)
CHEZ	21.10 (6)			VIS	24.62 (7)	PRENDRE	52.75 (15)	LES	31.65 (9)
NEUF	21.10 (6)			TE	24.62 (7)	COMPRENDS	45.72 (13)	VOIR	31.65 (9)
PARCE	21.10 (6)					TRÈS	45.72 (13)	RIZ	28.13 (8)
PENDANT	21.10 (6)					SENS	45.72 (13)	FILM	28.13 (8)
						LES	45.72 (13)	ALLER	28.13 (8)
						ALLÉ	42.20 (12)	UNE	28.13 (8)
						QUE	42.20 (12)	EN	28.13 (8)
						NP	42.20 (12)	CE	28.13 (8)
						TRAVAILLER	42.20 (12)	DEPUIS	24.62 (7)
						ÉTUDIER	42.20 (12)	TRISTE	24.62 (7)
						VOIS	42.20 (12)	LIGNE	21.10 (6)
						ME	42.20 (12)	QUE	21.10 (6)
						TRISTE	38.68 (11)	MES	21.10 (6)
						INQUIÈTE	35.17 (10)		
						VOIR	35.17 (10)		
						VOUDRAIS	35.17 (10)		
						BOIS	35.17 (10)		
						FAIS	35.17 (10)		
						DANS	31.65 (9)		
						UNE	31.65		

L3	L2	L1	KW	R1	R2	(9)	R3
					MANGE	31.65 (9)	
					TOUJOURS	31.65 (9)	
					SAIS	31.65 (9)	
					EN	24.62 (7)	
					REGARDE	24.62 (7)	
					TE	24.62 (7)	
					MANGER	24.62 (7)	
					PEUT	21.10 (6)	
					ÊTRE	21.10 (6)	
					NÉ	21.10 (6)	
					NÉE	21.10 (6)	
					PARTIR	21.10 (6)	
					PASSER	21.10 (6)	
					RESTÉE	21.10 (6)	
					BEAUCOUP	21.10 (6)	
					SUIS	21.10 (6)	
					CONNAIS	21.10 (6)	

表 69 母語話者による *je* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
JE	133.15 (69)	NP	162.09 (84)	QUE	362.77 (188)	JE	3849.64 (1995)	SUIS	910.79 (472)	PAS	256.64 (133)	PAS	372.42 (193)
À	104.2 (54)	CE	140.86 (73)	HESIT	256.64 (133)			VAIS	302.95 (157)	AI	175.6 (91)	À	223.84 (116)
ET	96.48 (50)	ET	119.64 (62)	NP	142.79 (74)			N'	262.43 (136)	LE	138.93 (72)	NP	165.95 (86)
EST	86.83 (45)	EN	88.76 (46)	ET	133.15 (69)			PRENDS	169.81 (88)	AIME	123.5 (64)	J'	104.2 (54)
NP	86.83 (45)	PARCE	88.76 (46)	DONC	113.85 (59)			REGARDE	133.15 (69)	À	121.57 (63)	L'	86.83 (45)
DE	69.47 (36)	HESIT	69.47 (36)	MOI	111.92 (58)			PEUX	131.22 (68)	UN	121.57 (63)	LE	79.12 (41)
LA	63.68 (33)	JE	67.54 (35)	NON	96.48 (50)			VOUDRAIS	113.85 (59)	LA	115.78 (60)	LA	77.19 (40)
HESIT	61.75 (32)	LE	65.61 (34)	MAIS	88.76 (46)			BOIS	111.92 (58)	DE	104.2 (54)	UN	69.47 (36)
EN	50.17 (26)	DE	59.82 (31)	QUAND	86.83 (45)			VIS	104.2 (54)	APPELLE	75.26 (39)	HESIT	65.61 (34)
APPELLE	48.24 (25)	À	57.89 (30)	JE	67.54 (35)			NE	102.27 (53)	FRANÇAISE	75.26 (39)	DE	61.75 (32)
AU	46.31 (24)	L'	50.17 (26)	BONJOUR	63.68 (33)			SAIS	100.34 (52)	QUE	73.33 (38)	EN	55.96 (29)
LE	46.31 (24)	MOI	50.17 (26)	ANS	61.75 (32)			PRÉFÈRE	98.41 (51)	DU	63.68 (33)	CHEZ	50.17 (26)
SUIS	42.45 (22)	UN	44.38 (23)	SI	46.31 (24)			M'	81.05 (42)	DES	63.68 (33)	ET	50.17 (26)
PAS	36.66 (19)	LA	40.52 (21)	OUI	38.59 (20)			FAIS	75.26 (39)	NÉE	57.89 (30)	POUR	46.31 (24)
EXCUSEZ	32.8 (17)	SUIS	40.52 (21)	ENSUITE	36.66 (19)			ME	69.47 (36)	ÉTUDIANTE	54.03 (28)	VILLE	40.52 (21)
CE	30.87 (16)	MAIS	40.52 (21)	PAS	34.73 (18)			PENSE	63.68 (33)	SUIS	50.17 (26)	DU	38.59 (20)
MAIS	30.87 (16)	CHEZ	38.59 (20)	GÉNÉRAL	34.73 (18)			POURRAIS	46.31 (24)	ALLER	48.24 (25)	BEAUCOUP	38.59 (20)
VAIS	30.87 (16)	AU	36.66 (19)	WEEKEND	32.8 (17)			TRAVAILLE	42.45 (22)	L'	46.31 (24)	MON	38.59 (20)
AI	27.02 (14)	EXCUSEZ	32.8 (17)	PUIS	28.94 (15)			SERAIS	34.73 (18)	EN	46.31 (24)	PEU	36.66 (19)
J'	23.16 (12)	DONC	32.8 (17)	UNIVERSITÉ	27.02 (14)			TE	34.73 (18)	HESIT	44.38 (23)	MAIS	34.73 (18)
PARCE	23.16 (12)	QUE	32.8 (17)	ALORS	25.09 (13)			MANGE	30.87 (16)	ALLÉE	40.52 (21)	OFFICIEL	32.8 (17)
DES	21.23 (11)	PAS	27.02 (14)	RESTAURANT	25.09 (13)			DOIS	28.94 (15)	AU	40.52 (21)	TRAM	30.87 (16)
IL	21.23 (11)	TRÈS	25.09 (13)	POURRAIS	23.16 (12)			T'	27.02 (14)	BEAUCOUP	36.66 (19)	SI	30.87 (16)
		PETIT	23.16 (12)	DÉJEUNER	23.16 (12)			L'	25.09 (13)	VRAIMENT	36.66 (19)	LES	28.94 (15)

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
		DES	21.23 (11)	LÀ	23.16 (12)			CHERCHE	23.16 (12)	RENTRE	34.73 (18)	PLUS	28.94 (15)
		BIEN	21.23 (11)	APRÈS	21.23 (11)			METS	23.16 (12)	DÉSOLÉE	34.73 (18)	FILMS	27.02 (14)
		MÊME	21.23 (11)	SINON	21.23 (11)			VEUX	23.16 (12)	TRÈS	32.8 (17)	LIGNE	27.02 (14)
								CROIS	23.16 (12)	FAIRE	32.8 (17)	RESTAURANT	27.02 (14)
								LIS	23.16 (12)	CÉLIBATAIRE	30.87 (16)	CAMPAGNE	25.09 (13)
								VOULAIS	21.23 (11)	DANS	28.94 (15)	DES	25.09 (13)
										FRANÇAIS	28.94 (15)	MOI	23.16 (12)
										NÉ	28.94 (15)	TRÈS	23.16 (12)
										FAIS	28.94 (15)	AVEC	23.16 (12)
										AVEC	27.02 (14)	C'	23.16 (12)
										ÉTUDIANT	27.02 (14)	CE	23.16 (12)
										DÉSOLÉ	25.09 (13)	MES	23.16 (12)
										ACTUELLEMENT	25.09 (13)	CAFÉ	21.23 (11)
										AVOIR	25.09 (13)	NATIONALITÉ	21.23 (11)
										CHEZ	23.16 (12)		
										ME	21.23 (11)		
										VOIS	21.23 (11)		

7.1.1.6. *prenez*

学習者の *prenez* の使用頻度は母語話者よりも多いことが特徴後分析によって明らかになったが、この語は学習者コーパスでは主に人称代名詞主語の *vous* が L1 の位置に多く現れていた。*vous prenez* は「あなたは(乗り物に)乗る」という意味も持ち、実際に移動方法の説明でこの MWUs が用いられていた。主に、タスク j009 と j035 でこの MWUs が多く使用されていた。

j009	あなたは旅行者から成田空港で、上野駅までの行き方を尋ねられました。 以下の目的地への行き方と所要時間を参照して説明してください。 成田 → 日暮里 → 上野 京成線 山手線 (1時間10分) (5分)		
j035	Gares de départ et d'arrivée	Horaires	Trains
	PARIS	14h40	TGV
	BORDEAUX	17h39	
	BORDEAUX	18h04	TGV
	TOULOUSE	20h12	
	表を参考にして、Paris から Toulouse までの道順と出発・到着時間を説明してください。		

一方、母語話者は同タスクにおいて、*il faut prendre*(～に乗らなくてはならない)という MWUs を使用する傾向にあった。つまり、母語話者は人称代名詞主語を用いずに非人称主語の方を用いている。学習者コーパスにはこの MWUs は一回も生起していない。このことは、学習者が人称代名詞主語を用いる構文を好むという裏付けにもなる。

表 70 タスク j009 と j033 に見られる MWUs

タスク番号	日本人学習者		フランス語母語話者	
	MWUs	生起回数	MWUs	生起回数
j009	vous prenez	15	il faut prendre	22
	prenez	4	je prends	7
	on prend	2	on prend	2
j035	vous prenez	15	il faut prendre	18
	on prend	9	on prend	8
	prenez	2	je prends	5
	tu prends	2		

表 71 学習者による *prenez* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
D'	52.75 (15)	ABORD	52.75 (15)	VOUS	158.25 (45)	PRENEZ	221.55 (63)	LE	133.63 (38)	NP	112.53 (32)	NP	73.85 (21)
		NP	35.17 (10)	ET	21.10 (6)			LA	59.78 (17)	LIGNE	77.37 (22)	DE	52.75 (15)
												LIGNE	28.13 (8)



表 72 母語話者による *prenez* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	-	-	prenez	13.51 (7)	-	-	-	-	-	-

#### 7.1.1.7. japonaise

学習者によって使用された *japonaise* を含む MWUs には、*je suis japonaise*(私は日本人です)という MWUs が 20 例含まれていることが分かる。これはもちろん言語要因による差ではなく国籍の差であるので、まず国籍の表現方法について比較を行う。学習者コーパスとフランス語母語話者コーパスの国籍の紹介パターンは、以下の通りである(表 73)。

表 73 国籍紹介のパターン

日本人学習者		フランス語母語話者	
MWUs	生起回数	MWUs	生起回数
<i>je suis japonais(e)</i>	63	<i>je suis français(e)</i>	47
* <i>je suis de nationalité Japon</i> <sup>54</sup>	1	<i>je suis de nationalité française</i>	9
* <i>mon nationalité est japonais</i>	1	<i>je suis nationalité française</i>	4
* <i>mon nationalité est Japon</i>	1		

日本人学習者は *je suis japonais(e)* というパターンで固定されており、その他のパターンはあまり見られなかった。フランス語母語話者コーパスには *je suis français(e)* という MWUs がほとんどを占めるものの、*nationalité* を使用した例も多く見られた。一方、学習者コーパス内には *nationalité* を使用した例がいくつか見られたが、いずれも正しい使用ではなかった。

また、学習者コーパスで最も多く見られた MWU は、*elle est japonaise* で、38 例見られた。そのうちタスク j025 で 32 例が使用されており、レンジは 82.05%にものぼる。母語話者が同タスクで国籍を紹介する際に使用した MWUs のうち、*elle est japonaise* という MWUs は 9 例のみであり、レンジは 23.07%であった。これは、このタスクの国籍紹介で母語話者によって最も多く使用された MWUs であるが、母語話者が同タスクで使用した MWUs は他にも多くの種類が見られたことを意味している。

<sup>54</sup> 以下アスタリスクがついているものは非文とする。

表 74 母語話者による国籍紹介のパターン

国籍紹介 MWUs	
elle est japonaise	8
c'est une japonaise	5
主語 + ÊTRE + de nationalité japonaise	4
NP de nationalité japonaise	2
NP qui est japonaise	2
ÊTRE + d'origine japonaise	2

表 75 学習者による *japonaise* の共起語

L1	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
NP	186.38 (53)	JE	161.77 (46)	SUIS	161.77 (46)	JAPONAISE	327.05 (93)	ELLE	144.18 (41)	A	126.60 (36)	TRENTE	98.47 (28)
		ELLE	130.12 (37)	EST	133.63 (38)			J	52.75 (15)	AI	52.75 (15)	DIX	38.68 (11)
								ET	24.62 (7)			VINGT	31.65 (9)

表 76 母語話者による *japonaise* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
NP	27.02 (14)	—	—	EST	23.16 (12)	JAPONAISE	77.19 (40)	ELLE	30.87 (16)	A	34.73 (18)	TRENTE	40.52 (21)

#### 7.1.1.8. vous

両コーパスとも *vous* が含まれる MWUs として *s'il vous plaît* (お願いします) が特徴的であると言える。つまり、この使用における *vous* は *s'il vous plaît* という決まり文句の一部として使用されており、人称代名詞主語としてではなく、間接目的語として使用されている *vous* である。そこで、*s'il vous plaît* という MWU 内で使用される *vous* の頻度とそれ以外の頻度についてまとめたところ、以下ようになった。

表 77 *s'il vous plaît* の頻度

	母語話者	学習者
<i>s'il vous plaît</i>	173	70
それ以外	103	227

以上について頻度差の検定を行ったところ、有意差が見られた ( $\chi^2=89.60, p<.001$ )。つまり、母語話者は学習者よりも *vous* を *s'il vous plaît* という MWUs 内で使用していることが分かる。一方、学習者は *vous* の後ろに動詞が後続していることから、*vous* を人称代名詞主語として使用している傾向があると言える。そこで、*vous* 全体の使用に対して人称代名詞主語として *vous* が用いられる割合を学習者と母語話者とで比較したところ、母語話者よりも学習者の方が人称代名詞としての *vous* を有意に多く使用していることが分かった。

表 78 人称代名詞としての *vous* の頻度

	母語話者	学習者	$\chi^2$ 統計量、有意確率
人称代名詞主語の <i>vous</i>	76	192	$\chi^2=145.82$ ; $p<.001$

また、母語話者が人称代名詞主語として *vous* を用いる場合、その動詞の多くは *pourriez* と *pouvez* であり、これは依頼表現の MWUs であると言える。依頼表現はタスク f017、f018、f047 で使用されていた。これらのタスクは、相手から情報を引き出すタスクであると言える。：

f017	Vous êtes dans la rue, et demandez à un passant où se trouve le magasin "bio Natura".
f018	Au restaurant, vous voyez un plat que vous ne connaissez pas. Comment vous-informeriez vous auprès du serveur ?
f047	Vous êtes au kiosque à Paris. Vous voulez acheter "l'officiel des spectacles". Que dites-vous ?

学習者コーパスと母語話者コーパスでこれらのタスクにおける依頼表現の比較を行った。依頼表現で *vous* と **POUVOIR** が使用される場合は疑問文となるが、イントネーション型、EST-CE QUE 型疑問文そして倒置型疑問文の頻度は分けずにカウントした。

表 79 POUVOIR の法による頻度の違い

		母語話者	学習者
pouvez	me dire	6	0
	m'indiquer	3	7
	m'expliquer	2	3
	me renseigner	2	0
	me donner	0	3
	expliquer	0	2
	計	13	15
pourriez	me dire	9	0
	m'indiquer	7	0
	me décrire	2	0
	m'expliquer	1	0
	me renseigner	1	0
	me donner	0	1
	expliquer	0	1
計	19	2	

まず、学習者と母語話者との間に、直説法である *pouvez* と条件法である *pourriez* の使用頻度について有意差があるかどうかを調べたところ、有意差が確認された ( $\chi^2=10.28$ ,  $p<.001$ )。内訳を見てみると、母語話者は依頼表現を使用するときに条件法を用いる傾向にあるが、学習者は直説法を用いていることが分かる。一般的に依頼表現として条件法を用いる方が直説法を用いるよりも丁寧さが増す。したがって、学習者の発話は母語話者と比べてポライトネスが欠けていると言える。

また、依頼表現とともに使用される動詞(句)にも差が見られた。母語話者に見られて学習者に見られなかった動詞(句)は、*me dire* と *me renseigner* (それぞれ母語話者 15 例と

3例)であり、その反対は *me donner* と *expliquer* (それぞれ学習者 4例と 3例)であった。実際に、POUVOIR *me donner* と POUVOIR *expliquer* はこのコンテキストでは非文となる。



### 7.1.1.9. *peux*

学習者コーパスにも母語話者コーパスにも *peux* の左側には、人称代名詞主語の *tu* と *je* が用いられていた。しかしそれらの頻度を比較してみると、学習者は *tu* と *je* と頻度がほぼ同じであるのに対して、母語話者は *je* の頻度が高いことが分かる。*tu peux* という MWUs が使用されたのは、主に以下のタスクであった。

j003	あなたは、今日携帯を忘れてしまいました。携帯を忘れたことを伝え、それを貸してくれるように友達に頼むときには何といいますか？
------	---

人称代名詞主語 *tu* と *peux* が同時に用いられると、「～してくれますか？」という依頼の意味も持つ。既出の通り、依頼表現を使用する際に、直説法と条件法を用いることが考えられ、主語が *vous* の場合母語話者は条件法を好むが、学習者は直説法を使用する傾向にあるということが分かった。そこで、このタスクにおいて主語が *tu* の場合も同じ傾向を持つのかどうかを比較した。今回の分析にも、イントネーション型、EST-CE QUE 型疑問文そして倒置型疑問文の頻度は分けずにカウントした。

表 82 *tu peux* と *tu pourrais* の頻度の比較

	母語話者	学習者
<i>tu peux</i>	13	21
<i>tu pourrais</i>	16	1

上記の直説法と条件法の使用頻度について、母語話者と学習者に有意差が見られた ( $\chi^2=14.43, p<.001$ )。話者間の関係が近い時に用いられる *tu* が主語の場合でも、母語話者は条件法である *pourrais* を好むが、学習者は *peux* を使用する傾向にあると言える。

また、学習者による *peux* の使用について、特徴的なものも見られた。たとえば、タスク j006 や j032 のような友達を誘うタスクにおいて *tu peux* という MWUs が使用されていた。

j006	今晚レストランで夕食を一緒に食べに行こうと、友人に誘ってください。
j032	あなたはこの週末一緒に出かけられるかどうかを聞こうと、友人に電話をしています。ところが、留守番電話に接続されました。「いつ」、「どこで」、「どんなことをしたいのか」を留守番電話に吹き込み、友人を誘ってください。

**tu peux** aller au restaurant ce soir avec moi (JT24 ; j006)

allô samedi soir je voudrais aller au restaurant mexicain avec toi à NP HESIT **tu peux** aller avec moi (JT28 ; j032)

このような、相手を誘うという状況において *tu peux* が使用された例は、学習者コーパスでは 15 例見られたのに対して、母語話者コーパスでは一例も見られなかった。これは、

*tu peux* を使用することによって、勧誘の機能というよりは依頼の機能を持つことになることに起因していると考えられる。一方、母語話者はこれらタスクにおいて *ça te dit* (25 例) や *ça te dirait* (9 例) という『～はどうですか、～しませんか』という勧誘を意味する MWUs を主に用いていた。

また、*peux* の L1 の位置には母語話者コーパスにも学習者コーパスにも否定文で用いられる副詞 *pas* が第一位に見られる。学習者コーパスの R1 の位置には否定辞 *ne* が見られるものの、母語話者コーパスにはこの否定辞は見られなかった。そこで、否定辞 *ne* を含む MWUs である主語 + *ne peux pas* と否定辞を含まない MWUs である主語 + *peux pas* の出現頻度を比較した。

表 83 否定辞 *ne* の保持と脱落

	母語話者コーパス	学習者コーパス
主語 + <i>ne peux pas</i>	5	37
主語 + <i>peux pas</i>	35	1

否定辞の使用頻度についてカイ二乗検定を行って有意差を調べたところ、0.1%水準で有意差が認められた ( $\chi^2=56.48$ )。

表 84 学習者による *peux* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
PARCE	66.82 (19)	JE	140.67 (40)	TU	168.80 (48)	PEUX	488.82 (139)	PAS	133.63 (38)	PRÊTER	56.27 (16)	TON	38.68 (11)
MAIS	42.2 (12)	QUE	94.95 (27)	JE	165.28 (47)			ME	73.85 (21)	ALLER	42.2 (12)	MOI	24.62 (7)
MON	31.65 (9)	PORTABLE	31.65 (9)	NE	137.15 (39)			ALLER	28.13 (8)	LE	35.17 (10)		
CE	31.65 (9)							FAIRE	21.10 (6)	AVEC	21.10 (6)		
QUE	21.10 (6)									BEAUCOUP	21.10 (6)		

表 85 母語話者による *peux* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
CE	25.09 (13)	QUE	44.38 (23)	JE	131.22 (68)	PEUX	171.74 (89)	PAS	77.19 (40)	TE	21.23 (11)	LE	28.94 (15)
				TU	28.94 (15)			ME	23.16 (12)				



### 7.1.1.1 0. *abord*

学習者コーパスとも母語話者コーパスとも、*abord* という名詞は *d'* とともに用いられており、*d'abord* という「まず、はじめに」を意味する副詞として使用されていた。しかし、両コーパスで使用頻度の差は大きく見られた。

学習者コーパスでは、46 例中 44 例が、母語話者コーパス中では 7 例中 6 例がタスク j008 と j009 で *d'abord* の使用が見られた。

j008	あなたの家から学校までの行き方と、それにかかる時間を説明してください。
j009	あなたは旅行者から成田空港で、上野駅までの行き方を尋ねられました。 以下の目的地への行き方と所要時間を参照して説明してください。 成田 → 日暮里 → 上野 京成線 山手線 (1時間10分) (5分)

学習者はこれらのタスクにおいて順序を表すこの副詞をより多く用いていたが、母語話者は *d'abord* 以外にも順序を明確に示す指標となる語をあまり用いていない。その代わりに状況補語である前置詞句を発話単位の初めに置くことによって、その後の文の内容の枠組みを示していると考えられる。

**pour aller à l'université** je dois prendre le bus neuf INT qui est à l'arrêt en face de chez moi puis je dois m'arrêter à NP (FT18; f008)

**pour aller à NP** il faut prendre la ligne NP pendant une heure et de HESIT une heure dix environ jusqu'à NP (FT20; f009)

使用頻度の差もさることながら、*d'abord* の出現位置も注目に値する。学習者は 46 例中 40 例について、*d'abord* は発話単位の頭で使用されているが、母語話者は全て発話単位の中で使用されていた。

pour aller de chez moi à la fac il faut **d'abord** sortir de moi à pied ça pour arriver au à l'arrêt de tram (FT31 ; f008)

pour aller de NP à NP il faut **d'abord** prendre la ligne NP à NP jusqu'à NP (FT26 ; f008)

**d'abord** je vais à la gare de NP en bicyclette (JT22 ; j008)

**d'abord** vous prenez le NP ligne à la gare de NP et prenez le NP ligne de NP à NP (JT35 ; j009)

表 86 学習者による *abord* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	D'	161.77 (46)	ABORD	161.77 (46)	VOUS	77.37 (22)	PRENEZ	52.75 (15)	LE	49.23 (14)
								JE	28.13 (8)	VAIS	21.10 (6)	À	42.20 (12)

表 87 母語話者による *abord* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
—	—	—	—	—	—	ABORD	13.51 (7)	—	—	—	—	—	—

#### 7.1.1.1 1. *football*

この語はタスクに大きく依存する語であると言える。*football* が現れたのは、以下のタスクである。

j033	あなたは、友人からサッカーの試合を見に行こうと誘われました。 あなたはサッカーが好きではないので誘いを断りたいと思っています。 理由を添えて、誘いを断ってください。
f033	Un(e) ami(e) vous propose d'aller voir un match de foot. Vous ne voulez pas y aller, parce que vous n'aimez pas le foot. Refusez sa proposition et trouvez des excuses.

したがって、発話数が多いフランス語母語話者がこの語を使用する割合は学習者と比べて低くなるのは当然である。しかし、生起頻度自体を見てみても、両グループ間に大きな隔たりがあることが分かる。これは、指示文で *football* ではなく、*foot* が使用されていた可能性が考えられる。そこで、フランス語母語話者がこのタスクを行った際に、どちらの語を使用していたのかを調べた。

表 88 *football* と *foot* の生起頻度

	母語話者	学習者
<i>football</i>	2	35
<i>foot</i>	36	15

*football* と *foot* の使用頻度について有意差検定を行ったところ、0.1%水準で有意差が認められた。学習者は母語話者に比べて、省略ありの *foot* よりも省略なしの *football* を多用している。

表 89 学習者による *football* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
AIME	31.65 (9)	PAS	28.13 (8)	LE	66.82 (19)	FOOTBALL	123.08 (35)	—	—	—	—	—	—
PAS	21.10 (6)	MATCH	24.62 (7)	DE	28.13 (8)								

表 90 母語話者による *football* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
—	—	—	—	—	—	FOOTBALL	3.86 (2)	—	—	—	—	—	—

### 7.1.1.1 2. université

この語もまた、タスクからの影響を受けた語であると考えられる。この語が使用されたタスクは、以下のタスクである。

j008	あなたの家から学校までの行き方と、それにかかる時間を説明してください。
f008	Expliquez comment aller à l'université de chez vous et dites combien de temps il faut pour arriver. Indiquez les étapes.

母語話者によってこのタスクで主に用いられていた *université* と同じ意味を持つ単語は、*fac* であった。そこで、両コーパス全体におけるこれらの二語の使用頻度について比較を行ったところ、有意差が見られた ( $\chi^2=26.96$ ;  $p<.001$ )。

表 91 *université* と *fac* の使用頻度

	母語話者	学習者
<i>université</i>	46	89
<i>fac</i>	21	2

また注目すべき点は、*football* の場合とは異なり母語話者の指示文には *université* と明記されているのにも関わらず、母語話者は *fac* の方を好んで使用していた点である。なお、学習者コーパスに見られた 2 例は全て同じ学習者によって使用されていた (JT04)。一方、母語話者によるこの語の使用レンジは 33.33% である。よって、頻度の違いだけではなく、レンジの差も確認できた。

次に、両コーパスとも *aller à l'université* という MWUs が多く使用されていることが分かる。そこで、この MWUs の使用について比較を行った。学習者も母語話者も前置詞 *pour* とともに用いられる例が多く見られたが、出現位置が異なることが分かった。母語話者は前置詞句を発話単位の頭で使用する例がほとんどであるのに対して、学習者は発話単位中もしくは発話単位の終わりで使用しており、使用頻度に有意差が見られた ( $\chi^2=9.41$ ;  $p<.01$ )。母語話者は、*pour aller à l'université* を発話単位の頭に置くことによって、発話内容の枠組みを示しているものと考えられる。

表 92 *aller à l'université* の使用位置

	母語話者	学習者
単位の初め	12	1
単位中、単位の終わり	1	4

**pour aller à l'université** je prends le tram il me faut environ HESIT trente minutes (FT10; f008)

**pour aller à l'université** c'est pas très compliqué à vrai dire j'habite à cinquante mètres de l'université (FT10; f008)

je je mets HESIT dix minutes pour **aller à l'université** je dois traverser HESIT un chemin puis HESIT je marche le long d'une route et j'arrive (FT06; f008)

**pour aller à l'université** je prends le train (JT27 ; j008)  
 je mis deux heures **pour aller à l'université** (JT16 ; j008)  
 ça fait seulement cinq minutes **pour aller à l'université** (JT39 ; j008)

表 93 学習者による *université* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
VAIS	66.82 (19)	À	228.58 (65)	L'	267.27 (76)	UNIVERSITÉ	312.98 (89)	EN	52.75 (15)	VÉLO	35.17 (10)	IL	21.10 (6)
ALLER	56.27 (16)	DE	24.62 (7)					ET	31.65 (9)	NP	21.10 (6)		
FRANÇAIS	35.17 (10)							À	31.65 (9)				

表 94 母語話者による *université* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
ALLER	28.94 (15)	À	82.97 (43)	L'	88.76 (46)	UNIVERSITÉ	88.76 (46)	JE	27.02 (14)	—	—	—	—

### 7.1.2. 過少使用語との共起語

過剰使用語の分析と同様に、前節で抽出した過少使用語について、調整頻度 20 回以上共起した語を前後 3 語ずつ抽出し、これらの MWUs について個別に分析を行っていく。

#### 7.1.2.1. HESIT

学習者は圧倒的にためらい語の使用頻度が低く、既に指摘した点ではあるが、これにはタスク遂行時に学習者には準備する時間が母語話者よりも長く設けられていたことに起因するかもしれない。また、学習者の 1 つの発話単位に含まれる語数自体が少なく、一文が短いことより、ためらい語が使用されなかったという理由も考えられる。しかし、コーパスに基づく分析では因果関係を示すことはできないため、今後別の観点からもためらい語の使用頻度について分析を行う必要がある。

ためらい語の出現位置について比較を行ったところ、母語話者は発話中にためらい語を置くことも多い。ためらい語の HESIT が発話単位内部で行われているものを調べたところ以下の通りとなった。

表 95 ためらい語を含む MWUs

HESIT を含む MWUs	生起頻度
et + HESIT	107
parce que + HESIT	37
mais + HESIT	27
donc + HESIT	22
excusez-moi + HESIT	14
alors + HESIT	12
c'est + HESIT	11

*et*+HESI、*parce que*+HESIT、*mais*+HESIT、*donc*+HESIT のように、母語話者は接続詞(句)の後ろにためらい語を使用していることが多い。これは、発話単位内でためらい語が用いられていることを示している。学習者コーパスでも、発話単位内で現れた例で最も多かったためらい語の位置は接続詞の後ろで 57 例中 7 例あった。

また、母語話者コーパスの中にはためらい語の後ろに *voilà* (16 例) を置いている例が見られた。これは発話の流れに淀みができた際に発話を終了させたり、いったん話を仕切り直したりする際のマーカーとして用いられている例である。

on est comme des frères et **HESIT voilà** (FT07)

c'est très très calme c'est beau **HESIT voilà** (FT34)

je te prête mon portable tu me le rends HESIT quand quand tu ne HESIT quand quand on a plus besoin et **HESIT voilà** juste une précaution fait attention à pas le casser (FT17)

e vis à NP depuis le moins de septembre HESIT parce que avant j'étais chez mes parents et que et **HESIT voilà** je suis venue juste pour pour étudier donc au mois de septembre à la rentrée (FT33)

しかし、学習者コーパスにはこのような使用例が一例も見られなかった。

表 96 学習者による *HESIT* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
—	—	A	24.62 (7)	NP	45.72 (13)	HESIT	200.45 (57)	JE	35.17 (10)	—	—	—	—

表 97 母語話者による *HESIT* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
À	77.19 (40)	NP	102.27 (53)	NP	225.77 (117)	HESIT	2313.64 (1199)	JE	256.64 (133)	EST	167.88 (87)	DE	119.64 (62)
NP	71.4 (37)	DE	90.69 (47)	ET	206.47 (107)			J'	148.58 (77)	NP	96.48 (50)	JE	61.75 (32)
HESIT	65.61 (34)	À	67.54 (35)	QUE	73.33 (38)			C'	125.43 (65)	SUIS	69.47 (36)	À	61.75 (32)
JE	65.61 (34)	ET	65.61 (34)	MAIS	52.1 (27)			ON	86.83 (45)	JE	69.47 (36)	UN	50.17 (26)
DE	63.68 (33)	PARCE	63.68 (33)	HESIT	50.17 (26)			IL	81.05 (42)	AI	63.68 (33)	EST	48.24 (25)
J'	54.03 (28)	EST	54.03 (28)	MOI	42.45 (22)			À	67.54 (35)	J'	48.24 (25)	NP	48.24 (25)
ET	40.52 (21)	UN	54.03 (28)	DONC	42.45 (22)			LE	61.75 (32)	QUE	42.45 (22)	J'	48.24 (25)
EST	38.59 (20)	EN	50.17 (26)	PAS	30.87 (16)			ET	57.89 (30)	AIME	40.52 (21)	QUE	44.38 (23)
AIME	36.66 (19)	JE	44.38 (23)	EST	28.94 (15)			ELLE	48.24 (25)	LE	40.52 (21)	LE	38.59 (20)
C'	36.66 (19)	LA	42.45 (22)	ANS	27.02 (14)			DE	46.31 (24)	C'	34.73 (18)	PAS	38.59 (20)
LA	34.73 (18)	LE	40.52 (21)	WEEKEND	23.16 (12)			DES	46.31 (24)	IL	32.8 (17)	ET	36.66 (19)
EN	32.8 (17)	CE	40.52 (21)	ALORS	23.16 (12)			CE	46.31 (24)	A	28.94 (15)	EN	30.87 (16)
IL	32.8 (17)	L'	36.66 (19)					POUR	44.38 (23)	LA	28.94 (15)	A	28.94 (15)
PAS	32.8 (17)	AIME	34.73 (18)					EN	40.52 (21)	ÉTAIT	27.02 (14)	UNE	28.94 (15)
LE	28.94 (15)	SUIS	34.73 (18)					NP	38.59 (20)	Y	27.02 (14)	BIEN	28.94 (15)
UN	25.09 (13)	DES	32.8 (17)					UN	36.66 (19)	ON	25.09 (13)	AI	27.02 (14)
AVEC	23.16 (12)	HESIT	27.02 (14)					ÇA	34.73 (18)	DE	25.09 (13)	C'	27.02 (14)
DES	23.16 (12)	AU	27.02 (14)					MAIS	30.87 (16)	ÇA	25.09 (13)	DES	27.02 (14)
L'	23.16 (12)	EXCUSEZ	27.02 (14)					VOILÀ	30.87 (16)	L'	25.09 (13)	LA	25.09 (13)
POUR	23.16 (12)	C'	25.09 (13)					LES	30.87 (16)	SE	25.09 (13)	SUIS	23.16 (12)
A	21.23 (11)	PAS	25.09 (13)					DONC	27.02 (14)	À	23.16 (12)	CE	23.16 (12)
MON	21.23 (11)	D'	25.09 (13)					AU	25.09 (13)	FAUT	23.16 (12)	AU	23.16 (12)
ON	21.23 (11)	DU	25.09 (13)					SI	23.16 (12)	HABITE	21.23 (11)	IL	21.23 (11)
		DEUX	25.09 (13)					UNE	23.16 (12)	DES	21.23 (11)		
		POUR	23.16 (12)					PARCE	21.23 (11)	CE	21.23 (11)		
		MES	23.16 (12)					DU	21.23 (11)	ME	21.23 (11)		
		J	21.23 (11)										

### 7.1.2.2. *on*

*on* は特に話し言葉で *nous* と同じ機能として使用され、「私たちは」という意味で使われる。したがって、学習者による *on* の過少使用は、*on* ではなく *nous* を使用したこと  
に起因すると考えられる。そこで、主語として使用された *on* と *nous* の使用頻度について比較を行ったところ、以下のようになった。

表 98 *on* と *nous* の生起頻度

	母語話者	学習者
<i>on</i>	521	94
<i>nous</i>	12	38

有意差検定を行ったところ、0.1%基準で有意差が確認できた( $\chi^2 = 107.14$ ;  $p < .001$ )。つまり、母語話者の方が学習者よりも *on* を使用する確率が高いことが分かる。

また、*on* と共に使用されている動詞を比較してみると、学習者が使用している動詞は母語話者も使用していることが分かる。そこで、共通して使用されていた *se retrouve*、*prendre*、*va* について、*on* との共起頻度を調べると同時に、*nous* と共起された場合の活用形 *nous retrouvons*、*prenons*、*allons* と *nous* との共起頻度も調べ、どちら代名詞主語を選択しているのかを比較した。

表 99 *on* と *nous* とともに使用された動詞の生起頻度

	<i>on</i>		<i>nous</i>	
	母語話者	学習者	母語話者	学習者
SE RETROUVER	12	17	0	0
PRENDRE	27	17	0	0
ALLER	39	9	0	11

すると、動詞 *SE RETROUVER* と *PRENDRE* については、両グループとも人称代名詞 *nous* を使用している例が見つからなかったが、*ALLER* に対して学習者は *on* よりも *nous* を使用していることが分かる。

次に、母語話者によって多く使用されていたが、学習者にはあまり使用されていなかった動詞 *a* (*AVOIR*)、*est* (*ÊTRE*)、*peut* (*POUVOIR*)、*arrive* (*ARRIVER*) について同様の比較を行った。

表 100 母語話者によって多く使用されていた動詞

	<i>on</i>		<i>nous</i>	
	母語話者	学習者	母語話者	学習者
<i>AVOIR</i>	96	5	7	13
<i>ÊTRE</i>	57	0	1	9
<i>POUVOIR</i>	43	4	0	0
<i>ARRIVER</i>	16	4	0	0

いずれの動詞についても、母語話者が *on* を使用した頻度は学習者よりも高い。また、動詞 *POUVOIR* と *ARRIVER* について、両グループとも人称代名詞主語 *nous* を使用していなかったが、*AVOIR* と *ÊTRE* について、母語話者と学習者で人称代名詞 *on* と *nous* の選択の傾向が逆転していることが分かる。*AVOIR* と *ÊTRE* は、過去形を作る助動詞として使用される。そこで、母語話者が過去形を用いるために *on a* もしくは *on est* を使用したタスクを調べたところ、以下のタスクとなった。

f030	Racontez votre dernier voyage.
j030	あなたが最近行った旅行について話してください。
f031	Racontez votre dernière fin de semaine.
j031	あなたの先週末の過ごし方について話してください。

以上のタスクで用いられた過去形について、*nous* と *on* の出現頻度の比較を行ったところ、以下の通りとなった。

表 101 *on* と *nous* を主語にとる過去形の生起頻度

	母語話者	学習者
<i>on a +PP</i>	67	4
<i>on (s') est +PP</i>	59	2
合計	126	6
<i>nous avons +PP</i>	3	12
<i>nous (nous) sommes +PP</i>	1	12
合計	4	24

以上について有意差を検定したところ、有意差が見られた ( $\chi^2=99.90$ ;  $p<.001$ )。つまり、過去形を使用する際も母語話者は学習者よりも有意に多く *on* を使用している。

表 102 学習者による *on* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
EST	21.10 (6)	A	31.65 (9)	NP	49.23 (14)	ON	330.57 (94)	SE	66.82 (19)	RETROUVE	59.78 (17)	NP	66.82 (19)
		CE	31.65 (9)	SI	31.65 (9)			PREND	56.27 (16)	LE	59.78 (17)	RESTAURANT	31.65 (9)
				ET	21.10 (6)			VA	31.65 (9)	AU	31.65 (9)	À	21.10 (6)
										A	21.10 (6)	LE	21.10 (6)
												DEVANT	21.10 (6)



表 103 母語話者による *on* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
NP	32.80 (17)	ET	65.61 (34)	QU	146.65 (76)	ON	1005.35 (521)	A	185.25 (96)	À	48.24 (25)	NP	55.96 (29)
ON	28.94 (15)	À	44.38 (23)	HESIT	86.83 (45)			SE	111.92 (58)	EST	36.66 (19)	À	44.38 (23)
ET	28.94 (15)	PARCE	38.59 (20)	ET	81.05 (42)			EST	109.99 (57)	FAIT	32.80 (17)	UN	40.52 (21)
À	23.16 (12)	NP	32.80 (17)	NP	55.96 (29)			PEUT	82.97 (43)	AU	30.87 (16)	DE	23.16 (12)
DE	23.16 (12)	ON	28.94 (15)	DONC	44.38 (23)			VA	69.47 (36)	PAS	27.02 (14)	LE	23.16 (12)
LA	21.23 (11)	LE	25.09 (13)	PUIS	28.94 (15)			S'	52.10 (27)	LE	28.94 (15)	BEAUCOUP	21.23 (11)
L'	21.23 (11)	HESIT	25.09 (13)	QUAND	21.23 (11)			PREND	52.10 (27)	PAS	27.02 (14)	ET	21.23 (11)
								ARRIVE	30.87 (16)	UN	27.02 (14)	HESIT	21.23 (11)
								AILLE	28.94 (15)	ALLER	25.09 (13)		
										VOIT	25.09 (13)		
										BEAUCOUP	23.16 (12)		
										RETROUVE	23.16 (12)		

### 7.1.2.3. *donc*

母語話者コーパスも学習者コーパスも、*donc* の後ろの R1 の位置に主語が続いている例が多く見られる。つまり、両グループとも発話単位の頭で *donc* を使用する傾向にあると言える。

また、母語話者コーパスに特徴的な例は、*donc* の R1 の位置に *voilà* という語が置かれる例であり、20 例見られた。

je suis désolé j'ai HESIT j'ai plus de batteries en plus j'ai pas de crédits donc il sera pas trop utile donc

**donc voilà** (FT07)

ma grand mère pour le moment elle est à l'hôpital **donc voilà** (FT22)

je stresse plus avant l'examen que pendant **donc voilà** (FT34)

ce matin j'ai pas eu le temps de manger **donc voilà** mais d'habitude c'est un fruit un thé un yaourt  
(FT03)

j'aime beaucoup lire les les fictions ou les les livres de littérature INT des littérature françaises  
anglaises et INT **donc voilà** mais j'aime pas trop lire les HESIT les policiers les les romans policiers  
(FT33)

これは、発話単位の先頭に用いられるのではなく、むしろ発話単位の後ろで用いられ、発話単位終了のディスコースマーカの役割を果たす。たとえば初めの 3 例は、文末に用いられており、発話終了を示している。*donc voilà* の次に *mais* が続く 2 例は、*mais* 以前と以降で話題が相反する。したがって、この場合も前半の発話単位の終了を示すマーカとして用いられていると言える。

また、発話単位の最後だけではなく、発話単位の中にも *donc voilà* は用いられることがある。この場合、意味としては *donc* とあまり変わりがないが、接続詞 *donc* が単独で使用されるよりは、前後で発話の区切りが見られる。そして、後ろにコメントとして理由を付加する機能を持つ。

c'était la folie **donc voilà** c'était fun (FT07)

c'est la dernière semaine avant la semaine de révision pour les exams **donc voilà** je me dis qu'il faut que je me mette à bosser à fond et puis voilà (FT16)

一方、学習者コーパスには *donc voilà* という MWU は、どちらの使用方法としても用いられていなかった。

表 104 学習者による *donc* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	-	-	DONC	144.18 (41)	JE	66.82 (19)	NE	24.62 (7)	-	-

表 105 母語話者による *donc* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
DE	38.59 (20)	LA	40.52 (21)	NP	55.96 (29)	DONC	617.49 (320)	JE	113.85 (59)	EST	63.68 (33)	PAS	36.66 (19)
À	34.73 (18)	DE	30.87 (16)	HESIT	27.02 (14)			C'	54.03 (28)	JE	32.80 (17)	NP	28.94 (15)
PAS	25.09 (13)	NP	25.09 (13)	ET	23.16 (12)			ON	44.38 (23)			DE	25.09 (13)
		LE	23.16 (12)					HESIT	42.45 (22)			UN	23.16 (12)
								J'	40.52 (21)				
								VOILA	38.59 (20)				
								ÇA	21.23 (11)				
								IL	21.23 (11)				

#### 7.1.2.4. *peu*

*un peu* (『すこし』) という MWU が母語話者コーパスには 114 例、学習者コーパスには 6 例見られる。そこで、*peu* の使用頻度に対して、*un peu* が使用される割合について学習者と母語話者で比較したところ、有意差が見られなかった ( $\chi^2=0.09$ ;  $p>.05$ )。しかし、両コーパスにおける *peu* の生起頻度の差自体がかなり大きいことは留意しなければならない。また、*un petit peu* という MWU は母語話者コーパスには 18 例見られた。これは、*un peu* の拡張表現である。拡張表現が学習者コーパスに 1 例のみしか見られなかった。

そして、母語話者コーパスにはこの他にも *peu* を含む MWUs が見られる。たとえば、*à peu près* は学習者コーパスには 1 例だけ見られたのに対して、母語話者コーパスには 27 例見られた。この MWU は『だいたい』という意味を持ち、曖昧性付加句の一種であ

る。先行研究の Altenberg (1998)では、母語話者はこのような付加句を多く使用していることが指摘していたが、その結論を追認する結果となった。

このように、学習者と母語話者の間で *peu* の出現頻度が大きく異なっている一つの原因は、このような MWUs の使用頻度の差であると考えられる。

表 106 学習者による *peu* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	UN	21.10 (6)	PEU	28.13 (8)	-	-	-	-	-	-

表 107 母語話者による *peu* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
JE	36.66 (19)	UN	34.73 (18)	UN	219.98 (114)	PEU	314.53 (163)	PRÈS	59.82 (31)	-	-	-	-
C'	28.94 (15)	EST	32.80 (17)	A	52.10 (27)			DE	27.02 (14)				
				PETIT	34.73 (18)								

#### 7.1.2.5. *bien*

両コーパスともに共通して *bien* と共に使用される MWUs は *bien sûr* と *très bien* であることが分かる。*bien* の生起頻度に対する *bien sûr* の頻度、そして *très bien* の頻度についてそれぞれ有意差があるのかどうかそれぞれ検定を行ったところ、以下の通りとなった。

表 108 *bien* を含む MWUs

	母語話者	学習者	$\chi^2$ 統計量、有意確率
<i>bien sûr</i>	13	22	$\chi^2=68.29; p<.001$
<i>très bien</i>	14	7	$\chi^2=5.70; p<.05$

それぞれ有意差が見られたことより、学習者の方が *bien* を使用する際、*bien sûr* や *très bien* という MWUs で使用される割合が高いと言える。つまり、学習者の方が母語話者と比べて *bien* の生起頻度は少ないだけでなく、使用パターンはかなり限られているということが分かる。

一方、母語話者は動詞 AIMER と共に *bien* を使用している。*aime bien* と *aimerais bien* について、動詞 AIMER と *bien* との共起強度を比較したところ、以下の通りとなった。

表 109 AIMER +*bien* の頻度の比較と共起頻度

	母語話者		学習者	
	頻度	共起強度	頻度	共起強度
<i>aime bien</i>	150	1011.74	3	3.65
<i>aimerais bien</i>	27	196.08	1	4.42

表より、学習者と比較して母語話者は *bien* とともに動詞 *AIMER* を多く使用していることが分かる。現在形の *aimer* + *bien* は、「～が好き」という意味を持つが、*bien* と共に *aime* を使用することによって *aime* を単独で使用するよりも意味を弱めることができる。また、*aimerais bien*+動詞句は、「～したい」という意味を持つ。母語話者によって *aimerais bien* が多く用いられていたことは、これとほぼ同じ意味を持つ *voudrais*+動詞句の頻度が学習者と比べて少なかったことの一因であると考えられる。

表 110 学習者による *bien* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
c'	21.10 (6)	—	—	OUI	45.72 (13)	BIEN	196.93 (56)	SÛR	77.37 (22)	VOILÀ	21.10 (6)	—	—
				TRÈS	24.62 (7)								

表 111 母語話者による *bien* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
WEEKEND	52.10 (27)	J'	345.41 (179)	AIME	289.45 (150)	BIEN	603.98 (313)	SORTIR	28.94 (15)	DES	23.16 (12)	PARCE	23.16 (12)
MAIS	28.94 (15)	C'	40.52 (21)	AIMERAIS	52.10 (27)			LE	27.02 (14)	JE	21.23 (11)		
HESIT	28.94 (15)			EST	40.52 (21)			SÛR	25.09 (13)				
QUE	27.02 (14)			TRÈS	27.02 (14)			LIRE	23.16 (12)				
C'	25.09 (13)							LA	23.16 (12)				
ÇA	23.16 (12)							LES	21.23 (11)				
ET	23.16 (12)												

#### 7.1.2.6. c'

これは、指示代名詞である *ce* の省略形であり、動詞 *ÊTRE* の前で用いられる。実際に、学習者による 2 例のエラーを除いて、全て *c* の後ろには動詞 *ÊTRE* が用いられていた。この指示代名詞と共に用いられる動詞で両グループに差があるとすると、時制の差である。そこで現在形である *c'est* と半過去形である *c'était* との時制の使用頻度差を検定したところ、5%水準で有意差が認められた ( $\chi^2=5.82$ )。それぞれ *ÊTRE* の時制の頻度の割合を見ても、学習者の方が現在形を多く用いていることが分かる。つまり、学習者は現在形の *c'est* をより多く使用しているということがいえる。

また、この指示代名詞の使用自体にも違いが見られる。たとえば、母語話者に多く見られた例が *c'est c'est* と繰り返し使用されている例である。母語話者コーパスにはこの繰り返しが 32 例見られたが、学習者コーパスには 5 例のみしか見られなかった。

また、この指示代名詞の R2 の頻度を見てみると、否定文で用いられる副詞 *pas* が 36 例見られるが、学習者コーパスの調整頻度 20 以上のリストには見られなかった。これは、*c'est* や *c'était* が否定文になる際に、否定辞 *ne* を省略しない *ce n'est pas* や *ce n'était pas* が使用されていたことに起因することが考えられる。そこでこれらの否定文について、使用頻度を比較した。

表 112 *ne* の脱落と保持(*c'est*, *c'était*)

	母語話者	学習者
<i>ce n'est pas</i>	1	9
<i>c'est pas</i>	31	1
<i>ce n'était pas</i>	1	0
<i>c'était pas</i>	5	0

以上の表に基づき、否定辞 *ne* が用いられる場合と省略される場合の使用頻度について両グループ間に有意差があるかどうか検定を行ったところ、有意差が確認された ( $\chi^2=32.18$ ;  $p<.001$ )。ここでもまた、母語話者が否定辞 *ne* を脱落する傾向にあることが分かった。*ne* の脱落によって、この指示代名詞が *ce* から *c* に変わる。この点についても、学習者による *c* の過少使用の一因となったと考えられる。

一方、指示代名詞 *c* を含む MWUs のうち、学習者コーパスに見られて母語話者コーパスは一例も見られなかった MWU もあった。それは、*bonjour c'est* という MWU であり、学習者コーパスの中では 11 例見られた。これは、以下のタスクで用いられていたものである。

j032	あなたはこの週末一緒に出かけられるかどうかを聞こうと、友人に電話をしています。ところが、留守番電話に接続されました。「いつ」、「どこで」、「どんなことをしたいのか」を留守番電話に吹き込み、友人を誘ってください。
------	---

学習者はこのタスクでまず *bonjour* と挨拶をしてから *c'est NP* と名前を名乗っていた。一方、母語話者が同タスクで名前を名乗る際に *c'est* を用いた際、挨拶として *bonjour* を使用せず、*salut* (10 例) もしくは *coucou* (1 例) を用いていた。確かに、このような状況において *bonjour* と挨拶から始めるのは慣例ではなく、学習者は正しい使用ができていないと言える。

また、*c* との共起語で学習者に特徴的なのは R2 の位置に見られる *très* である。そこで、R2 の位置に *très* が用いられるパターンの生起頻度を調べ、*c'est* および *c'était* と *très* との共起強度を比較したところ以下の通りとなった。

表 113 *c'est*, *c'était* と *très* との共起

	母語話者		学習者	
	頻度	共起強度	頻度	共起強度
<i>c'est très</i>	26	70.67	36	201.29
<i>c'était très</i>	17	89.17	14	123.98

共起強度より、学習者は *c'est* と *c'était* の生起頻度に対して *très* との共起頻度が高いことが分かる。

表 114 学習者による *c* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
EST	94.95 (27)	PARCE	101.98 (29)	QUE	196.93 (56)	C'	833.45 (237)	EST	752.57 (214)	TRÈS	175.83 (50)	PLAT	52.75 (15)
ADRESSE	38.68 (11)	CE	101.98 (29)	NP	105.50 (30)			ÉTAIT	73.85 (21)	NP	80.88 (23)	BONNE	45.72 (13)
NP	35.17 (10)	DE	45.72 (13)	BONJOUR	38.68 (11)					CE	49.23 (14)	NP	38.68 (11)
LE	24.62 (7)	À	35.17 (10)	ET	35.17 (10)					COMBIEN	45.72 (13)	BON	31.65 (9)
BEAUCOUP	21.10 (6)	DES	28.13 (8)							UNE	45.72 (13)	JE	28.13 (8)
		EN	24.62 (7)							LE	35.17 (10)	EST	21.10 (6)
		AU	24.62 (7)							UN	21.0 (6)	AMUSANT	21.10 (6)
										LA	21.10 (6)		

表 115 母語話者による *c* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
C'	67.54 (35)	PARCE	127.36 (66)	QUE	187.18 (97)	C'	1534.07 (795)	EST	1302.51 (675)	UN	104.2 (54)	NP	67.54 (35)
NP	59.82 (31)	C'	77.19 (40)	HESIT	125.43 (65)			ÉTAIT	231.56 (120)	UNE	88.76 (46)	EST	65.61 (34)
EST	48.24 (25)	ET	69.47 (36)	NP	117.71 (61)					LE	86.83 (45)	UN	50.17 (26)
ET	40.52 (21)	NP	65.61 (34)	ET	86.83 (45)					TRÈS	82.97 (43)	TRÈS	38.59 (20)
MON	38.59 (20)	EST	61.75 (32)	EST	65.61 (34)					PAS	69.47 (36)	HESIT	36.66 (19)
MA	36.66 (19)	EN	54.03 (28)	DONC	54.03 (28)					NP	57.89 (30)	PEU	28.94 (15)
À	32.8 (17)	LE	44.38 (23)	MAIS	48.24 (25)					VRAIMENT	48.24 (25)	ET	27.02 (14)
LA	30.87 (16)	HESIT	34.73 (18)	VOILÀ	30.87 (16)					ASSEZ	42.45 (22)	BIEN	25.09 (13)
HESIT	27.02 (14)	LA	30.87 (16)	PRÉFÉRÉE	28.94 (15)					LA	40.52 (21)	JOUR	21.23 (11)
QUE	27.02 (14)	SAISON	30.87 (16)	MOI	27.02 (14)					BIEN	40.52 (21)		
FAIRE	25.09 (13)	DERNIER	27.02 (14)	VOYAGE	27.02 (14)					QUELQU	36.66 (19)		
JE	23.16 (12)	À	23.16 (12)	SALUT	21.23 (11)					PLUS	32.8 (17)		
DE	21.23 (11)	DE	23.16 (12)	ENFIN	21.23 (11)					À	28.94 (15)		
UN	21.23 (11)	CE	23.16 (12)							L'	27.02 (14)		
										POUR	27.02 (14)		
										TROP	27.02 (14)		
										QUOI	25.09 (13)		
										HESIT	25.09 (13)		
										SUPER	21.23 (11)		

#### 7.1.2.7. *vraiment*

学習者による副詞 *vraiment* の使用頻度はかなり低い。一方副詞でも、*beaucoup* や *très* といった副詞や、時を表す副詞(*ce weekend* など)は学習者によってもよく使用されている。そこで、副詞の中でも *vraiment* 同様に *-ment* の形をした副詞で、母語話者の調

整頻度が 20 以上の副詞について比較を行った。母語話者コーパスの中でこの条件に該当した副詞は *vraiment* の他に *actuellement* と *justement* であった。

表 116 *-ment* 型の副詞の頻度

	母語話者コーパス	学習者コーパス
<i>vraiment</i>	202.61 (105)	21.10 (6)
<i>actuellement</i>	42.45 (22)	0 (0)
<i>justement</i>	27.02 (14)	3.52 (1)

総語彙数に対するこれらの副詞の生起頻度の割合について有意差があるかどうか調べたところ、有意差が見られた ( $\chi^2 = 61.09$ ;  $p < .001$ )。なお、学習者コーパスで最も多く使用されていた *-ment* 型の副詞は *seulement* で 12 例見られたが、母語話者コーパスでは 2 例のみであり、この副詞については学習者の方が有意に多く使用していた ( $\chi^2 = 15.48$ ;  $p < .001$ )。しかし、学習者によって二番目に多く使用されていた副詞は *vraiment* であったことより、*-ment* 型の副詞については全般的に、学習者が過少使用していると言える。

表 117 学習者による *vraiment* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	-	-	VRAIMENT	21.10 (6)	-	-	-	-	-	-

表 118 母語話者による *vraiment* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	C'	48.24 (25)	EST	36.66 (19)	VRAIMENT	202.61 (105)	-	-	-	-	-	-
		JE	36.66 (19)	PAS	32.80 (17)								
				SUIS	23.16 (12)								

#### 7.1.2.8. *tout*

母語話者コーパスの中で多く見られた *tout* を含む MWUs は、*tout le* + 名詞、(*pas*) *du tout*、*et tout* であった。*tout* の総頻度数に対するこれらの MWUs の生起回数の割合について、母語話者と学習者の頻度を比較して有意差を検定したところ、以下の通りとなった。

表 119 *tout* を含む MWUs

	母語話者コーパス	学習者コーパス	$\chi^2$ 統計量、有意確率
<i>tout le</i> + 名詞	22	3	$\chi^2 = 0.24$ ; $p > .05$
( <i>pas</i> ) <i>du tout</i>	21	2	$\chi^2 = 0.02$ ; $p > .05$
<i>et tout</i>	14	0	NA

*tout le* + 名詞と (*pas*) *du tout* については、MWUs の共起頻度と *tout* 自体の頻度との関係について、両グループ間で有意差が見られなかった。しかし、*et tout* という MWU に

ついて、学習者コーパスで一例も見られなかった点については、注目すべきである。*et tout* は話し言葉でよく使用される MWU である。「～など」という意味を持ち、文を曖昧にする曖昧性付加句としての機能を働く。このような機能として使われたのは、母語話者コーパスには 11 例見られたが、学習者コーパスには一例も見られなかった。

mec ça te dit pas qu'on aille bouffer un au resto ça serait plutôt sympa ça c'est fait un bail **et tout** (FT16)  
 des séries dessins animés HESIT **et tout** (FT29)  
 quand on se voit on est toujours ensemble **et tout** et HESIT elle est elle est super (FT35)

表 120 学習者による *tout* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
—	—	—	—	—	—	TOUT	49.23 (14)	—	—	—	—	—	—

表 121 母語話者による *tout* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
—	—	PAS	32.80 (17)	DU	40.52 (21)	TOUT	260.50 (135)	LE	42.45 (22)	—	—	—	—
				ET	27.02 (14)								

## 7.2. 自由会話コーパス

### 7.2.1. 過剰使用語との共起語

前節で抽出した自由会話コーパスに見られた過剰使用語について、調整頻度 20 回以上共起した語を前後 3 語ずつまとめた。これらの MWUs について、個々に分析を行っていく。なお、表における共起語および、特徴語の隣にある数字は調整頻度を示し、カッコ内の数字は生起回数を示している。

まず、表をもとに学習者コーパスと母語話者コーパスにそれぞれ特徴的なコロケーションおよびコリゲーションを選択して比較する。次に、その使用について学習者間で差があるかどうかを比較する。学習者のレベルについては、「語彙の豊かさ」で分類した「初中級学習者」と「上級学習者」の 2 レベルに基づくこととする。

#### 7.2.1.1. INT

自由会話において、学習者によってもっとも過剰使用されていたものは間投詞であった。間投詞は、主に話し言葉の単位である。感情を示し、話者の認知的変化を指し示す単位語もしくは短い定型句のことである。また、間投詞は隣接した節から分布的に独立しており、節レベルで統語関係を確立することはできないが、談話レベルでは重要な機能を持ち、日常会話で出来事を話す際の構成において、特別な役割を持つ(Norrick 2008 :439)。

間投詞の使用について母語話者との比較だけではなく、上級学習者と初中級学習者との間にも違いがあるかどうか調べた。生起頻度は上級学習者が 433 回、初中級学習者は



467回であった。それぞれ3グループの総語数とINTの数に有意差があるかどうか検定したところ、以下のような結果となった。

表 122 INTのグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
INT	612.89 (402)	1765.40 (433)	4674.21 (467)	$\chi^2=1168.76$ ; $p<.001$

そこで、ボンフェローニの補正を用いて有意水準を1.67%として多重比較を行ったところ、以下ようになった。

INT	初級学習者>上級学習者>母語話者
-----	------------------

したがってINTの多用は初中級学習者に特徴付けられたものであるといえる。

次に、MWUsレベルでの比較を行う。間投詞と共に多く用いられていたのは、*oui*、*ouais*、*non*といった副詞であった。これらの副詞との共起頻度を調べたところ、以下ようになった。

表 123 INTを含むMWUs(1)

	母語話者	上級学習者	初中級学習者
INT <i>oui</i>	95	116	49
INT <i>ouais</i>	77	26	2
INT <i>non</i>	29	15	7
総頻度	201	157	58

INTの総頻度において*oui/ouais/non*とともに使用されたINTの使用頻度の割合について、母語話者、上級学習者、初中級学習者との間で検定を行ったところ、有意差が見られた( $\chi^2=145.86$ ;  $p<.001$ )。そこでボンフェローニ補正を適応し多重比較を行ったところ、以下ようになった(有意水準1.67%)。

INT+ <i>oui/ouais/non</i>	初中級学習者<上級学習者<母語話者
---------------------------	-------------------

つまり、間投詞自体は学習者の方が使用しているものの、間投詞とこれらの副詞を共に使用する割合は母語話者が最も高く、初級学習者がもっとも低いということが分かる。

また両コーパスに共通して見られた間投詞を含むMWUsとして、INT+*bon*やINT+*d'accord*があった。これらについても同様に、INTに占める割合について有意差を調べた。

表 124 INTを含むMWUs(2)

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
INT <i>bon</i>	15	5	9	$\chi^2=6.66$ ; $p<.05$
INT <i>d'accord</i>	17	30	12	$\chi^2=9.99$ ; $p<.01$

多重比較を行ったところ、INT+*bon*については初中級学習者と母語話者との間に有意差は見られなかったが、上級学習者よりも母語話者の方が多くこのMWUを使用していた。また、INT+*d'accord*については、初中級学習者と上級学習者の間に有意差が見られ、その傾向は上級学習者に顕著であった。

INT+ <i>bon</i>	上級学習者<母語話者
INT+ <i>d'accord</i>	上級学習者>初中級学習者

また、学習者の間投詞の使用の表にはINT INTといった間投詞の連続も見られた。INTの総頻度におけるこの連続使用の頻度について、母語話者、上級学習者、初中級学習者の間に有意差があるかどうかを調べたところ、有意差が見られた。

表 125 INTの繰り返し

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
INT INT	5	6	19	$\chi^2=20.67$ ; $p<.001$

そこで、どのグループ間に有意差があるのかどうかを調べるため、ボンフェローニ補正を適応して多重比較を行ったところ、INTが使用される際、INT INTは初中級学習者によって有意に多く使われていることが分かる。

INT INT	初中級学習者>上級学習者=母語話者
---------	-------------------

以上より、INTの生起頻度は学習者の方が多く、初中級学習者によって特に多く用いられていたことを確認した。しかしMWUsレベルで見ると、INTとともに使用される語句には、母語話者に特徴的なものも見られるということが分かった。

表 126 学習者による INT の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
INT	37.66 (13)	INT	81.12 (28)	INT	78.22 (27)	INT	2607.34 (900)	OUI	478.01 (165)	EST	162.23 (56)	EST	121.68 (42)
C'	26.07 (9)	C'	31.87 (11)	NP	49.25 (17)			C'	170.93 (59)	ACCORD	121.68 (42)	C'	66.63 (23)
ET	23.18 (8)	BC	28.97 (10)	HESIT	49.25 (17)			D'	133.26 (46)	C'	118.78 (41)	JE	63.73 (22)
EST	20.28 (7)	EST	26.07 (9)	OUI	43.46 (15)			JE	95.6 (33)	OUI	75.32 (26)	OUI	60.84 (21)
J'	20.28 (7)	EN	23.18 (8)	BC	37.66 (13)			Ouais	81.12 (28)	JE	60.84 (21)	J'	60.84 (21)
LA	20.28 (7)	DE	20.28 (7)	EST	31.87 (11)			J'	63.73 (22)	J'	37.66 (13)	NP	46.35 (16)
				FW	23.18 (8)			NON	63.73 (22)	MAIS	37.66 (13)	DE	31.87 (11)
								NP	55.04 (19)	DE	34.76 (12)	BC	31.87 (11)
								FW	55.04 (19)	BC	34.76 (12)	A	26.07 (9)
								BC	40.56 (14)	AI	31.87 (11)	OK	26.07 (9)
								MAIS	40.56 (14)	HESIT	31.87 (11)	MAIS	23.18 (8)
								BON	40.56 (14)	ÉTAIT	28.97 (10)	TU	23.18 (8)
								TU	34.76 (12)	NON	28.97 (10)	BIEN	23.18 (8)
								DONC	31.87 (11)	NP	26.07 (9)	NON	20.28 (7)
								ET	28.97 (10)	TU	23.18 (8)	À	20.28 (7)
								LA	26.07 (9)	ET	23.18 (8)	QUE	20.28 (7)
								OK	26.07 (9)	Y	20.28 (7)		
								MOI	23.18 (8)	SUIS	20.28 (7)		
								LE	23.18 (8)	D'	20.28 (7)		
								HESIT	20.28 (7)	ÇA	20.28 (7)		
								À	20.28 (7)				
								DANS	20.28 (7)				
								EN	20.28 (7)				
								IL	20.28 (7)				

表 127 母語話者による INT の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	-	-	INT	612.89 (402)	OUI	144.84 (95)	C'	33.54 (22)	EST	28.96 (19)
								Ouais	120.44 (79)	OUI	32.02 (21)		
								C'	38.11 (25)	EST	32.02 (21)		
								NON	32.02 (21)	ACCORD	25.92 (17)		
								D'	27.44 (18)	MAIS	22.87 (15)		
								BON	22.87 (15)				

### 7.2.1.2. *oui*

まず、*oui* の過剰使用が学習者のレベルに起因するものかどうかを比較した。

表 128 *oui* のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>oui</i>	885.79 (581)	2907.00 (713)	3121.89 (321)	$\chi^2=638.41$ ; $p<.001$

有意差が見られたため、ボンフェローニの補正を用いて有意水準を 1.67% として多重比較を行ったところ、以下ようになった。

<i>oui</i>	初中級学習者 = 上級学習者 > 母語話者
------------	-----------------------

つまり、*oui* の出現頻度は初中級学習者と上級学習者間には有意差がなく、頻度の差があるのは学習者グループと母語話者グループの間であった。

次に、MWU レベルの使用に起因しているのかどうかについて分析を行う。MWUs レベルについて分析を行うため、学習者コーパスと母語話者コーパスを比較したところ、両グループとも *oui* を連続して使用していた。一般疑問文に対する答えである *oui* は「はい」という意味であるが、話し言葉では相づちの機能も持ち、むしろこの機能の方が多く使用されている。さらに *oui* を連続使用することによって、話者への相づちとしての意味合いが強くなる。*oui oui* と 2 回連続された場合、*oui oui oui* と 3 回連続された場合、そして *oui oui oui oui* と 4 回連続された場合の *oui* に占めるこれらの連続使用頻度の割合について、母語話者コーパスと学習者コーパスを比較した。

表 129 *oui* の繰り返しの頻度

	母語話者コーパス	上級学習者	初中級学習者
<i>oui oui</i>	41	43	29
<i>oui oui oui</i>	14	41	18
<i>oui oui oui oui</i>	2	3	4

初中級学習者、上級学習者、母語話者の 3 グループ間について、*oui* の出現頻度と連続使用の頻度の割合について有意差検定を行ったところ、*oui oui* の 2 連続 ( $\chi^2=3.07$ ;  $p>.05$ ) については有意差が見られなかったが、*oui oui oui* の 3 連続 ( $\chi^2=31.07$ ;  $p<.001$ )、*oui oui oui oui* の 4 連続使用 ( $\chi^2=14.09$ ;  $p<.001$ ) については有意差が見られた。そこで、ボンフェローニの補正を用いて有意水準を 1.67% として多重比較を行ったところ、以下ようになった。

<i>oui oui oui</i>	上級学習者 > 初級学習者 > 母語話者
<i>oui oui oui oui</i>	初級学習者 > 上級学習者 = 母語話者

このように学習者が *oui* を過剰使用した原因は、特に学習者による連続使用が一因となっている可能性が高い。

また、フランス語学習者による *oui* の過剰使用には、*ouais* (本章 7.2.2.10 を参照のこと) の過少使用との関連があると考えられる。*ouais* は *oui* のくだけた形で、非形式的な話し言葉によく見られるものである。そこで *oui* と *ouais* の生起頻度について、母語話者、上級学習者、初中級学習者で差が見られるかどうか検定を行ったところ、有意差が見られた ( $\chi^2=496.80$ ;  $p<.01$ )。そこで多重比較を行い、どのグループ間に差が見られるのかを観察する。その際、ボンフェローニ補正を適応することとし、有意水準を 1.67% と定めた。

表 130 *oui* と *ouais* の頻度

	母語話者コーパス	上級学習者	初中級学習者
<i>oui</i>	885.79 (581)	2907.00 (713)	3212.89 (321)
<i>ouais</i>	1405.68 (922)	937.74 (230)	370.33 (37)

その結果、全てのグループの間に有意差があり、*ouais* の使用頻度は母語話者 > 上級学習者 > 初中級学習者であることが分かった。この結果を生起頻度と照らし合わせると、学習者は *oui* を使用し、母語話者は *ouais* を使用する傾向にあると言える。以下の図は、初中級学習者、上級学習者、母語話者の *oui* と *ouais* の調整頻度を示したものである。

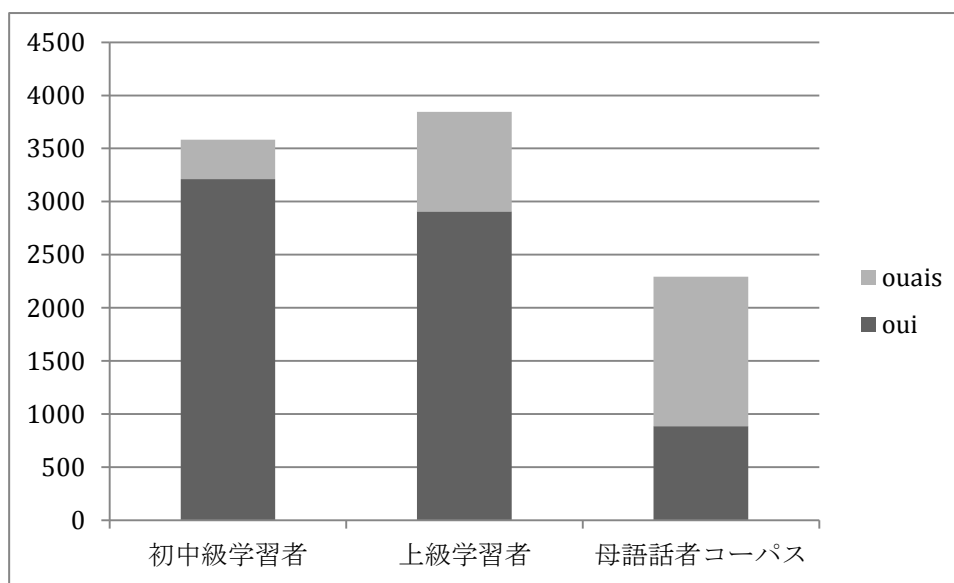


図 16 *oui* と *ouais* の頻度

以上の表から、初中級学習者、上級学習者、母語話者の順に *oui* の頻度が低くなり、上級学習者と母語話者の *ouais* の使用頻度の差も大きいですが、レベルが上がるにしたがって、*ouais* の使用頻度が上がっているということが分かる。

次にコーパス全体の頻度における *oui* と *ouais* を合わせた頻度の割合について、3 グループを比較したところ、有意差が見られた ( $\chi^2=183.15$ ;  $p<.01$ )。そこでボンフェローニ補正を用いて多重比較を行い、どのグループ間に差が見られるのかを観察したところ、初

中級学習者と上級者の間には有意差が見られなかったが、学習者グループと母語話者との間には有意差が確認された(1.67%水準)。

<i>oui</i> と <i>ouais</i> の頻度	初級学習者 = 上級学習者 > 母語話者
-------------------------------	----------------------

以上を総合すると、学習者はレベルを問わず *oui* を過剰使用しているが、レベルが上がると *ouais* の使用割合も増えている。つまり、初中級学習者よりも上級学習者の方が非形式的な用法を好む傾向が見られる。そして、学習者は *ouais* を過少使用しているものの、*oui* と *ouais* を合わせた頻度は母語話者よりも高く、特に相づちとしての用法に多く用いられていると考えられる。

表 131 学習者による *oui* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
C'	72.43 (25)	OUI	254.94 (88)	OUI	637.35 (220)	OUI	2995.54 (1034)	C'	373.72 (129)	EST	341.85 (118)	ÇA	141.95 (49)
INT	60.84 (21)	INT	75.32 (26)	INT	480.91 (166)			JE	107.19 (37)	C'	104.29 (36)	EST	104.29 (36)
OUI	57.94 (20)	EST	66.63 (23)	NP	60.84 (21)			MAIS	95.6 (33)	AI	55.04 (19)	C'	84.01 (29)
EST	37.66 (13)	NP	28.97 (10)	HESIT	46.35 (16)			ET	92.71 (32)	J'	55.04 (19)	VRAI	66.63 (23)
DE	26.07 (9)	DE	26.07 (9)	ET	37.66 (13)			J'	78.22 (27)	JE	49.25 (17)	J'	57.94 (20)
A	20.28 (7)	BC	23.18 (8)	ÇA	31.87 (11)			IL	46.35 (16)	NP	49.25 (17)	JE	55.04 (19)
BC	20.28 (7)			BC	23.18 (8)			BC	46.35 (16)	MAIS	43.46 (15)	HESIT	55.04 (19)
								DONC	46.35 (16)	Y	37.66 (13)	AI	40.56 (14)
								INT	43.46 (15)	ÉTAIT	31.87 (11)	MAIS	34.76 (12)
								HESIT	40.56 (14)	QUE	28.97 (10)	NP	31.87 (11)
								PARCE	37.66 (13)	PENSE	28.97 (10)	TU	31.87 (11)
								TU	28.97 (10)	ET	28.97 (10)	A	31.87 (11)
								À	28.97 (10)	HESIT	28.97 (10)	ET	28.97 (10)
								EN	26.07 (9)	DONC	26.07 (9)	IL	28.97 (10)
								ON	23.18 (8)	EN	26.07 (9)	TRÈS	26.07 (9)
								MOI	23.18 (8)	BC	23.18 (8)	EN	26.07 (9)
								D'	20.28 (7)	SUIS	20.28 (7)	BIEN	23.18 (8)
								ÇA	20.28 (7)	FAIT	20.28 (7)	QUE	23.18 (8)
												FAIT	20.28 (7)

表 132 母語話者による *oui* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
OUI 24.39 (16)	OUI 57.93 (38)	INT 144.84 (95)	OUI 885.79 (581)	C' 73.18 (48)	EST 71.66 (47)	EST 28.96 (19)
	INT 25.92 (17)	OUI 143.31 (94)		MAIS 45.74 (30)	C' 27.44 (18)	C' 22.87 (15)
		ET 39.64 (26)		HESIT 36.59 (24)	Y 24.39 (16)	HESIT 22.87 (15)
				IL 28.97 (19)	ACCORD 21.34 (14)	
				JE 24.39 (16)		
				D' 21.34 (14)		

### 7.2.1.3. NP

自由会話コーパスのため、固有名詞の内訳は様々である。そのため質的分析行わない。ここでは、固有名詞の生起頻度を比較する。初中級学習者、上級学習者の生起頻度を調べたところ、以下の通りとなった。

表 133 NP のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
NP	1248.65 (819)	2894.77 (701)	4494.04 (449)	$\chi^2=607.58 ; p<.001$

有意差が見られたため、ボンフェローニ補正を用いて多重比較を行い、どのグループ間に差が見られるのかを観察したところ、以下の通りとなった。

NP	初級学習者 > 上級学習者 > 母語話者
----	----------------------

初中級学習者と母語話者の間、上級学習者と母語話者の間、そして異なるレベルの学習者間すべてで有意差が見られた。つまり、レベルがより低い話者による発話にはより多くの固有名詞が含まれるということである。

表 134 学習者による NP の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
NP	141.95 (49)	NP	295.5 (102)	À	518.57 (179)	NP	3331.6 (1150)	ET	295.50 (102)	EST	185.41 (64)	EST	72.43 (25)
LE	133.26 (46)	ALLÉE	110.09 (38)	DE	408.48 (141)			C'	176.72 (61)	ET	69.53 (24)	A	72.43 (25)
SUIS	115.88 (40)	DE	86.91 (30)	EN	391.1 (135)			MAIS	98.5 (34)	JE	66.63 (23)	ET	66.63 (23)
DE	104.29 (36)	ET	75.32 (26)	NP	240.45 (83)			HESIT	86.91 (30)	Y	57.94 (20)	C'	57.94 (20)
JE	95.6 (33)	HESIT	72.43 (25)	AU	205.69 (71)			JE	78.22 (27)	HESIT	55.04 (19)	JE	55.04 (19)
À	72.43 (25)	À	63.73 (22)	LA	124.57 (43)			IL	66.63 (23)	LA	49.25 (17)	LE	46.35 (16)
EST	69.53 (24)	EST	57.94 (20)	LE	124.57 (43)			DE	63.73 (22)	C'	46.35 (16)	DE	40.56 (14)
EN	66.63 (23)	VISITÉ	52.15 (18)	ET	86.91 (30)			OUI	60.84 (21)	AI	37.66 (13)	MAIS	34.76 (12)
ET	60.84 (21)	ALLÉ	52.15 (18)	L'	81.12 (28)			J'	55.04 (19)	IL	37.66 (13)	HESIT	28.97 (10)
J'	60.84 (21)	OUI	49.25 (17)	AUX	63.73 (22)			À	52.15 (18)	J'	37.66 (13)	J'	28.97 (10)
TU	60.84 (21)	LA	46.35 (16)	INT	55.04 (19)			DONC	49.25 (17)	ÉTAIT	37.66 (13)	IL	26.07 (9)
AI	57.94 (20)	LE	43.46 (15)	A	46.35 (16)			INT	49.25 (17)	LE	34.76 (12)	PAS	23.18 (8)
C'	57.94 (20)	C'	43.46 (15)	EST	43.46 (15)			TU	40.56 (14)	EN	31.87 (11)	Y	23.18 (8)
L'	57.94 (20)	EN	34.76 (12)	VISITÉ	40.56 (14)			EN	40.56 (14)	OUI	28.97 (10)	LA	23.18 (8)
LA	55.04 (19)	ALLER	34.76 (12)	DU	31.87 (11)			OU	40.56 (14)	TRÈS	26.07 (9)	AI	23.18 (8)
HESIT	46.35 (16)	QU	34.76 (12)	QUE	31.87 (11)			ON	37.66 (13)	MAIS	26.07 (9)	POUR	23.18 (8)
INT	46.35 (16)	A	31.87 (11)	APPELLE	26.07 (9)			EST	31.87 (11)	TU	26.07 (9)	CE	23.18 (8)
A	40.56 (14)	VAIS	31.87 (11)	HESIT	23.18 (8)			D'	31.87 (11)	À	23.18 (8)	À	20.28 (7)
AU	40.56 (14)	QUE	28.97 (10)	SUR	20.28 (7)			LE	28.97 (10)	DONC	20.28 (7)		
ES	40.56 (14)	MAIS	28.97 (10)					POUR	26.07 (9)	ON	20.28 (7)		
OUI	31.87 (11)	JE	26.07 (9)					PENDANT	23.18 (8)				
PAS	28.97 (10)	TU	26.07 (9)					FW	23.18 (8)				
Y	28.97 (10)	INT	26.07 (9)					NON	23.18 (8)				
IL	26.07 (9)	MUSÉE	26.07 (9)					PAR	23.18 (8)				
PAR	26.07 (9)	S'	26.07 (9)					AUSSI	23.18 (8)				
POUR	26.07 (9)	EXEMPLE	23.18 (8)					L'	20.28 (7)				
QUE	26.07 (9)	AI	20.28 (7)					ÉTAIT	20.28 (7)				
TRÈS	23.18 (8)	PAS	20.28 (7)					OUAIS	20.28 (7)				
D'	20.28 (7)	PAR	20.28 (7)										
LES	20.28 (7)	AS	20.28 (7)										
ON	20.28 (7)	ÉTAIT	20.28 (7)										
VISITÉ	20.28 (7)												



表 135 母語話者による NP の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
C'	51.84 (34)	NP	56.41 (37)	À	115.87 (76)	NP	1248.65 (819)	ET	73.18 (48)	EST	50.31 (33)	EST	27.44 (18)
LA	33.54 (22)	DE	48.79 (32)	EN	88.43 (58)			C'	59.46 (39)	HESIT	25.92 (17)		
EST	32.02 (21)	EST	47.26 (31)	NP	86.9 (57)			IL	35.07 (23)	ÉTAIT	24.39 (16)		
NP	30.49 (20)	HESIT	47.26 (31)	L'	86.9 (57)			HESIT	35.07 (23)	C'	21.34 (14)		
DE	28.97 (19)	À	28.97 (19)	LE	68.61 (45)			QUI	33.54 (22)				
IL	27.44 (18)	ET	24.39 (16)	DE	65.56 (43)			JE	28.97 (19)				
LE	27.44 (18)	QUE	24.39 (16)	LA	53.36 (35)			POUR	24.39 (16)				
HESIT	21.34 (14)	UN	21.34 (14)	AU	47.26 (31)			OU	22.87 (15)				
				SUR	38.11 (25)								
				D'	30.49 (20)								
				DES	30.49 (20)								
				DU	30.49 (20)								
				HESIT	27.44 (18)								
				AVEC	27.44 (18)								
				MONSIEUR	25.92 (17)								
				LES	24.39 (16)								
				ET	22.87 (15)								
				UN	22.87 (15)								

#### 7.2.1.4. FW

外国語 FW については、学習者の発話と母語話者の発話で転写規則が異なるため、生起頻度の差が生まれたのは不可避であった。たとえば、学習者が英語発音で語を発した場合は FW に含めたが、母語話者の発話では英語としてその語を発言しているのか、借用語として使用しているのかの明確な判断基準がないため、フランス語として扱った。そのため、母語話者コーパスで外国語起源の単語として扱われたものは主に日本語であった。

一方、学習者コーパスの中にも多く日本語が含まれていたが、学習者が言葉に詰まった際には日本語の使用は許可されていた。そこで、日本人学習者が FW を使用するときは、主にフランス語の単語がわからない場合であったといえる。そこで、初中級学習者と上級学習者による FW の生起頻度を調べたところ、それぞれ 194 回と 76 回であることがわかった。レベルの異なる学習者間で総語数における FW の生起頻度の割合に有意差があるかどうか調べたところ、0.01%水準で有意差が確認された( $\chi^2 = 243.60$ ;  $p < .001$ )。これより、特に初中級学習者がフランス語の単語を知らない場合や思い出せない場合に FW をより多く使用する傾向にあるといえる。

表 136 学習者による FW の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
—	FW	INT	FW	ET	EST	—
—	(11)	(19)	(270)	(18)	(13)	—
—	C'	FW	—	C'	ET	—
—	(8)	(17)	—	(9)	(8)	—
—	HESIT	LE	—	INT	JE	—
—	(8)	(17)	—	(9)	(8)	—
—	—	BC	—	—	—	—
—	—	(16)	—	—	—	—
—	—	DE	—	—	—	—
—	—	(15)	—	—	—	—
—	—	EST	—	—	—	—
—	—	(8)	—	—	—	—
—	—	ET	—	—	—	—
—	—	(8)	—	—	—	—
—	—	NP	—	—	—	—
—	—	(8)	—	—	—	—

表 137 母語話者による FW の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
—	—	—	FW	—	—	—
—	—	—	(16)	—	—	—

#### 7.2.1.5. très

*très* について、学習者と母語話者の間には生起頻度に差が見られたが、初中級学習者、上級学習者、母語話者との間の生起頻度にも有意差が見られるかどうかについて、検定を行った。

表 138 *très* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
167.71 (110)	668.65 (164)	860.77 (86)	$\chi^2=202.82$ ; $p<.001$

その結果、*très* の使用頻度に有意差が見られた。そこで、ボンフェローニ補正を適応し、1.67%水準でグループ間の差があるかどうかを検定したところ、以下の結果となった。レベルに関係なく、学習者は *très* を過剰使用していることが分かる。

<i>très</i> の頻度	初級学習者 = 上級学習者 > 母語話者
-----------------	----------------------

次に、MWUs レベルで比較を行う。学習者コーパスの左側には *très* が L1 から L3 の位置まで見られるため、*très* を連続して使用している可能性が考えられる。そこで、*très* の連続使用が *très* の使用のどれくらいの割合を占めているのかについて母語話者コーパスと比較を行った。

表 139 *très* の繰り返しの頻度

	母語話者	上級学習者	初中級学習者
<i>très très</i>	10	11	9
<i>très très très</i>	0	1	1
<i>très très très très</i>	0	0	1

3 グループともに使用が見られた 2 連続使用の頻度と単独使用頻度との関係について、母語話者、上級学習者、初中級学習者との間に有意差があるかどうかを検定したところ、3 グループ間に有意差が見られなかった ( $\chi^2= 2.56$  ;  $p>.05$ )。したがって、学習者の *très* の過剰使用は連続使用が原因ではない。

次に *très* とある語句との共起関係によって *très* が過剰使用されたかどうかを分析する。学習者コーパスにも母語話者コーパスにも、*très* の左側の位置に *c'est* という MWU を多く使用していることが分かる。そこで、*c'est* と *très* の共起強度を調べた。また、学習者コーパスには *c'est* の半過去形である *c'était* も頻度上位に見られるため、これについても同様に調べた。

共起強度の比較を行った結果、*c'est* と *c'était* のどちらの場合も学習者の共起強度の方がはるかに高く、共起強度の順は初級学習者 > 上級学習者 > 母語話者であることが分かる。

表 140 *très* と *c'est*, *c'était* との共起

	母語話者		上級学習者		初中級学習者	
	頻度	共起強度	頻度	共起強度	頻度	共起強度
<i>c'est très</i>	22	60.34	45	124.26	33	164.25
<i>c'était très</i>	4	10.21	18	71.34	19	111.61

つまり、学習者は *c'est très* や *c'était très* といった MWUs を多く使用しており、その傾向は初中級学習者に顕著である。またこのことが、学習者による *très* の過剰使用の一因となった可能性があるといえる。

表 141 学習者による *très* の共起語

	L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度	
EST	89.81 (31)	C'	281.01 (97)	EST 231.76 (80)	TRÈS 724.26 (250)	BIEN 63.73 (22)	MAIS 40.56 (14)	EST 34.76 (12)
C'	43.46 (15)	ÉTAIT	37.66 (13)	ÉTAIT 121.68 (42)	BON 43.46 (15)	ET 26.07 (9)	NP 28.97 (10)	
MAIS	31.87 (11)	EST	34.76 (12)	TRÈS 78.22 (27)	INTÉRESSANT 26.07 (9)	C' 23.18 (8)	ET 20.28 (7)	
TRÈS	28.97 (10)	NP	26.07 (9)	PAS 34.76 (12)	C' 26.07 (9)	PARCE 23.18 (8)	QUE 20.28 (7)	
OUI	26.07 (9)	UN	23.18 (8)	SONT 23.18 (8)	HESIT 23.18 (8)	DE 23.18 (8)		
HESIT	23.18 (8)	TRÈS	20.28 (7)		CÉLÈBRE 20.28 (7)	POUR 23.18 (8)		
		ILS	20.28 (7)			EST 23.18 (8)		

表 142 母語話者による *très* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
—	—	35.07 (23)	38.11 (25)	167.71 (110)	38.11 (25)	—
	C'	EST	TRÈS	BIEN	—	—
		PAS				
		22.87 (15)				

### 7.2.1.6. *visité*

この語が過剰使用になった一因として、母語話者が使用したトピックと学習者が選択したトピックが異なっていた点が考えられる。母語話者はトピックの選択が自由であったのに対して、学習者は決まった 8 つのトピックから選択しなければならなかった。そのうち、「訪れた場所」を選択した学習者が多かったため「～を訪れた」に相当する (AVOIR) *visité* が多く用いられたものだと考えられる。3 グループ間の生起頻度を比較したところ、以下のような結果となった。

表 143 *visité* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
1.52 (1)	220.17 (54)	200.18 (20)	$\chi^2=137.26$ ; $p<.001$

有意差が見られたため多重比較を行ったところ、学習者グループ間には有意差が見られなかったが、学習者グループと母語話者の間には有意差が見られた(初中級学習者＝上級学習者>母語話者)。

また、学習者の R1 位置に現れた共起語を見てみると NP が最も多いが、à も多いことがわかる。しかし、動詞 VISITER は他動詞であるため、冠詞を伴わずに訪れた場所である目的語が直接後続する。一方、同じような意味で使用される「～に行った」を示す (ÊTRE) *allé(e)* は自動詞であり前置詞を伴う。このような自動詞と他動詞の使用区別について、学習者は問題があるのかもしれない。そこで、R1 位置以外にも動詞 VISITER に前置詞が用いられているケースを調べたところ、上級学習者は全 54 例中 11 例、初中級学習者は全 20 例中 4 例見られ、どちらも使用の 2 割程度がエラーであることが分かった。

表 144 学習者による *visité* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度	
TU	23.18 (8)	J	107.19 (37)	AI	112.98 (39)	214.38 (74)	
		TU	46.35 (16)	AS	37.66 (13)	NP	46.33 (14)
				A	23.18 (8)	NP	52.15 (18)
						NP	20.28 (7)
						À	26.07 (9)
						HESIT	20.28 (7)

表 145 母語話者による *visité* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度
-	-	-	VISITÉ	1.52 (1)	-	-	-

### 7.2.1.7. *beaucoup*

*beaucoup* の使用頻度の割合について、母語話者、上級学習者、初中級学習者の間で有意差があるのかどうか検定を行った。すると、以下のような結果となった。

表 146 *beaucoup* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
123.49 (81)	497.41 (122)	480.43 (48)	$\chi^2=123.21$ ; $p<.001$

有意差が見られたため、ボンフェローニ補正を適応して多重比較を行ったところ、学習者のレベルを問わず、母語話者よりも有意に多く *beaucoup* を使用していることが分かった。

<i>beaucoup</i>	初中級学習者 = 上級学習者 > 母語話者
-----------------	-----------------------

次に、MWUs レベルの比較を行う。学習者が使用した *beaucoup* を含む MWUs で特に特徴的なものは、*il y a beaucoup* または、これの半過去形である *il y avait beaucoup* であった。そこで、これらの MWUs について、*beaucoup* の出現頻度との関係について有意差があるかどうかを調べたところ、*il y a beaucoup* には有意差は見られなかったが、*il y avait beaucoup* には有意差が見られた。

表 147 *beaucoup* を含む MWUs (1)

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>il y a beaucoup</i>	8	19	9	$\chi^2=2.23$ ; $p>.05$
<i>il y avait beaucoup</i>	2	1	6	$\chi^2=14.03$ ; $p<.001$

そこで、*il y avait beaucoup* について有意水準 1,67% で多重比較を行ったところ、上級学習者と母語話者の間にのみ有意差が見られた。しかし母語話者と上級学習者について、この MWU の生起頻度自体が少なかった点は留意しなければならない。

<i>il y avait beaucoup</i>	初中級学習者 > 上級学習者
----------------------------	----------------

次に、*beaucoup* の右側に注目すると *d(e)* が後続していることが分かる。*beaucoup* の頻度と *beaucoup d'(e)* の頻度について 3 グループ間で有意差を検定したところ、有意差が見られた ( $\chi^2=10.40$ ;  $p<.01$ )。そこで、ボンフェローニ補正を行い多重比較したところ、

学習者グループ間には有意差が見られなかったが、母語話者と学習者グループの間に有意差が確認された。

<i>beaucoup d'(e)</i>	初級学習者 = 上級学習者 > 母語話者
-----------------------	----------------------

つまり *beaucoup d'(e)* という MWU はレベルを問わず、母語話者よりも学習者が有意に多く使用していたということが出来る。次に、共起強度を比較した。

表 148 *beaucoup* を含む MWUs (2)

	母語話者		上級学習者		初中級学習者	
	頻度	共起強度	頻度	共起強度	頻度	共起強度
<i>beaucoup d'(e)</i>	34	130.51	78	365.80	30	154.03

以上の表より、*beaucoup* と *d(e)* の共起強度は、上級学習者 > 初中級学習者 > 母語話者の順となった。したがって、特に上級学習者に特徴付けられる使用であるといえる。

一方、母語話者コーパスのリストに見られて、学習者コーパスのリストにはない共起語は *plus* である。*beaucoup plus* と使用することで、「より多くの」という意味になり、*plus* を強調する役割として *beaucoup* が使用されている。そこで、学習者による *beaucoup plus* の使用頻度を調べたところ、母語話者の使用では 15 例見られたのに対して、学習者コーパスには初中級学習者には 1 例も見られない上、上級学習者にも 3 例のみしか見られなかった。上級学習者と母語話者の *beaucoup* の頻度における *beaucoup plus* の頻度を調べたところ、有意差が見られた ( $\chi^2 = 15.54$ ;  $p < .001$ )。 *beaucoup* 自体の使用頻度は母語話者の方が少ないものの、*beaucoup plus* という MWU については母語話者に特徴的である。

表 149 学習者による *beaucoup* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
IL	104.29 (36)	Y	101.40 (35)	PAS	98.50 (34)	BEAUCOUP	492.50 (170)	DE	283.91 (98)	GENS	31.87 (11)	DE	34.76 (12)
J'	55.04 (19)	AI	46.35 (16)	A	95.60 (33)			D'	28.97 (10)	DE	23.18 (8)	ET	23.18 (8)
Y	34.76 (12)	A	34.76 (12)	DE	23.18 (8)								
JE	28.97 (10)	J'	31.87 (11)	AVAIT	20.28 (7)								
ON	20.28 (7)	ON	28.97 (10)	AIME	20.28 (7)								
		BEAUCOUP	28.97 (10)										

表 150 母語話者による *beaucoup* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	A	22.87 (15)	BEAUCOUP	123.49 (81)	DE	38.11 (25)	-	-	-	-
								PLUS	22.87 (15)				

### 7.2.1.8. BC

BCは相づちのことである。相づちを使用することによって対話者の会話を聞いているという態度を示すことができる。母語話者によってもかなり高い頻度で用いられているが、これについても学習者は母語話者よりも多用していることが分かった。そこで、母語話者、上級学習者、初中級学習者の間に、BCの使用頻度の割合の差があるのかどうかの検定を行ったところ、有意差が見られた。

表 151 BCのグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
1154.12 (757)	1088.60 (267)	4063.66 (406)	$\chi^2=547.92$ ; $p<.001$

そこで多重比較を行ったところ、初中級学習者と上級学習者間、そして初級学習者と母語話者間に有意差が見られた。

BC	初中級学習者>上級学習者=母語話者
----	-------------------

したがって学習者によるBCの過剰使用の原因は、初中級学習者によるBCの使用頻度が高いことが一因となっているといえる。相づちとしての*oui*の使用について、学習者の方がその使用頻度が高かったことは、本章7.2.1.2.で述べた。ここでも、学習者による相づちの使用頻度の高さを追認することが出来た。

次に、MWUsレベルでの比較を行うため、BCの左側のコンコーダンスラインに注目したところ、特にL1の位置でBCが用いられていたことが分かる。つまり、BCは繰り返し使用されると考えられる。そこで、BCの総頻度に対するBCの繰り返しの使用頻度の割合について比較を行ったところ、2連続使用( $\chi^2=148.33$ ;  $p<.001$ )、3連続使用( $\chi^2=10.31$ ;  $p<.001$ )と、4連続使用( $\chi^2=11.31$ ;  $p<.001$ )について有意差が見られた。

表 152 BCの繰り返し

	母語話者	上級学習者	初中級学習者
BC BC	172	74	29
BC BC BC	4	4	6
BC BC BC BC	1	1	3

そこで、初中級学習者、上級学習者、母語話者の3グループ間について、どのグループ間で有意差が現れるのかを多重比較を行ったところ、以下のような結果となった(ボンフェローニ補正、1.67%水準)。

BC BC	上級学習者>母語話者>初中級学習者
BC BC BC	初中級学習者=上級学習者>母語話者
BC BC BC BC	初中級学習者>母語話者

これより BC が繰り返し使用される際、特に上級学習者は 2 連続使用を、初中級学習者は 3 連続以上の連続使用を好む傾向にあるということが分かった。しかし、3 連続以上現れた生起頻度はそれほど高くないため、学習者の BC の過剰使用の主な原因は単独使用であると考えられる。

表 153 学習者による BC の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
INT 31.87 (11)	BC 69.53 (24)	BC 365.03 (126)	BC 1949.71 (673)	JE 49.25 (17)	EST 55.04 (19)	JE 26.07 (9)
BC 20.28 (7)	INT 34.76 (12)	OUI 46.35 (16)		FW 43.46 (15)	INT 28.97 (10)	EST 20.28 (7)
	OUI 23.18 (8)	INT 40.56 (14)		C' 37.66 (13)	OUI 23.18 (8)	OUI 20.28 (7)
	JE 20.28 (7)	EST 23.18 (8)		INT 37.66 (13)	JE 20.28 (7)	
		ET 20.28 (7)		J' 26.07 (9)		
				IL 23.18 (8)		
				OUI 23.18 (8)		
				ET 20.28 (7)		

表 154 母語話者による BC の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
—	—	BC 279.00 (183)	BC 1154.12 (757)	Ouais 35.07 (23)	—	—

### 7.2.1.9. *allée*

学習者は ALLER (行く)の女性形の過去分詞である *allée* を過剰使用していたが、男性形の *allé* も過剰使用リストに見られていた (コロケートは 7.2.1.12.を参照のこと)。そこで、*allée* と *allé* の両方をまとめて、初中級学習者、上級学習者、母語話者の間で使用頻度の差が見られるかどうかを分析した。

表 155 *allée* と *allé* のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>allée</i>	33.54 (22)	126.39 (31)	510.46 (51)	
<i>allé</i>	9.15 (6)	142.70 (35)	140.13 (14)	
計	42.69 (28)	269.69 (66)	650.59 (65)	$\chi^2=226.96$ ; $p<.001$

そこで、ボンフェローニ補正を用いて多重比較を行い、グループ間の有意差があるかどうかについて分析したところ、以下の通りとなった。

<i>allé(e)</i>	初中級学習者 > 上級学習者 > 母語話者
----------------	-----------------------



これより、初中級学習者が最も *allé(e)* を用いていたということが分かる。実際に、この過去分詞の過剰使用は自由会話の主題に大きく影響を受けたと考えられる。なぜなら、*visité* のケースと同様に、自由会話のテーマとして与えられていた「訪れた場所」を主題として選択した学習者が多かったことに起因している可能性があるからである。そして、特に主題の影響を受けていたのは初中級学習者であった。

表 156 学習者による *allée* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
ET 20.28 (7)	JE 153.54 (53)	SUIS 162.23 (56)	ALLÉE 237.56 (82)	EN 60.84 (21)	NP 110.09 (38)	—
	TU 49.25 (17)	ES 49.25 (17)		À 55.04 (19)		
				AU 26.07 (9)		

表 157 母語話者による *allée* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
—	—	SUIS 22.87 (15)	ALLÉE 33.54 (22)	—	—	—

#### 7.2.1.1 0. *français*

「フランス人」や「フランス語」の意味である *français* について、両コーパスのコロケートを見てみると、どちらも主題に大きく依存している語であることが分かる。特に、フランス語母語話者についてはこの傾向は顕著であり、使用の 4 割近くが *ancien* と共に使用されており、「古フランス語」についての話題が行われていたことにより *français* という単語が出現したものだと考えられる。一方、学習者は、フランス語の授業についてや、フランス語の学習について、フランス人についての話題が多かった。これより、この語の過剰使用の原因は語の使用というよりは、発話主題の偏りによるものであるといえる。

表 158 学習者による *français* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R2 頻度	R3 頻度
—	COURS 20.28 (7)	LE 104.29 (36)	FRANÇAIS 298.40 (103)	ET 37.66 (13)	—	—	—
		EN 43.46 (15)					
		DE 34.76 (12)					
		LES 26.07 (9)					

表 159 母語話者による *français* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
—	—	ANCIEN 21.34 (14)	FRANÇAIS 59.46 (39)	—	—	—

### 7.2.1.1 1. japonais

「日本語」や「日本人」の意味である *japonais* の過剰使用についても、発話の主題が大きく影響しているといえる。特に、フランス語母語話者の場合は必ずしも日本や日本語に慣れ親しんだ話者である訳ではなく、そのような話者の場合会話の中に「日本語」や「日本人」自体が生起することはまれであるといえる。

学習者の使用で興味深いのが *en japonais* という MWUs である。特に、「日本語では～という」というような使用は 7 例見られた。そして、それらの全ては初中級学習者によって使用されたものであった。

c'est quoi **en japonais** FW (JD08)

**en japonais** c'est NP (JD13)

表 160 学習者による *japonais* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度				
-	-	EN	37.66 (13)	JAPONAIS	173.82 (60)	C'	20.28 (7)	EST	20.28 (7)	-	-
-	-	LE	28.97 (10)								

表 161 母語話者による *japonais* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度
-	-	-	JAPONAIS	15.25 (10)	-	-	-

### 7.2.1.1 2. allé

*allé* の女性形である *allée* と共に分析を行った。本章 7.2.1.9. を参照のこと。

表 162 学習者による *allé* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度				
-	ON	46.35 (16)	EST	60.84 (21)	ALLÉ	141.95 (49)	À	40.56 (14)	NP	52.15 (18)	-
-	JE	40.56 (14)	SUIS	49.25 (17)			EN	23.18 (8)			

表 163 母語話者による *allé* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度
-	-	-	ALLÉ	9.15 (6)	-	-	-

### 7.2.1.1 3. *pendant*

「～の間に」を意味する *pendant* について学習者によるコロケートを見てみると、左側に *les vacances* と続くケースが多く見られることが分かる(10 例)。全てのケースについて、この MWU は「訪れた場所」が主題の時に使用されていた。つまり、主題の差が *pendant* の過剰使用の一因であることが分かる。しかし、その他の共起語では、～年間という意味を示す MWUs (ex. *pendant deux ans* : 二年間)以外は、特に定型の MWU は見られない。したがって、この過剰使用は主に WUs レベルでの過剰使用であると考えられる。

表 164 学習者による *pendant* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度		
-	-	-	NP	23.18 (8)	PENDANT	217.28 (75)	LES 34.76 (12)	ANS 28.97 (10)	ET 31.87 (11)
					DEUX 20.28 (7)	VACANCES 28.97 (10)			
					LE 20.28 (7)				

表 165 母語話者による *pendant* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度
-	-	-	PENDANT	38.11 (25)	-	-	-

### 7.2.1.1 4. *est*

*est* は動詞 ÊTRE の三人称単数現在形である。まず、母語話者、上級学習者、初中級学習者の間でこの使用頻度の差があるのかどうかを検定した。

表 166 *est* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
2942.48 (1930)	4020.06 (986)	3483.13 (348)	$\chi^2=67.47$ ; $p<.001$

有意差が見られたため、どのグループ間に有意差が見られたのかどうか多重比較を行ったところ、以下のようになった。

<i>est</i>	初中級学習者 = 上級学習者 > 母語話者
------------	-----------------------

これより、母語話者と比較して学習者グループが *est* を多用していることが分かった。

次に MWUs レベルでの比較を行う。*est* と最も共起されたコロケートは左右のとも *c'* であることが分かる。*c'est* は一般的に指示機能を持ち、*c'est* + X で「(それは) X である」のようなコピュラ文を作る。また、*c'est* はコピュラ文以外にも強調構文の一部として用

いられることがある。そこで *c'est* という MWU について、母語話者、上級学習者、初中級学習者の *est* の頻度における *c'est* の割合について比較を行った。

表 167 *c'est* のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>c'est</i>	2294.52 (1505)	3302.48 (810)	2081.87 (208)	$\chi^2=74.66$ ; $p<.001$

有意差が見られたため、どのグループの間に有意差が見られるのかについて、多重比較を行ったところ、以下のような結果となった。

<i>c'est</i>	上級学習者 > 母語話者 > 初中級学習者
--------------	-----------------------

これより、*est* が用いられる際に *c'est* として現れる頻度が最も高かったのは上級学習者であり、母語話者、初中級学習者と続いた。したがって、上級学習者による *est* の過剰使用の主な原因はこの *c'est* という MWU に起因していると考えられる。次に、L3 の位置にも *c* が見られるため、*c'est c'est* と MWU が 2 つ連続しておこった可能性があることが分かる。そこで、*est* の過剰使用がこの MWU の連続にあるのかどうかを分析するため、*est* の総頻度と 2 連続、3 連続の共起頻度の関係について比較を行った。

表 168 *c'est* の繰り返しの頻度

	母語話者	上級学習者	初中級学習者
<i>c'est c'est</i>	56	87	15
<i>c'est c'est c'est</i>	4	5	3

その結果、2 連続使用 ( $\chi^2=93.92$ ;  $p<.001$ ) と 3 連続使用 ( $\chi^2=12.70$ ;  $p<.01$ ) ともに有意差が見られた。そこで、多重比較を行ったところ、以下のような結果となった。

<i>c'est c'est</i>	上級学習者 > 母語話者 = 初中級学習者
<i>c'est c'est c'est</i>	初中級学習者 = 上級学習者 > 母語話者

これより、学習者による *est* の過剰使用の一因は主に上級学習者による *c'est* の連続使用であるといえる。

一方、初中級学習者の *est* の過剰使用の一因として考えられるのは、別にあると考えられる。実際に、初中級学習者コーパスで多く見られた MWU は *est-ce que* であった。これは、疑問文を生成する際に使用される 1 つの形である。諾否疑問文も疑問詞先行型の疑問文も *est-ce que* が用いられることがある。このため、諾否疑問文と疑問詞疑問文内での *est-ce que* の使用頻度が、*est* の頻度内でどれほどの割合を占めるのかを比較した。

表 169 *est-ce que* のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>est-ce que</i>	45.74 (30)	40.77 (10)	540.49 (54)	$\chi^2=230.97$ ; $p<.001$

有意差が見られたため、多重比較を行ったところ、以下のような結果となった。

<i>est-ce que</i>	母語話者 = 上級学習者 < 初中級学習者
-------------------	-----------------------

これより EST-CE QUE 型の疑問文の使用が初級学習者に特徴的であり、この多用が初中級学習者の *est* の過剰使用の一因となっていることが分かった。これは、先行研究 Debrock et al. (1999)の結果を支持するものであった。

表 170 学習者による *est* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
C'	359.23 (124)	EST	417.17 (144)	C'	2949.19 (1018)	EST	3864.65 (1334)	C'	370.82 (128)	QUE	214.38 (74)	C'	185.41 (64)
EST	185.41 (64)	OUI	341.85 (118)	IL	147.75 (51)			CE	260.73 (90)	C'	173.82 (60)	TU	144.85 (50)
INT	121.68 (42)	NP	185.41 (64)	QU	121.68 (42)			PAS	252.04 (87)	PEU	84.01 (29)	DE	130.37 (45)
OUI	104.29 (36)	INT	162.23 (56)	N'	110.09 (38)			TRÈS	231.76 (80)	ET	72.43 (25)	TRÈS	89.81 (31)
DE	81.12 (28)	QUE	141.95 (49)	ON	84.01 (29)			ÇA	182.51 (63)	OUI	66.63 (23)	QUE	81.12 (28)
EN	78.22 (27)	DONC	127.47 (44)	ELLE	60.84 (21)			UN	159.34 (55)	DE	66.63 (23)	NP	69.53 (24)
ET	75.32 (26)	MAIS	124.57 (43)	QUI	40.56 (14)			VRAI	112.98 (39)	HESIT	63.73 (22)	HESIT	63.73 (22)
NP	72.43 (25)	HESIT	112.98 (39)	NP	31.87 (11)			LE	110.09 (38)	ÇA	57.94 (20)	IL	60.84 (21)
HESIT	57.94 (20)	ET	107.19 (37)	S'	28.97 (10)			LA	101.4 (35)	NP	57.94 (20)	LA	60.84 (21)
PENSE	57.94 (20)	ÇA	78.22 (27)	OÙ	26.07 (9)			HESIT	92.71 (32)	À	52.15 (18)	ET	57.94 (20)
LE	55.04 (19)	CE	55.04 (19)					BIEN	75.32 (26)	UN	43.46 (15)	LE	55.04 (19)
MAIS	46.35 (16)	BC	55.04 (19)					À	69.53 (24)	IL	40.56 (14)	UN	52.15 (18)
LA	43.46 (15)	OUAIS	49.25 (17)					UNE	69.53 (24)	MAIS	40.56 (14)	MAIS	52.15 (18)
PEUT	43.46 (15)	NON	46.35 (16)					ALLÉ	60.84 (21)	ÉTAIT	40.56 (14)	OUI	37.66 (13)
À	37.66 (13)	ÊTRE	40.56 (14)					DOMMAGE	49.25 (17)	DIRE	37.66 (13)	ÉTAIT	31.87 (11)
JE	34.76 (12)	FW	37.66 (13)					IL	49.25 (17)	EN	37.66 (13)	PAS	31.87 (11)
TRÈS	34.76 (12)	PAS	34.76 (12)					POUR	43.46 (15)	TRÈS	34.76 (12)	ON	31.87 (11)
D'	31.87 (11)	DE	26.07 (9)					NP	43.46 (15)	LE	34.76 (12)	QUI	31.87 (11)
PAS	31.87 (11)	TRÈS	23.18 (8)					DIFFICILE	37.66 (13)	POUR	34.76 (12)	L'	31.87 (11)
IL	28.97 (10)	BIEN	23.18 (8)					EN	34.76 (12)	PAS	31.87 (11)	ÇA	28.97 (10)
L'	28.97 (10)	FAIT	23.18 (8)					JE	34.76 (12)	LA	31.87 (11)	À	28.97 (10)
QUE	28.97 (10)	LE	20.28 (7)					QUOI	34.76 (12)	BIEN	31.87 (11)	JE	28.97 (10)
CE	26.07 (9)	LA	20.28 (7)					ELLE	34.76 (12)	JE	28.97 (10)	EN	26.07 (9)
NON	26.07 (9)	ELLE	20.28 (7)					INT	31.87 (11)	BON	28.97 (10)	D'	26.07 (9)

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
UN	26.07 (9)	VOILÀ	20.28 (7)					BON	28.97 (10)	DONC	28.97 (10)	POUR	23.18 (8)
PARCE	23.18 (8)	JAPONAIS	20.28 (7)					SUPER	26.07 (9)	QU	26.07 (9)	DONC	23.18 (8)
BC	20.28 (7)							ASSEZ	23.18 (8)	INT	26.07 (9)	UNE	23.18 (8)
ÉTAIT	20.28 (7)							TROP	23.18 (8)	CE	23.18 (8)	PEU	20.28 (7)
								COMME	23.18 (8)	J'	20.28 (7)	INT	20.28 (7)
								L'	23.18 (8)	OUAIS	20.28 (7)	J'	20.28 (7)
								BC	23.18 (8)			PARCE	20.28 (7)
								FW	23.18 (8)			LES	20.28 (7)
								VRAIMENT	20.28 (7)				
								PLUS	20.28 (7)				
								QUI	20.28 (7)				

表 171 母語話者による *est* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
C'	102.15 (67)	HESIT	160.08 (105)	C'	2294.52 (1505)	EST	2942.48 (1930)	PAS	286.62 (188)	QUE	186 (122)	C'	94.53 (62)
EST	91.48 (60)	QUE	147.89 (97)	QUI	137.21 (90)			VRAI	213.44 (140)	C'	79.28 (52)	QUE	70.13 (46)
ET	65.56 (43)	MAIS	131.12 (86)	IL	97.57 (64)			UN	147.89 (97)	ÇA	70.13 (46)	QUI	68.61 (45)
NON	53.36 (35)	EST	117.39 (77)	ELLE	80.8 (53)			C'	103.67 (68)	QU	70.13 (46)	DE	65.56 (43)
EN	48.79 (32)	ÇA	97.57 (64)	ON	79.28 (52)			CE	96.05 (63)	PEU	47.26 (31)	IL	60.98 (40)
MAIS	47.26 (31)	OUAIS	86.9 (57)	QU	51.84 (34)			ÇA	82.33 (54)	NP	47.26 (31)	UN	44.21 (29)
HESIT	44.21 (29)	DONC	74.71 (49)	N'	48.79 (32)			LE	73.18 (48)	DE	42.69 (28)	HESIT	41.16 (27)
OUAIS	44.21 (29)	OUI	71.66 (47)	S'	47.26 (31)			UNE	71.66 (47)	HESIT	41.16 (27)	JE	39.64 (26)
L'	39.64 (26)	ET	70.13 (46)					HESIT	70.13 (46)	QUI	41.16 (27)	LA	36.59 (24)
LA	36.59 (24)	CE	70.13 (46)					BIEN	60.98 (40)	LE	39.64 (26)	ÇA	35.07 (23)
LES	33.54 (22)	NP	50.31 (33)					QUE	57.93 (38)	UN	36.59 (24)	TU	35.07 (23)
PAS	33.54 (22)	NON	47.26 (31)					À	48.79 (32)	EN	36.59 (24)	LE	33.54 (22)
LE	32.02 (21)	BON	41.16 (27)					LA	47.26 (31)	DIRE	30.49 (20)	NP	32.02 (21)
PARCE	32.02 (21)	LÀ	36.59 (24)					DES	47.26 (31)	LA	30.49 (20)	L'	28.97 (19)
DE	28.97 (19)	ENFIN	35.07 (23)					POUR	39.64 (26)	LES	30.49 (20)	HEIN	28.97 (19)
INT	28.97 (19)	FAIT	33.54 (22)					COMME	39.64 (26)	À	28.97 (19)	D'	28.97 (19)
OUI	28.97 (19)	INT	32.02 (21)					TRÈS	38.11 (25)	QUOI	25.92 (17)	PAS	28.97 (19)
TOUT	28.97 (19)	QUI	27.44 (18)					EN	38.11 (25)	MAIS	25.92 (17)	J	27.44 (18)
UN	28.97 (19)	MOI	24.39 (16)					LÀ	38.11 (25)	L'	24.39 (16)	EN	25.92 (17)
ÇA	27.44 (18)	QU	21.34 (14)					PLUS	33.54 (22)	ÊTRE	22.87 (15)	ON	25.92 (17)
NP	27.44 (18)	VOILÀ	21.34 (14)					VRAIMENT	33.54 (22)	UNE	22.87 (15)	QU'	24.39 (16)

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
QUE	27.44 (18)	SI	21.34 (14)					TROP	28.97 (19)	BIEN	22.87 (15)	LES	22.87 (15)
VRAI	27.44 (18)							TOUJOURS	27.44 (18)	MÊME	22.87 (15)	DU	22.87 (15)
CE	25.92 (17)							ASSEZ	24.39 (16)	IL	21.34 (14)	POUR	21.34 (14)
MÊME	24.39 (16)							DE	22.87 (15)			DES	21.34 (14)
								PAREIL	22.87 (15)				
								PEUT	21.34 (14)				
								QUOI	21.34 (14)				
								TOUT	21.34 (14)				

### 7.2.1.1 5. *c'*

*c'*は指示代名詞 *ce* のエリジオンがおこった形であり、原則的に動詞 **ÊTRE** の主語として用いられる。まず、動詞 **ÊTRE** の主語として生起された *c'* の頻度を母語話者、初中級学習者、上級学習者で比較する。

表 172 *c'* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
2727.51 (1789)	3930.36 (964)	2792.51 (279)	$\chi^2=90.05$ ; $p<.001$

その結果、有意差が見られた。そこで、ボンフェローニ補正を行い、多重比較を行ったところ、以下のような結果となった。

<i>c'</i>	上級学習者 > 初中級学習者 = 母語話者
-----------	-----------------------

これより、上級学習者のみがこの指示代名詞を過剰使用しているということが分かった。この結果は、先述の上級学習者による *c'est* の過剰使用に起因しているものと考えられる。次に、*c'* に後続する動詞のほとんどが *est* と *était* であったことより、現在形である *c'est* と半過去形である *c'était* の頻度の割合の比較を行う。*c'est* については前節で分析を行い、上級学習者による過剰使用を確認した。そこで、*c'était* についての結果を以下に示す。

表 173 *c'était* のグループ別頻度

	母語話者	上級学習者	初中級学習者
<i>c'était</i>	428.41 (281)	619.73 (152)	710.64 (71)

その結果、*c'était* に有意差が見られた ( $\chi^2=17.27$  ;  $p<.001$ )。ボンフェローニ補正を用いて有意水準 1.67% で多重比較を行ったところ、半過去形である *c'était* は初級学習者が有意に多く用いていることが分かった。

<i>c'était</i>	初中級学習者 > 上級学習者 = 母語話者
----------------	-----------------------

初中級学習者の *c'* の使用頻度については母語話者との有意差は見られなかったものの、*c'était* と使用される割合は母語話者よりも高いことが分かった。これまでの結果より、上級学習者は *c'est* の使用に特徴づけられ、初中級学習者の使用には *c'était* に特徴付けられるといえる。つまり、学習者は *ce+ÊTRE* の MWUs を使用する傾向が高いといえる。

*ce+ÊTRE* の使用方法について分析を行う。*ce+ÊTRE* の形は主にコピュラ文と強調文で使用される。そこで、より特殊な構文である強調構文の使用頻度について比較を行うこととする。

表 174 強調構文のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
強調構文	21	1	1	$\chi^2=10.18; p<.01$

有意差が見られたため多重比較を行ったところ、以下のような結果となった。

強調構文	母語話者 > 上級学習者
------	--------------

上級学習者は *c'est* を過剰使用しているが、強調構文の使用としては母語話者の方が多く用いていた。また、初中級学習者と母語話者との有意差はなかったが、初中級学習者によって使用された強調構文の頻度は 1 回のみであった点は看過することはできないだろう。

次に、学習者の R2 のコロケートを見てみると *très* の頻度が高いのが分かるが、特に初中級学習者による使用の中には *c'est* および *c'était* と *très* が多く共起することは *très* の項で議論した (7.2.1.5 を参照のこと)。一方、母語話者の R2 で最も頻度が高いのは否定の副詞である *pas* であった。*c* とともに *pas* が共起されるということは、否定辞 *ne* が脱落しているということである。なぜなら、否定辞 *ne* が脱落しない場合、主語は *ce* となるからである。*c'est pas* と *c'était pas* の否定辞が脱落した頻度と、*ce n'est pas* と *ce n'était pas* の否定辞が保持されていた頻度を比較した。

表 175 *ne* の保持と脱落 (*c'est, c'était* の否定文)

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>c'est pas</i>	144	58	5	$\chi^2=13.26; p<.01$
<i>c'était pas</i>	22	3	2	
合計	166	61	7	
<i>ce n'est pas</i>	18	8	5	
<i>ce n'était pas</i>	3	0	1	
合計	21	8	6	

否定辞脱落の頻度と否定辞保持の頻度について有意差が見られたため、ボンフェローニ補正を行って多重比較を行った。その結果、母語話者と上級学習者に有意差は見られな



かったが、初中級学習者と母語話者、初中級学習者と上級学習者にそれぞれ有意差が見られた。

否定辞 <i>ne</i> の保持	初中級学習者 > 上級学習者 = 母語話者
-------------------	-----------------------

初中級学習者の否定辞の脱落および保持の場合の生起頻度を見てみると、否定辞保持の割合が高いことが分かる。つまり、初中級学習者は *c'est* や *c'était* の否定文で *ne* を保持する傾向にあるといえる。そして、レベルが上がると母語話者のように *ne* の脱落が起こったと考えられる。

表 176 学習者による *c* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
C'	205.69 (71)	C'	431.66 (149)	OUI	373.72 (129)	C'	3601.02 (1243)	EST	2949.19 (1018)	TRÈS	281.01 (97)	EST	359.23 (124)
EST	185.41 (64)	EST	173.82 (60)	EST	370.82 (128)			ÉTAIT	646.04 (223)	PAS	197 (68)	ÉTAIT	86.91 (30)
JE	110.09 (38)	INT	118.78 (41)	NP	176.72 (61)					ÇA	179.62 (62)	ET	78.22 (27)
OUI	84.01 (29)	OUI	104.29 (36)	INT	170.93 (59)					UN	147.75 (51)	PEU	75.32 (26)
ET	66.63 (23)	EN	84.01 (29)	QUE	153.54 (53)					LE	121.68 (42)	OUI	72.43 (25)
HESIT	66.63 (23)	PENSE	66.63 (23)	DONC	150.65 (52)					VRAI	115.88 (40)	ÇA	57.94 (20)
INT	66.63 (23)	DE	63.73 (22)	MAIS	130.37 (45)					LA	110.09 (38)	NP	57.94 (20)
MAIS	57.94 (20)	LE	60.84 (21)	HESIT	124.57 (43)					BIEN	104.29 (36)	DE	57.94 (20)
NP	57.94 (20)	PEUT	49.25 (17)	ET	81.12 (28)					HESIT	92.71 (32)	HESIT	49.25 (17)
PAS	57.94 (20)	NP	46.35 (16)	ÇA	66.63 (23)					UNE	81.12 (28)	TRÈS	43.46 (15)
DE	55.04 (19)	HESIT	43.46 (15)	ÉTAIT	60.84 (21)					À	69.53 (24)	BIEN	43.46 (15)
DONC	52.15 (18)	ET	40.56 (14)	OUAIS	60.84 (21)					DOMMAGE	55.04 (19)	POUR	43.46 (15)
LE	52.15 (18)	À	40.56 (14)	NON	52.15 (18)					POUR	49.25 (17)	MAIS	40.56 (14)
QUE	40.56 (14)	LA	37.66 (13)	ÊTRE	46.35 (16)					EN	46.35 (16)	EN	37.66 (13)
ÊTRE	34.76 (12)	L'	37.66 (13)	VOILÀ	37.66 (13)					NP	43.46 (15)	BON	37.66 (13)
À	31.87 (11)	ÉTAIT	34.76 (12)	BC	37.66 (13)					BON	43.46 (15)	DIRE	34.76 (12)
EN	28.97 (10)	MAIS	28.97 (10)	PAS	31.87 (11)					COMMENT	43.46 (15)	UN	34.76 (12)
LA	28.97 (10)	D'	28.97 (10)	TRÈS	26.07 (9)					DIFFICILE	40.56 (14)	LA	34.76 (12)
UN	26.07 (9)	JE	26.07 (9)	FW	26.07 (9)					JE	40.56 (14)	IL	34.76 (12)
BIEN	20.28 (7)	NON	26.07 (9)	FAIT	23.18 (8)					VRAIMENT	40.56 (14)	LE	31.87 (11)
		PAS	23.18 (8)	JAPONAIS	20.28 (7)					QUOI	40.56 (14)	DIFFICILE	31.87 (11)
		PARCE	23.18 (8)	LÀ	20.28 (7)					TROP	34.76 (12)	JE	31.87 (11)
		TRÈS	23.18 (8)	BIEN	20.28 (7)					INT	31.87 (11)	À	28.97 (10)
		QUE	20.28 (7)							QUAND	31.87 (11)	J'	28.97 (10)
		UN	20.28 (7)							SUPER	26.07 (9)	QUE	28.97 (10)
										L'	26.07 (9)	INT	26.07 (9)
										ASSEZ	26.07 (9)	OUAIS	23.18 (8)
										COMME	26.07 (9)	PARCE	20.28 (7)
										FW	23.18 (8)		
										ELLE	23.18 (8)		
										INTÉRESSANT	23.18 (8)		
										IL	20.28 (7)		
										JUSTE	20.28 (7)		
										CE	20.28 (7)		

表 177 母語話者による *c* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
EST	94.53 (62)	C'	118.92 (78)	QUE	181.43 (119)	C'	2727.51 (1789)	EST	2294.52 (1505)	PAS	254.61 (167)	QUE	135.69 (89)
C'	70.13 (46)	EST	79.28 (52)	HESIT	155.51 (102)			ÉTAIT	428.41 (281)	VRAI	211.92 (139)	EST	102.15 (67)
ET	48.79 (32)	OUAIS	53.36 (35)	MAIS	135.69 (89)					UN	163.13 (107)	ÇA	71.66 (47)
JE	47.26 (31)	NON	51.84 (34)	EST	103.67 (68)					LE	86.9 (57)	PEU	54.89 (36)
HESIT	38.11 (25)	ET	50.31 (33)	OUAIS	99.1 (65)					ÇA	80.8 (53)	NP	51.84 (34)
MAIS	38.11 (25)	MAIS	47.26 (31)	ÇA	89.95 (59)					UNE	80.8 (53)	QU	51.84 (34)
EN	32.02 (21)	EN	45.74 (30)	DONC	80.8 (53)					HESIT	60.98 (40)	UN	47.26 (31)
TU	27.44 (18)	HESIT	44.21 (29)	OUI	73.18 (48)					BIEN	59.46 (39)	LE	38.11 (25)
DE	25.92 (17)	PARCE	41.16 (27)	ET	68.61 (45)					QUE	56.41 (37)	DE	38.11 (25)
OUAIS	24.39 (16)	LES	33.54 (22)	NP	59.46 (39)					LA	53.36 (35)	LA	36.59 (24)
L'	22.87 (15)	INT	33.54 (22)	NON	54.89 (36)					DES	50.31 (33)	QUI	36.59 (24)
OUI	22.87 (15)	VRAI	30.49 (20)	BON	45.74 (30)					VRAIMENT	47.26 (31)	HESIT	32.02 (21)
QUE	22.87 (15)	L'	28.97 (19)	INT	38.11 (25)					À	44.21 (29)	LES	32.02 (21)
À	21.34 (14)	PAS	28.97 (19)	ENFIN	36.59 (24)					POUR	42.69 (28)	DIRE	30.49 (20)
LA	21.34 (14)	OUI	27.44 (18)	LÀ	35.07 (23)					COMME	36.59 (24)	BIEN	28.97 (19)
		ÇA	27.44 (18)	FAIT	32.02 (21)					TRÈS	35.07 (23)	QUOI	25.92 (17)
		DE	25.92 (17)	MOI	27.44 (18)					TROP	32.02 (21)	MAIS	25.92 (17)
		LA	24.39 (16)	VOILÀ	22.87 (15)					CE	28.97 (19)	UNE	24.39 (16)
		LE	24.39 (16)	SI	22.87 (15)					QUOI	28.97 (19)	À	24.39 (16)
		MÊME	24.39 (16)							ASSEZ	27.44 (18)	ÊTRE	21.34 (14)
		TU	22.87 (15)							PLUS	25.92 (17)		
		NP	21.34 (14)							L'	24.39 (16)		
										DE	24.39 (16)		
										TOUT	22.87 (15)		
										PAREIL	22.87 (15)		
										LES	21.34 (14)		
										LÀ	21.34 (14)		
										TOUJOURS	21.34 (14)		

### 7.2.1.1 6. *musique*

この語については、主題に大きく依存する語であるといえる。特に、学習者による *musique* とのコロケートを見てみると、*classique* が左側に続く例が 7 例見られ、これは *musique* の生起頻度の 2 割を占める。さらに、この語は学習者のうち 7 名によって使用されているが、ある一つのペアによって *musique* が使用された生起頻度は 28 例で、*musique* の生起頻度の 8 割にも上った。つまり、このペアの主題が音楽であり、そのことによってこの語の過剰使用が引き起こされたのだと考えられる。

表 178 学習者による *musique* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度
-	-	LA	43.46 (15)	MUSIQUE	101.4 (35)	CLASSIQUE	20.28 (7)
		DE	23.18 (8)				

表 179 母語話者による *musique* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度
-	-	-	-	MUSIQUE	4.57 (3)	-	-

### 7.2.1.1 7. *j'*

*j'*は人称代名詞主語 *je* であるが、後続の語が母音または無音の *h* である際にエリジオンがおこった形である。そこで、*je* も過剰使用語として抽出されているため、まずはこの二つの形について、母語話者、上級学習者、初中級学習者に使用頻度の違いが見られるか、また合計頻度に違いがあるかどうかについて比較を行った。

表 180 *j(e)* のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>j</i>	1169.37 (767)	2140.50 (525)	1801.62 (180)	$\chi^2=129.60$ ; $p<.001$
<i>je</i>	1870.68 (1227)	2181.27 (535)	3553.20 (355)	$\chi^2=119.27$ ; $p<.001$
合計	3040.05 (1994)	4321.77 (1060)	5354.82 (535)	$\chi^2=184.31$ ; $p<.001$

いずれも有意差が見られたため、ボンフェローニ補正を行い多重比較を行ったところ、以下の通りとなった。

<i>j'</i>	初中級学習者 = 上級学習者 > 母語話者
<i>je</i>	初中級学習者 > 上級学習者 > 母語話者
<i>j'+je</i>	初中級学習者 > 上級学習者 > 母語話者

これより、レベルを問わず学習者は一人称代名詞主語を過剰使用しているということが分かる。また、その傾向は初中級学習者に顕著であった。

次に MWUs レベルでの比較を行うため、右側のコロケートを見たところ、母語話者は R1 位置で中性代名詞 *en* と *y* を多く用いていることが分かる。学習者のコーパスにも R1 の位置に *y* が見られているが、*en* は頻度上位の語ではなかった。そこで、*j'* と共に使用されたこれらの中性代名詞の頻度について比較を行う。

表 181 中性代名詞との共起頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>y</i>	36.59 (24)	36.69 (9)	80.07 (8)	$\chi^2=4.39$ ; $p>.05$
<i>en</i>	41.16 (27)	4.08 (1)	0 (0)	NA

その結果、*y* との共起頻度については有意差が見られなかった。しかし、*j'y* に後続する語について調べたところ、使用方法に違いが見られることが分かった。まず、学習者コーパスには *j'y* の連続使用が 4 例見られた(上級学習者 1 例、初中級学習者 3 例)。そのうちの 1 例は *j'y* の 3 連続使用であり、初中級学習者によるものであった。次に、*j'y* と一緒に使用される動詞にも相違が見られた。どのグループとも動詞 ALLER が後続していたが、その内訳は以下の通りであった。

表 182 *j'y* の後続語

<i>j'y</i> の後続語	母語話者	上級学習者	初中級学習者
<i>vais</i>	5	2	1
<i>suis allé(e)</i>	2	4	0
<i>aïlle</i>	2	1	0

また、母語話者コーパス内には動詞 CROIRE の活用形である *crois* も 4 例見られたが、学習者コーパス内には *j'y crois* という MWU は一例も見られなかった。母語話者によるこの MWU の使用の全ては、*j'y crois* の後ろに否定辞 *pas* を用いたものであった。

Ça en fait **j'y crois pas** trop (FL18)

*en* については初中級学習者による頻度が 0 であったが、上級学習者も *j'en* という MWU は一例しか見られなかった。母語話者と上級学習者の間に有意差があるかどうか検定したところ、有意差が見られた( $\chi^2=16.30$ ;  $p<.001$ )。母語話者に一番多く見られた MWU は *en avoir marre*(飽き飽きする)という表現であり、4 例見られた。

**j'en ai** fait des plus gros parce qu'à la fin **j'en avais marre** (FL01)

tu pars avec tes parents tu fais plein de voyages à quinze ans tu vois **j'en ai marre** (FL03)

以上までを総合すると、学習者は *j* を過剰使用しているが、この中心語と中性代名詞との共起頻度は母語話者に比べて少ないということがいえる。

そこで、学習者の *j* の多用の原因を調べるため、左右のコロケートで一番多かった動詞が動詞 AVOIR の一人称単数現在形の *ai* であることに注目した。また、左側のコロケートには *j* が多く見られることもあり、*j'ai j'ai* という *j'ai* が連続して使用されたケースについて頻度の比較を行う。

表 183 *j'ai* の繰り返しの頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>j'ai j'ai</i>	12	22	9	$\chi^2=22.07$ ; $p<.001$
<i>j'ai j'ai j'ai</i>	0	4	0	NA

以上の通り、2連続使用について3グループ間に有意差が見られたため多重比較を行った。その結果、以下の通りとなった。

<i>j'ai j'ai</i>	初中級学習者 = 上級学習者 > 母語話者

したがって、学習者はレベルを問わず *j'ai* の連続使用が多いことが分かった。さらに上級学習者には *j'ai* の3連続使用が見られた。これより、*j* の過剰使用の一因は *j'ai* という MWU の連続使用であるといえる。

その他の原因について考えられるのは、主題の影響である。学習者の会話の主題は、「観た映画」「読んだ本」「訪れた場所」「大切な思い出」など、過去に関連する主題が多い。このことより、過去に関する話題では過去形の使用が必須となり、複合過去形を形成する助動詞の役割を持つ *ai* が *j* に後続することによって、*j* が過剰使用された可能性が高い。実際に R2 を見てみると、学習者の R2 には調整頻度 20 以上の過去分詞が多く現れていることが分かる。

表 184 学習者による *j* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
J'	107.19 (37)	J'	199.9 (69)	AI	141.95 (49)	J'	1752.71 (605)	AI	392.00 (392)	VISITÉ	107.19 (37)	AI	110.09 (38)
INT	60.84 (21)	ET	72.43 (25)	ET	107.19 (37)			AIME	159.34 (55)	PAS	78.22 (27)	À	72.43 (25)
OUI	57.94 (20)	AI	69.53 (24)	HESIT	104.29 (36)			ÉTAIS	133.26 (46)	VU	66.63 (23)	LE	66.63 (23)
ET	49.25 (17)	OUI	55.04 (19)	QUAND	89.81 (31)			AVAIS	66.63 (23)	BIEN	52.15 (18)	UN	60.84 (21)
AI	43.46 (15)	JE	49.25 (17)	QUE	89.81 (31)			Y	49.25 (17)	JE	49.25 (17)	NP	60.84 (21)
DE	37.66 (13)	HESIT	46.35 (16)	OUI	78.22 (27)			HABITAIS	26.07 (9)	ENTENDU	34.76 (12)	BEAUCOUP	55.04 (19)
HESIT	37.66 (13)	EN	46.35 (16)	DONC	78.22 (27)			HABITE	23.18 (8)	HESIT	34.76 (12)	HESIT	52.15 (18)
JE	31.87 (11)	INT	37.66 (13)	JE	69.53 (24)					EN	34.76 (12)	DE	52.15 (18)
MAIS	31.87 (11)	MAIS	37.66 (13)	MAIS	66.63 (23)					BEAUCOUP	31.87 (11)	LA	46.35 (16)
C'	28.97 (10)	NP	37.66 (13)	INT	63.73 (22)					FAIT	31.87 (11)	ET	34.76 (12)
EN	28.97 (10)	MOI	37.66 (13)	NP	55.04 (19)					OUBLIÉ	31.87 (11)	JE	23.18 (8)
NP	28.97 (10)	PARCE	26.07 (9)	MOI	40.56 (14)					DÉJÀ	28.97 (10)	QUE	23.18 (8)
DONC	26.07 (9)	DONC	23.18 (8)	AUSSI	37.66 (13)					LE	28.97 (10)	UNE	23.18 (8)
PAS	26.07 (9)	EST	20.28 (7)	BC	26.07 (9)					COMMENCÉ	26.07 (9)	AIME	23.18 (8)
QUE	26.07 (9)			AIME	20.28 (7)					MANGÉ	26.07 (9)	Y	20.28 (7)
LA	23.18 (8)									TROUVÉ	23.18 (8)	EN	20.28 (7)
EST	20.28 (7)									UN	23.18 (8)	INT	20.28 (7)
L'	20.28 (7)									L'	20.28 (7)	LES	20.28 (7)
										À	20.28 (7)		

表 185 母語話者による *j* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
EST	27.44 (18)	HESIT	32.02 (21)	QUE	138.74 (91)	J	1169.37 (767)	AI	661.68 (434)	PAS	120.44 (79)	DE	33.54 (22)
JE	25.92 (17)	PARCE	32.02 (21)	HESIT	76.23 (50)			AVAIS	111.3 (73)	BIEN	50.31 (33)	ÇA	33.54 (22)
HESIT	22.87 (15)	J'	30.49 (20)	MOI	68.61 (45)			ÉTAIS	76.23 (50)	FAIT	50.31 (33)	HESIT	30.49 (20)
J'	22.87 (15)	EN	30.49 (20)	ET	42.69 (28)			AIME	60.98 (40)	TROUVÉ	35.07 (23)	UN	28.97 (19)
		ET	25.92 (17)	MAIS	39.64 (26)			EN	41.16 (27)	AI	35.07 (23)	LE	28.97 (19)
		CE	24.39 (16)	QUAND	25.92 (17)			Y	36.59 (24)	EU	30.49 (20)	LA	27.44 (18)
				FAIT	22.87 (15)			AURAI	30.49 (20)	VU	27.44 (18)	UNE	24.39 (16)
				AI	21.34 (14)					UN	25.92 (17)	AI	21.34 (14)
										HESIT	22.87 (15)		
										DIT	22.87 (15)		

### 7.2.1.1 8. difficile

まず、WU レベルでのグループ別の違いがあるかどうかを調べるため、それぞれ母語話者コーパス、上級学習者コーパス、初中級学習者コーパスにおける *difficile* の生起頻度を比較したところ、以下の通り有意差が見られた。

表 186 *difficile* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
15.25 (10)	106.01 (26)	200.18 (20)	$\chi^2=67.61$ ; $p<.001$

多重比較を行い、どのグループ間で有意差があるのかどうかを調べたところ、以下の表の通りとなった(ボンフェローニ補正、 $p<.0167$ )。

<i>difficile</i>	初中級学習者 = 上級学習者 > 母語話者

したがって、学習者内におけるレベルの差は関係なく学習者によって *difficile* が多用されていることが分かる。

次に、MWUs レベルの比較を行う。学習者コーパスのコロケートを見ると、*c'est* や *c'était* と共に *difficile* が多く使用されていることが分かる。そこで、*difficile* の頻度における *c'est* や *c'était* との共起頻度の割合を比較した。また、*c'est* および *c'était* と *difficile* の間に副詞が含まれている場合や否定文として使用されていた場合についても、数に含めた。まず、*c'est* について有意差検定を行ったところ、有意差は見られなかった ( $\chi^2=1.68$  ;  $p>.05$ )。 *c'était* については、母語話者の生起頻度が0であったため母語話者との有意差検定を行わなかったが、上級学習者の頻度の割合と学習者の頻度の割合には有意差が見られなかった ( $\chi^2=0.94$  ;  $p>.05$ )。

表 187 *difficile* と *c'est*, *c'était* との共起頻度

	母語話者	上級学習者	初中級学習者
<i>c'est</i> + <i>difficile</i>	7	12	11
<i>c'était</i> + <i>difficile</i>	0	7	3

以上の結果より、*difficile* の頻度はレベルを問わず学習者の方が有意に高いが、*c'est* と共起される際の頻度は母語話者とあまり変わらない。つまり、頻度は異なるが使用方法には違いが見られず、WUs レベルでの過剰使用であると考えられる。

表 188 学習者による *difficile* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度
C' 31.87 (11)	C' 40.56 (14)	EST 37.66 (13)	DIFFICILE	133.26 (46)	DE 26.07 (9)	-	-
	ÉTAIT 23.18 (8)	PEU 20.28 (7)					



表 189 母語話者による *difficile* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度
-	-	-	DIFFICILE	15.25 (10)	-	-	-

### 7.2.1.1.9. *je*

一人称代名詞主語である *je* の過剰使用は既に *j* の項(7.2.1.17)で分析を行い、母語話者、上級学習者、初中級学習者の順にこの代名詞が現れる頻度が高くなることが分かった。そこで、本項ではその原因を探る。

中心語である *je* の左側を見てみると、L1 から L3 で *je* が現れていることが分かる。そこで、*je* の連続使用について比較を行う。*je* の 2 回連続から 4 回連続までの使用を比較したところ、以下の通りとなった。

表 190 *je* の繰り返しの頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>je je</i>	21	42	33	$\chi^2=115.00; p<.001$
<i>je je je</i>	0	4	0	NA
<i>je je je je</i>	0	1	0	NA

3 連続使用と 4 連続使用については、上級学習者のみにしか見られなかった。また、2 連続使用については、有意差が見られた。2 連続使用について、多重比較を行ったところ、以下のような結果となった。

<i>je je</i>	初中級学習者 = 上級学習者 > 母語話者
--------------	-----------------------

その結果、学習者は初中級学習者も上級学習者も *je* の 2 連続使用を過剰使用していることが分かった。これが *je* の過剰使用の一因であると考えられる。

次に、*je* と共起する語の傾向について分析する。*je* と共起される調整頻度 20 以上の動詞の頻度についてまとめた表が以下である。

表 191 *je* との共起頻度(1)

順位	母語話者		学習者	
	コロケート	調整頻度 (頻度)	コロケート	調整頻度 (頻度)
1	<i>sais</i>	286.62 (188)	<i>suis</i>	402.69 (139)
2	<i>suis</i>	172.28 (113)	<i>pense</i>	260.73 (90)
3	<i>crois</i>	120.97 (80)	<i>vais</i>	170.93 (59)
4	<i>vais</i>	94.53 (62)	<i>sais</i>	144.85 (50)
5	<i>pense</i>	86.90 (57)	<i>crois</i>	69.53 (24)
6	<i>trouve</i>	57.93 (38)	<i>voudrais</i>	69.53 (24)
7	<i>fais</i>	35.07 (23)	<i>veux</i>	49.25 (17)
8	<i>veux</i>	30.49 (23)	<i>peux</i>	49.25 (17)
9	<i>dis</i>	28.97 (19)	<i>dois</i>	49.25 (17)
10	<i>peux</i>	25.92 (17)	<i>connais</i>	43.46 (15)
11	<i>vois</i>	25.92 (17)	<i>voulais</i>	34.76 (12)
12			<i>pensais</i>	26.07 (9)
13			<i>trouve</i>	23.18 (8)
14			<i>parle</i>	20.28 (7)

頻度 20 以上の動詞は、母語話者コーパスに 11、学習者コーパスに 14 あったが、そのうちの 8 つは 2 つのコーパスに共通するものであった。そこで、*je* の総頻度あたりに対する共通している 8 つの動詞の生起頻度について、母語話者、上級学習者、初中級学習者の頻度に有意差があるかどうかを調べたところ、以下の通りとなった。

表 192 *je* との共起頻度(2)

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量 有意確率	
<i>sais</i>	286.62 (188)	154.93 (38)	120.11 (12)	$\chi^2=51.64 ; p<.001$	初中級 = 上級 < 母語話者
<i>suis</i>	172.28 (113)	269.09 (66)	630.57 (63)	$\chi^2=20.40 ; p<.001$	初中級 > 母語話者
<i>crois</i>	121.97 (80)	77.47 (19)	50.05 (5)	$\chi^2=18.24 ; p<.001$	初中級 = 上級 < 母語話者
<i>vais</i>	94.53 (62)	138.62 (34)	250.23 (25)	$\chi^2=2.57 ; p>.05$	有意差なし
<i>pense</i>	86.90 (57)	236.47 (58)	320.29 (32)	$\chi^2=24.96 ; p<.001$	初中級 = 上級 > 母語話者
<i>trouve</i>	57.93 (38)	32.62 (8)	0 (0)	NA	上級 = 母語話者
<i>veux</i>	30.49 (23)	40.77 (10)	70.06 (7)	$\chi^2=0.02 ; p>.05$	有意差なし
<i>peux</i>	25.92 (17)	44.85 (11)	60.05 (6)	$\chi^2=1.08 ; p>.05$	有意差なし

レベルを問わず、母語話者よりも学習者による *je* との共起頻度が高かった動詞は *pense* のみであった。また、初中級学習者のみ *suis* の共起頻度が母語話者よりも高かった。しかし、その他の 6 つの単語については、母語話者の共起頻度の方が高いか、有意差が見られなかった。

次に、片方のコーパスにおいて調整頻度 20 以上であったが、もう一方のコーパスの上位には現れなかった動詞について比較を行う。

表 193 *je* との共起頻度(3)

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量 有意確率	
<i>fais</i>	35.07 (23)	20.39 (5)	0 (0)	NA	上級=母語話者
<i>dis</i>	28.97 (19)	0 (0)	0 (0)	NA	上級=初中級<母語話者
<i>vois</i>	25.92 (17)	4.08 (1)	10.01 (1)	$\chi^2=7.84$ ; $p<.05$	有意差なし
<i>voudrais</i>	6.10 (4)	40.77 (10)	20.02 (2)	$\chi^2=12.04$ ; $p<.001$	上級>母語話者
<i>dois</i>	6.10 (4)	32.62 (8)	90.08 (9)	$\chi^2=15.53$ ; $p<.001$	初中級=上級>母語話者
<i>voulais</i>	7.62 (5)	48.93 (12)	0 (0)	NA	上級>母語話者
<i>pensais</i>	4.57 (3)	28.54 (7)	20.02 (2)	$\chi^2=7.48$ ; $p<.05$	上級>母語話者
<i>parle</i>	4.57 (3)	20.39 (5)	20.02 (2)	$\chi^2=3.85$ ; $p<.05$	有意差なし

母語話者による使用頻度がより高かったものは、*dis* だけであった。学習者全体で母語話者よりも共起頻度が高かった動詞は *dois* であり、*voudrais*, *voulais*, *pensais* については、上級学習者による共起頻度が母語話者を上回っていた。

以上の点を総合すると、学習者全般で母語話者よりも有意に頻度が高かったものは、*je pense*, *je dois* であった。初中級学習者によって多用されていたのは *je suis* であった。また、上級者によって多く使用されていた MWUs は *je voudrais*, *je voulais* そして *je pensais* であった。これらの MWUs の使用により、学習者による *je* の過剰がおこった可能性は高い。

しかし一方で、母語話者の方が多く *je* と共に使用する動詞もあった。*je sais*, *je crois*, *je dis* という MWUs は母語話者の方が学習者よりも多く使用していた。特に、*je sais* と *je crois* は別の共起語とともに多く使用されていた。*je sais* の頻度は 188 回であったが、そのうち 124 回は *pas* と共起されていた。つまり、母語話者は *je sais pas* という MWU を多く使用していた。これは、*je ne sais pas* の *ne* の脱落の形である。また、*je crois* は従位接続詞 *que* とともに用いられ、「～と思う」という文として用いられた例が 29 例見られた。このように、*je* 自体の頻度は学習者の方が有意に多いが、母語話者に特徴付けられた *je* を含む MWUs もあることは注目に値する。

表 194 学習者による *je* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
JE	133.26 (46)	JE	176.72 (61)	JE	249.15 (86)	JE	2578.36 (890)	SUIS	402.69 (139)	QUE	211.48 (73)	PAS	185.41 (64)
HESIT	81.12 (28)	MAIS	107.19 (37)	MAIS	170.93 (59)			PENSE	260.73 (90)	PAS	202.79 (70)	C'	110.09 (38)
ET	66.63 (23)	HESIT	92.71 (32)	QUE	139.06 (48)			VAIS	170.93 (59)	ALLÉE	153.54 (53)	À	104.29 (36)
INT	63.73 (22)	ET	78.22 (27)	HESIT	124.57 (43)			NE	153.54 (53)	AI	121.68 (42)	NP	95.6 (33)
NP	55.04 (19)	NP	66.63 (23)	ET	112.98 (39)			SAIS	144.85 (50)	SAIS	52.15 (18)	EN	92.71 (32)
OUI	55.04 (19)	INT	60.84 (21)	OUI	107.19 (37)			N'	84.01 (29)	SUIS	49.25 (17)	HESIT	57.94 (20)
MAIS	49.25 (17)	OUI	49.25 (17)	INT	95.6 (33)			ME	72.43 (25)	J'	49.25 (17)	LE	52.15 (18)
EN	40.56 (14)	J'	49.25 (17)	QUAND	86.91 (30)			CROIS	69.53 (24)	LE	49.25 (17)	LA	49.25 (17)
DONC	37.66 (13)	PARCE	43.46 (15)	DONC	81.12 (28)			J'	69.53 (24)	ALLÉ	40.56 (14)	AI	46.35 (16)
QUE	37.66 (13)	QUE	40.56 (14)	NP	78.22 (27)			VOUDRAIS	69.53 (24)	AIME	37.66 (13)	EST	34.76 (12)
C'	31.87 (11)	C'	40.56 (14)	MOI	69.53 (24)			VEUX	49.25 (17)	À	37.66 (13)	QUE	31.87 (11)
EST	28.97 (10)	DONC	34.76 (12)	AI	49.25 (17)			PEUX	49.25 (17)	PEUX	34.76 (12)	J'	31.87 (11)
BC	26.07 (9)	EN	28.97 (10)	BC	49.25 (17)			DOIS	49.25 (17)	EN	34.76 (12)	DE	31.87 (11)
À	23.18 (8)	EST	28.97 (10)	NON	37.66 (13)			CONNAIS	43.46 (15)	HESIT	31.87 (11)	L'	31.87 (11)
J'	23.18 (8)	AI	26.07 (9)	EST	34.76 (12)			L'	37.66 (13)	DE	28.97 (10)	AU	31.87 (11)
LA	23.18 (8)	MOI	26.07 (9)	APRÈS	31.87 (11)			VOULAIS	34.76 (12)	VAIS	28.97 (10)	SUIS	28.97 (10)
PARCE	23.18 (8)	D'	26.07 (9)	PAS	28.97 (10)			M'	31.87 (11)	FAIRE	26.07 (9)	BEAUCOUP	28.97 (10)
PAS	20.28 (7)	PEUT	26.07 (9)	NE	28.97 (10)			PENSAIS	26.07 (9)	C'	26.07 (9)	UN	26.07 (9)
SUIS	20.28 (7)	À	23.18 (8)	SI	28.97 (10)			HESIT	26.07 (9)	NP	26.07 (9)	NE	26.07 (9)
		NON	23.18 (8)	MAINTENANT	26.07 (9)			TROUVE	23.18 (8)	Y	23.18 (8)	PLUS	26.07 (9)
		BC	20.28 (7)	ÊTRE	26.07 (9)			TE	20.28 (7)	PARLER	23.18 (8)	VAIS	23.18 (8)

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
		PAS	20.28 (7)	VAIS	23.18 (8)			PARLE	20.28 (7)	L'	23.18 (8)	ET	20.28 (7)
		FW	20.28 (7)							ALLER	20.28 (7)		
										SOUVIENS	20.28 (7)		
										BC	20.28 (7)		
										UN	20.28 (7)		

表 195 母語話者による *je* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
JE	64.03 (42)	HESIT	48.79 (32)	QUE	195.15 (128)	JE	1870.68 (1227)	SAIS	286.62 (188)	PAS	274.43 (180)	PAS	88.43 (58)
IL	42.69 (28)	EN	38.11 (25)	HESIT	114.34 (75)			SUIS	172.28 (113)	QUE	140.26 (92)	ÇA	48.79 (32)
EST	39.64 (26)	JE	35.07 (23)	MOI	103.67 (68)			CROIS	121.97 (80)	AI	82.33 (54)	C'	47.26 (31)
DE	28.97 (19)	ET	33.54 (22)	ET	77.75 (51)			ME	94.53 (62)	PLUS	59.46 (39)	À	45.74 (30)
ET	25.92 (17)	MAIS	30.49 (20)	MAIS	65.56 (43)			VAIS	94.53 (62)	QU'	51.84 (34)	IL	45.74 (30)
HESIT	25.92 (17)	QUE	30.49 (20)	ENFIN	54.89 (36)			PENSE	86.90 (57)	SAIS	45.74 (30)	HESIT	42.69 (28)
EN	22.87 (15)	PARCE	27.44 (18)	ÇA	35.07 (23)			NE	68.61 (45)	SUIS	42.69 (28)	EN	36.59 (24)
		SAIS	22.87 (15)	SI	33.54 (22)			TROUVE	57.93 (38)	DIS	35.07 (23)	SI	32.02 (21)
		CE	22.87 (15)	JE	32.02 (21)			L'	56.41 (37)	EN	33.54 (22)	LE	28.97 (19)
		LA	21.34 (14)	LÀ	32.02 (21)			LES	42.69 (28)	LE	28.97 (19)	J	25.92 (17)
				QUAND	32.02 (21)			LUI	39.64 (26)	UN	25.92 (17)	LA	24.39 (16)
				Ouais	30.49 (20)			FAIS	35.07 (23)	DIRE	22.87 (15)	PLUS	22.87 (15)
				NON	30.49 (20)			VEUX	30.49 (20)	DE	21.34 (14)	UN	21.34 (14)
				NP	28.97 (19)			DIS	28.97 (19)			DE	21.34 (14)
				OUI	24.39 (16)			TE	27.44 (18)			LES	21.34 (14)
				DONC	22.87 (15)			PEUX	25.92 (17)				
				BON	21.34 (14)			M'	25.92 (17)				
								VOIS	24.39 (16)				
								LE	22.87 (15)				
								T'	21.34 (14)				

### 7.2.1.2 0. *comment*

まず、*comment*の生起頻度について、母語話者、上級学習者、初中級学習者の間で有意差があるかどうか検定を行った。

表 196 *comment*のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
71.66 (47)	252.78 (62)	200.18 (20)	$\chi^2=49.89$ ; $p<.001$

その結果、以上の通り有意差が見られたため、ボンフェローニ補正を適応し、多重比較を行ったところ、上級学習者と初中級学習者の間には有意差が見られなかったが、学習者グループと母語話者の間には有意差が見られた。

<i>comment</i>	初中級学習者 = 上級学習者 > 母語話者
----------------	-----------------------

つまり、*comment*は全体的に学習者によって過剰使用されていたといえる。

次に、MWUsレベルの使用についての差を調べるため、学習者のコロケートに注目すると、*comment dire*というMWUsが特徴的であることが分かる。そこでこのMWUが*comment*の総頻度に対してどれくらいの割合を占めるのかをグループ間で比較した。

表 197 *comment dire*のグループ別頻度

	母語話者	上級学習者	初中級学習者
<i>comment dire</i>	6	37	5

その結果、グループの間に有意差が見られた( $\chi^2=26.69$ ;  $p<.001$ )。さらにボンフェローニ補正を行い、有意水準1.67%で多重比較を行ったところ、以下のような結果となった。

<i>comment dire</i>	上級学習者 > 初中級学習者 = 母語話者
---------------------	-----------------------

以上の結果より、*comment dire*は上級学習者に特徴づけられるMWUである。*comment dire*は言いたい単語や文が出てこない際に使用されることが多い。会話の流れを止めないためなどの方略的使用といえる。上級学習者による過剰使用は、この方略的使用を多用していたからだと考えられる。一方、初中級学習者は*comment dire*をそれほど多く使用していない。しかし、母語話者に比べて*comment*を多く使用しているのは、別の原因があると考えられる。そこで、初中級学習者のコンコーダンスラインを見たところ、*c'était comment*が多く使用されていることが分かった。

表 198 *c'était comment* のグループ別頻度

	母語話者	上級学習者	初中級学習者
<i>c'était comment</i>	3	4	9

そこで、グループ間におけるこの MWU の使用頻度について比較を行ったところ、有意差が見られた ( $\chi^2=23.15$ ;  $p<.001$ )。そして、どのグループの間に有意差が見られるのかを多重比較を行って検定したところ、初級学習者が他の 2 グループよりも有意に多くこの MWU を使用していることが分かった。

<i>c'était comment</i>	初中級学習者 > 上級学習者 = 母語話者
------------------------	-----------------------

つまり、上級学習者が *comment* を多用する一因は *comment dire* という MWU であり、初中級学習者による過剰使用の一因は *c'était comment* にあるといえる。

表 199 学習者による *comment* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
—	C 43.46 (15)	ÉTAIT 37.66 (13)	COMMENT 249.15 (86)	DIRE 121.68 (42)	—	—
				ÇA 23.18 (8)		

表 200 母語話者による *comment* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
—	—	—	COMMENT 71.66 (47)	—	—	—

### 7.2.2. 過少使用語との共起語

過剰使用語の分析と同様に、前節で抽出した過少使用語について、調整頻度 20 回以上共起した語を前後 3 語ずつ抽出し、これらの MWUs について個別に分析を行っていく。

#### 7.2.2.1. *enfin*

まず、*enfin* の生起頻度について、母語話者、上級学習者、初中級学習者の間で比較を行ったところ、以下の通りとなった。

表 201 *enfin* のグループ別頻度

	母語話者	上級学習者	初中級学習者
<i>enfin</i>	521.27 (336)	48.93 (12)	0 (0)

初中級学習者の生起頻度が 0 であったため、母語話者と上級学習者の間で有意差検定を行ったところ、有意差が認められた ( $\chi^2=99.63$ ;  $p<.001$ )。



Hansen (2005 :64)は *enfin* について 3つの用法があると指摘している。時間的意味を持つもの(「最後に」、「結局」)、統合する意味をもつもの(「要するに」)のような意味を持つ)、前言の補足や訂正を行うもの(「というよりは」という意味を持つ)である。3つ目の機能を持つ *enfin* の使用が話し言葉では最もよく見られるものである (Beeching 2007 : 81)。実際に母語話者の L1 の単語を見てみるとためらい語が先攻しており、言い直しの機能を持つ *enfin* が多いことが分かる。そこでこのようなディスコースマーカの使用について頻度の比較を行う。母語話者によって生起頻度が 20 以上であった後続語 *je, c', tu, il* について、*enfin* の副詞的用法とディスコースマーカ的な用法の生起頻度の比較を行うこととする。

表 202 *enfin* の副詞的用法と DM 的用法の比較

	母語話者	上級学習者
<i>enfin je</i>		
副詞	0	0
ディスコースマーカ	39	1
<i>enfin c'</i>		
副詞	0	0
ディスコースマーカ	26	1
<i>enfin tu</i>		
副詞	0	0
ディスコースマーカ	17	0
<i>enfin il</i>		
副詞	0	0
ディスコースマーカ	19	0

#### ディスコースマーカ的使用

non mais il y avait eu un problème je pense **enfin** je HESIT je crois (FL01)  
 en fait je viens de je viens d'NP **enfin** je je suis originaire de NP (FL05)  
 mais les phénomènes surnaturels c'est **enfin** c'est bizarre (FL17)  
 ouais ouais mais bon **enfin** tu expliques pourquoi et tu donnes une autre solution (FL10)  
 et alors il faisait nuit je crois **enfin** il pleuvait (FL18)  
 c'est pas très animé parce que c'est **enfin** c'est au début des au début novembre (JA12)  
**enfin** je sais pas mais n'importe quoi mais HESIT ils sont muets les étudiants (JA12)

このように、母語話者による全ての使用はディスコースマーカ的な使用であった。一方、上級学習者にもディスコースマーカ的な使用が見られたが、少数であった。上級学習者コーパスでは、上記の 2 例を含めてディスコースマーカ的な使用が 8 例のみで、副詞的用法は 4 例のみ見られた。そのうちの 1 例は *enfin* の繰り返しであったため、実質 3 例が副詞的使用であった。なお、以下の例の JA01 と JA12 はペアであり、実際にこの順番でインタラクションが行われていた。

donc j'allais au supermarché seulement quatre cinq fois HESIT par une année **enfin** (JA08)  
 et **enfin** (JA01)  
**enfin enfin** on a trouvé cette fille entre guillemets malade (JA12)

表 203 学習者による *enfin* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
						ENFIN	34.76 (12)						

表 204 母語話者による *enfin* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
				HESIT	33.54 (22)	ENFIN	521.27 (336)	JE	54.89 (36)	SAIS	35.07 (23)	PAS	39.64 (26)
								C'	36.59 (24)	EST	35.07 (23)	A	22.87 (15)
								TU	25.92 (17)				
								IL	24.39 (16)				

### 7.2.2.2. *hein*

*hein* は話し言葉で顕著に現れ、相手の言葉を聞き返す「何？」という意味だけでなく、同意を促す機能(「そうですね？」という意味)のようなディスコースマーカ―としての機能をもつ。学習者にはこのような使用が 1 例しか見られず、これは上級学習者による使用であった。

一方、母語話者は *hein* を多く使用し、聞き手とのインタラクションをより活発にしていると考えられる。特に、一番多く見られた *hein* を含む MWU は、*c'est pas*+形容詞/副詞 *hein* というもので、18 例見られた。これは、「(形容詞/副詞の否定)でしょう？」という意味になる。

ディスコースマーカ―的使用

*c'est vrai hein c'est c'est très c'est très animé* (JA12)

*non HESIT enfin c'est pas c'est pas clair hein HESIT* (FL14)

*ce n'est pas mal hein* (FL17)

表 205 学習者による *hein* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	-	-	HEIN	2.9 (1)	-	-	-	-	-	-

表 206 母語話者による *hein* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
EST	28.97 (19)	PAS	27.44 (18)	-	-	HEIN	368.95 (242)	-	-	-	-	-	-

### 7.2.2.3. *qui*

まず、この語の頻度について、母語話者、上級学習者、初中級学習者を比較した。その結果、以下の通りとなった。

表 207 *qui* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
973.63 (615)	309.86 (76)	190.17 (19)	$\chi^2=142.37$ ; $p<.001$

有意差が見られたため、グループ間の比較を行った。多重比較を行ったところ、初中級学習者と上級学習者の間に有意差は見られなかったが、学習者グループと母語話者の間に有意差が見られた。

<i>qui</i>	初中級学習者 = 上級学習者 < 母語話者

MWUs レベルでの比較を行うため、中心語 *qui* の左右の語を比較すると、学習者コーパスには *qui* の連続が見られることが分かる。そこで *qui* の生起回数における *qui* の連続使用頻度について比較を行った。

表 208 *qui* の繰り返しの頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>qui qui</i>	13	6	2	$\chi^2=24.29$ ; $p<.001$
<i>qui qui qui</i>	2	3	0	NA
<i>qui qui qui qui</i>	0	1	0	NA

*qui qui* の 2 連続使用では、3 グループ間に有意差が見られたため、多重比較を行ったところ、初中級学習者と上級学習者との間には有意差は見られなかったが、学習者グループと母語話者の間には有意差が見られた。

<i>qui qui</i>	初中級学習者 = 上級学習者 > 母語話者

3 連続使用については、初中級学習者コーパスで 1 例も見られなかったため、母語話者と上級学習者の比較を行ったところ、母語話者よりも上級学習者が 3 連続使用をしていることが分かった ( $\chi^2=37.61$ ;  $p<.001$ )。さらに、*qui* の 4 連続使用については、上級学習者コーパスで 1 例見られたが、他のコーパスには一例も見られなかった。つまり以上を総合すると、学習者は *qui* を過少使用しているが、*qui* の連続使用をしているということが分かった。

次に *qui* の機能について詳しく見ていく。*qui* は主に関係代名詞、疑問詞、強調構文の一部という 3 つの機能を持つ。そこで、両コーパスの R1 位置に見られる調整頻度 20 以上の動詞である *est* と *a* について、どの機能で使用されているかについて比較した。

表 209 *qui est, qui a* の機能

	母語話者	上級学習者	初中級学習者
<i>est</i>			
関係代名詞	83	9	2
強調構文	4	1	0
疑問詞	3	0	2
<i>a</i>			
関係代名詞	24	4	0
強調構文	10	4	1
疑問詞	2	0	0

学習者による *qui* の使用頻度がもともと低いため、*est* や *a* との共起頻度は低い。しかし関係代名詞にしる、強調構文にしる、複文を構成する *qui* の使用は母語話者と比較してかなり頻度が低いといえる。このように、複文が少ないことが *qui* の使用頻度の低さの一因となっていることが示唆できる。

表 210 学習者による *qui* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
EST	31.87 (11)	QUI	43.46 (15)	QUI	49.25 (17)	QUI	275.22 (95)	EST	40.56 (14)	-	-	-	-
A	20.28 (7)	QUELQU	20.28 (7)	UN	20.28 (7)			A	26.07 (9)				
QUI	20.28 (7)			EST	20.28 (7)								

表 211 母語話者による *qui* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
EST	68.61 (45)	DES	77.75 (51)	CE	93 (61)	QUI	973.63 (615)	EST	137.21 (90)	UN	33.54 (22)	EN	21.34 (14)
A	47.26 (31)	UN	54.89 (36)	NP	33.54 (22)			SONT	85.38 (56)	EST	27.44 (18)		
C'	36.59 (24)	EST	41.16 (27)	GENS	32.02 (21)			A	53.36 (35)	A	25.92 (17)		
Y	27.44 (18)	DE	35.07 (23)	ET	28.97 (19)			ÉTAIT	36.59 (24)	PAS	25.92 (17)		
DE	24.39 (16)	LES	32.02 (21)	HESIT	28.97 (19)			S'	32.02 (21)	HESIT	24.39 (16)		
		EN	30.49 (20)	QUI	25.92 (17)			M'	32.02 (21)	EN	24.39 (16)		
		UNE	25.92 (17)	A	21.34 (14)			SE	32.02 (21)				
		QUI	24.39 (16)					VA	24.39 (16)				
		LA	22.87 (15)					ONT	22.87 (15)				
		TOUT	22.87 (15)										

#### 7.2.2.4. *quoi*

まず *quoi* の頻度について母語話者、上級学習者、初中級学習者の間に有意差があるか調べたところ、有意差が見られることが分かった。

表 212 *quoi* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
588.50 (386)	114.16 (28)	120.11 (12)	$\chi^2=119.23; p<.001$

そこで、どのグループ間に差が見られるかどうかについて多重比較を行ったところ、学習者間の頻度については有意差が見られなかったが、学習者と母語話者の間には有意差が見られた。

<i>quoi</i>	初中級学習者 = 上級学習者 < 母語話者
-------------	-----------------------

次に、MWUs レベルの比較を行い、過少使用の原因について分析を行う。*quoi* と共に使用される調整頻度 20 以上の語について、学習者の表を見てみると、*c'est* が左側に多く使用されていることが分かる。そこで、*c'est quoi* について比較を行う。また、イントネーション型である *c'est quoi* は、EST-CE QUE 型になると *qu'est-ce que c'est* という形になり、EST-CE QUE 型を使用した場合は *quoi* の生起頻度が下がる。そのため、この形についても頻度を調べて *c'est quoi* と *qu'est-ce que c'est* の使用頻度を比較する。

表 213 疑問文の型による頻度分布

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>c'est quoi</i>	13	5	7	$\chi^2=0.06; p>.05$
<i>qu'est-ce que c'est</i>	2	1	1	

この結果、*c'est quoi* と *qu'est-ce que c'est* の使用については有意差が見られなかったため、この使用によって過少使用が引き起こされたとはいえない。そこで、過少使用の原因と考えられるのはディスコースマーカースとしての *quoi* である。ディスコースマーカースの *quoi* は間投詞の一種であり、談話終了を示す機能を持つ。このような機能を持つ *quoi* は話し言葉でよく使用されている。そこで *quoi* の総頻度に対する、母語話者、上級学習者、初中級学習者によるこれらの使用頻度を比較した。

表 214 DM としての *quoi* のグループ別頻度

	母語話者	上級学習者	初中級学習者
<i>quoi</i> (DM)	307	12	0

初中級学習者による使用頻度が一例も見られなかったため、母語話者と上級学習者の使用頻度について比較を行ったところ、母語話者と上級学習者の間には有意差が見られた ( $\chi^2=19.86; p<.001$ )。母語話者の使用の多くがこのディスコースマーカースであるということから、学習者による *quoi* の過少使用はこのディスコースマーカースの過少使用によるものであるといえる。

表 215 学習者による *quoi* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	C'	40.56 (14)	EST	34.76 (12)	QUOI	115.88 (40)	-	-	-	-	-	-

表 216 母語話者による *quoi* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
C'	25.92 (17)	LA	30.49 (20)	EST	21.34 (14)	QUOI	588.50 (386)	-	-	-	-	-	-
		C'	28.97 (19)										
		EST	25.92 (17)										
		DE	21.34 (14)										

### 7.2.2.5. *les*

名詞の複数形の前におかれる定冠詞である *les* の過少使用について調べるため、まず母語話者、上級学習者、初中級学習者の頻度を調べた。

表 217 *les* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
1359.94 (892)	640.11 (157)	440.40 (44)	$\chi^2=129.27$ ; $p<.001$

有意差が見られたため、多重比較を行ったところ、初級学習者と上級学習者の間には有意差が見られなかったが、学習者グループと母語話者との間には有意差が見られた。

<i>les</i>	初中級学習者 = 上級学習者 < 母語話者
------------	-----------------------

次に MWUs レベルでの比較を行うため、コロケートについて概観する。定冠詞の後ろには様々な名詞がおかれる可能性があるため、左側の語についてのみ分析を行うこととする両方のリストとも、L1 で最も使用されていたのは、*tous* であった。そこで、*tous les* という MWU について比較を行う。

表 218 *tous les* のグループ別頻度

	母語話者	上級学習者	初中級学習者
<i>tous les</i>	80.80 (53)	53.00 (13)	0 (0)

初中級学習者によるこの使用は 1 例も見られなかったため、母語話者と上級学習者について *les* の生起頻度に対する *tous les* の生起頻度に有意差があるかどうか検定を行ったところ、使用頻度に有意差が見られなかった ( $\chi^2=1.24$ ;  $p>.05$ )。しかし、母語話者コーパスの L1 には女性形名詞の前におかれる *toute les* という MWU が見られることがわかる。これについては、学習者コーパスの中には一例も見られなかった。

また、母語話者コーパスの L1 位置に見られる語で特筆すべきものは、*je* や *tu* といった人称代名詞主語である。人称代名詞主語のすぐ後ろに *les* がおかれた場合、*les* は代名詞である。そこで、人称代名詞主語の後ろで使用される代名詞 *les* の頻度について比較を行う。

表 219 人称代名詞の後ろで使用される代名詞 *les* の頻度

	je	tu	il	elle	on	nous	vous	ils	elles	合計
母語話者	18	19	3	0	13	0	1	2	0	66
上級学習者	1	0	0	0	0	0	0	0	0	1
初中級学習者	1	0	0	0	0	0	0	1	0	2

*les* の頻度における人称代名詞主語と使用される代名詞の合計数について有意差検定を行ったところ、有意差が見られた( $\chi^2=10.56$ ;  $p<.01$ )。そこで多重比較を行った結果、母語話者と上級学習者の使用に有意差が見られた( $\chi^2=10.21$ ;  $p<.01$ )。母語話者の代名詞の使用頻度と初中級学習者の使用頻度の間に有意差が見られなかったが、初中級学習者の生起頻度は1回のみということは留意しなければならない。これより、代名詞としての *les* の使用の過少使用が、*les* 全体の過少使用を引き起こしているといえる。

表 220 学習者による *les* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
LE	23.18 (8)	QUE	28.97 (10)	TOUS	37.66 (13)	LES	582.3 (201)	VACANCES	31.87 (11)	DE	72.43 (25)	NP	20.28 (7)
EST	20.28 (7)	HESIT	20.28 (7)	LES	34.76 (12)			ÉTUDIANTS	28.97 (10)	ET	20.28 (7)		
ET	20.28 (7)			PENDANT	34.76 (12)			FRANÇAIS	26.07 (9)				
HESIT	20.28 (7)			QUE	34.76 (12)			GALETTES	20.28 (7)				
J'	20.28 (7)			ET	28.97 (10)								
				SUR	23.18 (8)								
				POUR	23.18 (8)								

表 221 母語話者による *les* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3
LES	54.89 (36)	LES	39.64 (26)	TOUS	80.80 (53)	LES	1359.94 (892)	ÉTUDIANTS	86.9 (57)	DE	82.33 (54)	LES
C'	32.02 (21)	ON	32.02 (21)	POUR	79.28 (52)			GENS	42.69 (28)	ET	50.31 (33)	EST
TU	30.49 (20)	EST	30.49 (20)	DANS	64.03 (42)			DEUX	41.16 (27)	HESIT	44.21 (29)	L'
DE	28.97 (19)	ET	27.44 (18)	QUE	60.98 (40)			NP	24.39 (16)	C'	33.54 (22)	DE
ET	28.97 (19)	HESIT	27.44 (18)	LES	50.31 (33)					QUI	32.02 (21)	
À	25.92 (17)	PAS	24.39 (16)	JE	42.69 (28)							
IL	25.92 (17)	DE	22.87 (15)	AVEC	42.69 (28)							
ON	25.92 (17)	POUR	22.87 (15)	ET	35.07 (23)							
EST	22.87 (15)	C'	21.34 (14)	TOUTES	33.54 (22)							
JE	21.34 (14)			HESIT	33.54 (22)							
QUE	21.34 (14)			PAS	30.49 (20)							
				SUR	30.49 (20)							
				TU	28.97 (19)							

#### 7.2.2.6. *ça*

まず、指示代名詞である *ça* の生起頻度について、母語話者、上級学習者、初中級学習者を比較する。

表 222 *ça* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
1156.62 (1021)	954.05 (234)	350.32 (35)	$\chi^2=127.77$ ; $p<.001$

有意差が見られたため、多重比較を行った。その結果、母語話者と学習者グループの間に有意差があり、上級学習者と初中級学習者との間にも有意差が見られた。

<i>ça</i>	初中級学習者 < 上級学習者 < 母語話者
-----------	-----------------------

次に、MWUs レベルで比較を行う。学習者と母語話者のリストでの違いが顕著なのは、R1 位置であるといえる。なぜなら、母語話者コーパスの中には *ça* に後続する語の頻度上位には動詞が多いが、学習者コーパスにはそれほど動詞が見られない。そこで、各コーパス内の R1 位置に現れた生起頻度 20 以上の動詞をもとに、これらの動詞の時制が異なってもレマを単位として、右側に *ça* を主語にとるこれらの動詞が使用される頻度をカウントする。なお、*a* については動詞として使用される場合と、助動詞として使用される場合があるため、分けて比較を行うこととする。



表 223 *ça* に後続する動詞の頻度

	FAIRE	ALLER	POUVOIR	DEVOIR	AVIOR 助動詞	AVIOR 動詞	S'APPELER	合計
母語話者	66	64	23	22	20	10	8	213
上級学習者	10	12	1	1	4	0	7	35
初中級学習者	1	2	0	0	0	0	0	3

*ça* の頻度に対するこれらの動詞の頻度について母語話者、上級学習者、初中級学習者の間で比較を行った結果、有意差が見られた( $\chi^2=6.96$ ;  $p<.05$ )。そこで、多重比較を行ったところ、以下の通りとなった。

<i>ça</i>	上級学習者 < 母語話者
-----------	--------------

これは、一部の動詞についての結果であるが、上級学習者はこの指示代名詞を主語として使用する頻度が母語話者よりも少ないことが示唆できる。また、初級学習者と母語話者の間に有意差は見られなかったが、初級学習者の *ça* の頻度は低い上に、主語として使用される頻度も 3 回のみだったことを考慮すると、やはり学習者による *ça* の過少使用は、指示代名詞主語としての頻度が低いということが一因となっていると考えられる。

表 224 学習者による *ça* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
OUI	141.95 (49)	C'	182.51 (63)	EST	185.41 (64)	ÇA	779.3 (269)	C'	66.63 (23)	EST	78.22 (27)	LA	26.07 (9)
C'	57.94 (20)	EST	57.94 (20)	COMME	130.37 (45)			ET	46.35 (16)	APPELLE	23.18 (8)		
EST	28.97 (10)	ÇA	23.18 (8)	ÇA	34.76 (12)			VA	40.56 (14)	ÇA	23.18 (8)		
		ET	20.28 (7)	POUR	26.07 (9)			S'	34.76 (12)				
		INT	20.28 (7)	HESIT	26.07 (9)			FAIT	31.87 (11)				
				COMMENT	23.18 (8)			OUI	31.87 (11)				
				MAIS	23.18 (8)			MAIS	26.07 (9)				
				OUI	20.28 (7)			M'	26.07 (9)				
								QUE	23.18 (8)				

表 225 学習者による *ça* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
C'	71.66 (47)	C'	80.8 (53)	COMME	143.31 (94)	ÇA	1156.62 (1021)	C'	89.95 (59)	EST	97.57 (64)	PAS	56.41 (37)
JE	48.79 (32)	EST	70.13 (46)	QUE	118.92 (78)			VA	88.43 (58)	ÊTRE	51.84 (34)	QUE	30.49 (20)
EST	35.07 (23)	ET	35.07 (23)	EST	82.33 (54)			FAIT	83.85 (55)	A	47.26 (31)	EST	27.44 (18)
ET	35.07 (23)	PAS	33.54 (22)	TOUT	77.75 (51)			M'	70.13 (46)	PAS	45.74 (30)	DE	27.44 (18)
J'	33.54 (22)	HESIT	25.92 (17)	HESIT	56.41 (37)			ME	48.79 (32)	UN	27.44 (18)	ÇA	25.92 (17)
ÇA	25.92 (17)	ÇA	24.39 (16)	ET	51.84 (34)			JE	35.07 (23)	C'	27.44 (18)	UN	24.39 (16)
HESIT	24.39 (16)	AI	22.87 (15)	MAIS	47.26 (31)			S'	32.02 (21)	HESIT	22.87 (15)	ÊTRE	21.34 (14)
UN	24.39 (16)	DE	21.34 (14)	ÇA	38.11 (25)			A	32.02 (21)	FAIT	21.34 (14)	HESIT	21.34 (14)
EN	22.87 (15)			DONC	36.59 (24)			HESIT	30.49 (20)			LE	21.34 (14)
TU	21.34 (14)			POUR	36.59 (24)			DOIT	28.97 (19)			BIEN	21.34 (14)
				OUAIS	32.02 (21)			ET	27.44 (18)				
				PAS	25.92 (17)			SE	25.92 (17)				
				DE	25.92 (17)			PEUT	21.34 (14)				
								QUE	21.34 (14)				

7.2.2.7. *des*

まず、*des* の使用頻度の差について調べるため、母語話者、上級学習者、初中級学習者の頻度についてまとめた。

表 226 *des* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
992.51 (651)	448.49 (110)	330.30 (33)	$\chi^2=97.37$ ; $p<.001$

使用頻度に有意差が見られたため、多重比較を行いどのグループ間に有意差が現れるのかを調べたところ、以下の通りとなった。

<i>des</i>	初中級学習者 = 上級学習者 < 母語話者
------------	-----------------------

学習者間には有意差が見られなかったが、学習者と母語話者との間に有意差が見られた。つまり、レベルを問わず学習者は *des* を過少使用している。

次に、MWUs レベルでの比較を行う。両コーパスとも、*des* の連続使用が見られる。そこで、連続使用の生起頻度について、母語話者、上級学習者、初中級学習者で比較したところ、以下のような結果となった。

表 227 *des* の繰り返しの頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>des des</i>	24	16	1	$\chi^2=48.60$ ; $p<.001$
<i>des des des</i>	4	1	0	NA

2 連続使用の頻度について有意差が見られたため、多重比較を行ったところ、上級学習者が過剰使用していることが分かった。3 連続使用では初中級学習者の出現頻度が一例もなかったため、母語話者と上級学習者の頻度を比較したところ、有意差は見られなかった ( $\chi^2=0.38$ ;  $p>.05$ )。

<i>des des</i>	上級学習者 > 初中級学習者 = 母語話者
----------------	-----------------------

また、コロケートを見てみると、両コーパスとも *il y a des* という MWU があることがわかる。*des* の生起頻度におけるこの MWU の生起頻度について 3 グループ間の比較を行ったところ、有意差は見られなかった。つまり、この MWU は *des* の過少使用を説明するものではない。

表 228 *il y a des* のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>il y a des</i>	120.44 (79)	53.00 (13)	40.04 (4)	$\chi^2=0.01$ ; $p>.05$

以上を総合すると、上級学習者も初中級学習者も *des* の使用頻度が低いが、*des* の繰り返しの使用は上級学習者に多く見られた。しかし、共通して見られた MWUs については有意差が見られなかった。つまり、学習者による過少使用は MWUs レベルというよりは単語レベルでの過少使用である可能性が高いといえる。たとえば、複数名詞自体が過少使用されていて、それに先行する不定冠詞 *des* も過少使用されたという可能性がある。

表 229 学習者による *des* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
IL 63.73 (22)	Y 55.04 (19)	DES 52.15 (18)	DES 414.28 (143)	AMIS 34.76 (12)	DE 28.97 (10)	—
Y 23.18 (8)	A 28.97 (10)	A 52.15 (18)				
	DES 20.28 (7)	AVEC 26.07 (9)				
		HESIT 20.28 (7)				

表 230 母語話者による *des* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
IL	128.07 (84)	Y	126.54 (83)	A	118.92 (78)	DES	992.51 (651)	GENS	48.79 (32)	QUI	77.75 (51)	SONT	25.92 (17)
QUE	28.97 (19)	C'	50.31 (33)	DES	48.79 (32)			ÉTUDIANTS	41.16 (27)	DE	73.18 (48)	HESIT	22.87 (15)
DES	27.44 (18)	LA	36.59 (24)	EST	47.26 (31)			NP	30.49 (20)	HESIT	36.59 (24)	DE	21.34 (14)
ET	24.39 (16)	DES	33.54 (22)	AVEC	30.49 (20)			FOIS	30.49 (20)	À	24.39 (16)		
DE	22.87 (15)	TU	27.44 (18)	QUE	24.39 (16)			CHOSES	27.44 (18)	ET	24.39 (16)		
Y	22.87 (15)	A	25.92 (17)					COURS	22.87 (15)				
EST	21.34 (14)	DE	24.39 (16)					TRUCS	21.34 (14)				

### 7.2.2.8. *ils*

まず、母語話者、上級学習者、初中級学習者の間で生起頻度の比較を行ったところ、以下の通りとなった。

表 231 *ils* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
728.76 (478)	350.63 (86)	70.06 (7)	$\chi^2=93.99; p<.001$

有意差が見られたため多重比較を行ったところ、母語話者による使用頻度が最も高く、上級学習者、初級学習者と続いた。

<i>ils</i>	初中級学習者 < 上級学習者 < 母語話者
------------	-----------------------

次に、MWUs レベルの比較を行う。学習者の使用で特徴的なのは、*ils* の連続使用である可能性が高い。そこで *ils* の連続使用について比較を行った。

表 232 *ils* の繰り返しの頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>ils ils</i>	5	6	0	NA
<i>ils ils ils</i>	0	2	0	NA

2 連続使用と 3 連続使用とも初中級学習者の生起頻度が 1 例も見られず、3 連続使用では上級学習者にしか見られなかったため、2 語連鎖について上級学習者と母語話者の生起頻度を比較したところ、上級学習者が有意に多く 2 語連鎖を使用していることが分かった ( $\chi^2=27.36; p<.001$ )。

初級学習者については、この語の使用頻度はかなり低く、上級学習者については連続使用が多く用いられているということが分かった。つまり、この語についても MWUs レベルで過剰使用されているのではなく、単語レベルで過少使用されているということが考えられる。実際に、*ils* に後続する調整頻度 20 回以上の単語を見てみると、どちらのコーパ

スとも動詞が後続していることがわかるが、学習者の動詞の種類は少ない上に生起頻度もかなり低いことが分かる。つまり、学習者は第三人称複数主語をあまり用いていないということが考えられる。

表 233 学習者による *ils* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	ILS	34.76 (12)	HESIT	23.18 (8)	ILS	269.42 (93)	SONT	89.81 (31)	TRÈS	20.28 (7)	-	-
				ILS	23.18 (8)			ÉTAIENT	23.18 (8)				
								ONT	20.28 (7)				

表 234 母語話者による *ils* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	CE	21.34 (14)	QU	112.82 (74)	ILS	728.76 (478)	ONT	143.31 (94)	PAS	47.26 (31)	DE	35.07 (23)
				ET	38.11 (25)			SONT	88.43 (58)			PAS	30.49 (20)
				HESIT	38.11 (25)			FONT	30.49 (20)			À	24.39 (16)
								VONT	28.97 (19)				
								SE	27.44 (18)				
								ME	25.92 (17)				
								AVAIENT	22.87 (15)				

#### 7.2.2.9. *fait*

*fait* の生起頻度について、母語話者、上級学習者、初中級学習者を比較したところ、以下のような結果となった。

表 235 *fait* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
815.66 (535)	350.63 (86)	260.23 (26)	$\chi^2=85.88$ ; $p<.001$

有意差が見られたため多重比較を行ったところ、初中級学習者と上級学習者の間には有意差が見られなかったが、学習者グループと母語話者との間には有意差が見られた。

<i>fait</i>	初中級学習者 = 上級学習者 < 母語話者
-------------	-----------------------

次に、MWUs レベルで比較を行う。両コーパスとも L1 の位置に *en* があることが分かる。*en fait* という MWU はもともと「実際は」という意味を持つ表現であるが、話し言葉では前言の補足や訂正を行う機能を持つディスコースマーカースとして使用されることが多い。そこで、*fait* の頻度におけるこの MWU の生起頻度の比較を行う。

表 236 *en fait* のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>en fait</i>	217	48	2	$\chi^2=19.71$ ; $p<.001$

その結果、母語話者に比べて上級学習者はこの MWU を多用しており、初級学習者は母語話者よりも過少使用していることが分かった。

<i>en fait</i>	上級学習者 > 母語話者 > 初級学習者
----------------	----------------------

次に、母語話者のリストを見てみると、*tout à fait* という MWU が含まれることが分かる。そこで、*fait* の頻度におけるこの MWU の使用頻度を比較したところ、以下の通りとなった。

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
36.59 (24)	16.31 (4)	0 (0)	NA

初中級学習者にはこの MWU は 1 例も見られなかった。そこで、上級学習者と母語話者の使用頻度を比較したところ、有意差は見られなかった ( $\chi^2=0.00$ ;  $p>.05$ )。

次に、動詞の現在形、および過去分詞形として使用された頻度を比較する。まず現在形として生起された *fait* の生起頻度を数えたところ、以下のような結果となった。

表 237 *fait*(現在形)のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
動詞の現在形	138	19	13	$\chi^2=8.40$ ; $p<.05$

有意差が見られたため多重比較を行ったところ、初中級学習者が有意に多く動詞としての *fait* を使用していることが分かった。そして、学習者と母語話者の間には有意差は見られなかった。

現在形	初中級学習者 > 上級学習者 = 母語話者
-----	-----------------------

次に、過去分詞として *fait* が使用された頻度を比較したところ、以下の通りとなった。なお、*fait* は動詞 AVOIR とともに複合過去の一部となる場合と、動詞 ÊTRE とともに受身の過去分詞を形成する場合があるため、両方のケースを別々にカウントする。

表 238 *fait*(過去分詞)のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
過去分詞 (過去)	91	14	11	$\chi^2=10.97$ ; $p<.001$
過去分詞 (受動態)	16	1	0	NA

過去形の過去分詞 *fait* の使用頻度に有意差が見られた。そのため、多重比較を行ったところ、これについても初中級学習者は上級学習者と母語話者よりも過去形の使用頻度が有意に高いことが分かった。

過去分詞(過去)	初中級学習者>上級学習者=母語話者
----------	-------------------

初中級学習者による受動態の *fait* の使用が見られなかったため、上級学習者と母語話者間で有意差を検定したところ、有意差は見られなかった( $\chi^2=0.93$ ;  $p>.05$ )。

以上の結果についてまとめると、母語話者と比べて上級学習者も初中級学習者も *fait* を過少使用しているが、上級学習者は *fait* を使用する際には *en fait* という使用を多用しており、初中級学習者は動詞として *fait* を使用している割合が高いといえる。

表 239 学習者による *fait* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
QUI	20.28 (7)	J'	31.87 (11)	EN	144.85 (50)	FAIT	324.47 (112)	C'	23.18 (8)	EST	23.18 (8)	-	-
QUE	20.28 (7)	TU	23.18 (8)	AI	31.87 (11)								
		MAIS	20.28 (7)	ÇA	31.87 (11)								
		OUI	20.28 (7)	IL	31.87 (11)								
				AS	23.18 (8)								

表 240 母語話者による *fait* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
HESIT	32.02 (21)	J'	50.31 (33)	EN	330.84 (217)	FAIT	815.66 (535)	HESIT	41.16 (27)	EST	33.54 (22)	-	-
ET	32.02 (21)	ET	41.16 (27)	ÇA	83.85 (55)			C'	32.02 (21)				
QUE	27.44 (18)	TOUT	38.11 (25)	AI	50.31 (33)			UN	30.49 (20)				
		ON	35.07 (23)	A	48.79 (32)			TU	27.44 (18)				
		HESIT	33.54 (22)	À	36.59 (24)			DE	24.39 (16)				
		ÇA	21.34 (14)	PAS	25.92 (17)			J'	22.87 (15)				
				ON	21.34 (14)			QUE	22.87 (15)				
				IL	21.34 (14)			UNE	22.87 (15)				

#### 7.2.2.1 0. *ouais*

7.2.1.2.の項で既に指摘した通り、学習者による *ouais* の過少使用は *oui* の過剰使用の影響を受けていた。まず、グループの間で *ouais* の使用頻度に有意差があるかどうか調べたところ、以下のような結果となった。

表 241 *ouais* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
1405.68 (922)	937.74 (230)	380.94 (38)	$\chi^2=95.02$ ; $p<.001$

有意差が見られたため多重比較を行ったところ、初中級学習者による *ouais* の使用が最も少なく、上級学習者も母語話者より有意に少なく使用していた。

<i>ouais</i>	初中級学習者 < 上級学習者 < 母語話者
--------------	-----------------------

次に、MWUs レベルでの比較を行う。両方のコーパスに *ouais* の連続使用が見られるため、その頻度を比較した。

表 242 *ouais* の繰り返しの頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>ouais ouais</i>	61	12	2	$\chi^2=1.46$ ; $p>.05$
<i>ouais ouais ouais</i>	4	2	1	$\chi^2=10.32$ ; $p<.01$

その結果、*ouais* の 2 連続使用には有意差が見られなかった。3 連続使用については以下ようになった。

<i>ouais ouais ouais</i>	初中級学習者 > 上級学習者 = 母語話者
--------------------------	-----------------------

初中級学習者が他のグループよりも過剰使用が見られたが、生起頻度は 1 度だけであることを加味すると、*ouais* の過剰使用は、単語レベルでの過少使用であるといえる。

表 243 学習者による *ouais* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
C'	23.18 (8)	Ouais	26.07 (9)	INT	81.12 (28)	Ouais	773.51 (267)	C'	60.84 (21)	EST	49.25 (17)	-	-
		EST	20.28 (7)	Ouais	75.32 (26)			ET	31.87 (11)				
				HESIT	23.18 (8)			MAIS	26.07 (9)				
				NP	20.28 (7)			HESIT	23.18 (8)				
								DONC	20.28 (7)				



表 244 母語話者による *ouais* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
				INT	117.39 (77)	OUAIS	1405.68 (922)	C'	99.1 (65)	EST	86.9 (57)	EST	44.21 (29)
				OUAIS	106.72 (70)			MAIS	73.18 (48)	C'	53.36 (35)	C'	24.39 (16)
				HESIT	54.89 (36)			VOILÀ	35.07 (23)				
				BC	35.07 (23)			ÇA	32.02 (21)				
								JE	30.49 (20)				
								HESIT	27.44 (18)				
								IL	25.92 (17)				

### 7.2.2.1 1. *qu'*

*qu'*は *que* の後ろに母音が続いた際にエリジオンがおこった形である。母語話者、上級学習者、初中級学習者の間に *qu'*の使用頻度の差があるかどうかについて、まず比較を行う。

表 245 *qu'*のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
727.23 (477)	199.78 (49)	540.49 (54)	$\chi^2=86.51$ ; $p<.001$

使用頻度に有意差が見られたため多重比較を行ったところ、上級学習者が初中級学習者と母語話者と比較して使用頻度が低いことが分かった。

<i>qu'</i>	上級学習者 < 初中級学習者 = 母語話者
------------	-----------------------

次に、コロケートの頻度に注目し、MWUs レベルでの分析を行う。母語話者の R1 位置の単語に注目すると、母音で始まる人称代名詞主語が続いていることが分かる。これらは母音で始まる人称代名詞主語が後続することによって *que* のエリジオンがおこり *qu'*となったケースであるといえる。そこで、これらの人称代名詞主語が後続した場合において、*que* がエリジオンした場合とエリジオンしない場合についての頻度を比較する。

表 246 *que* のエリジョンした場合としない場合の頻度比較

	母語話者	上級学習者	初中級学習者
<i>qu'il</i>	147	14	14
<i>qu'elle</i>	24	1	2
<i>qu'on</i>	131	6	0
<i>qu'ils</i>	74	3	2
<i>qu'elles</i>	4	0	0
合計	380	24	18
<i>que il</i>	32	22	1
<i>que elle</i>	3	5	0
<i>que on</i>	11	9	0
<i>que ils</i>	8	2	0
<i>que elles</i>	0	0	0
合計	54	38	1

エリジョンの有無についての頻度を比較したところ、有意差が見られた( $\chi^2=89.66$ ;  $p<.001$ )。有意差が見られたため多重比較を行ったところ、上級学習者は他のグループよりもエリジョンを行っていないことが分かった。

エリジョン無し	上級学習者 > 初中級学習者 = 母語話者
---------	-----------------------

以上を総合すると、上級学習者が *qu'* を過少使用したのはエリジョンを行っていないからであると考えられる。

表 247 学習者による *qu'* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	EST	26.07 (9)	PARCE	40.56 (14)	QU'	298.40 (103)	EST	121.68 (42)	CE	118.78 (41)	QUE	110.09 (38)
				CE	37.66 (13)			IL	81.12 (28)	Y	55.04 (19)	A	49.25 (17)
				ET	31.87 (11)					NP	34.76 (12)		
										A	20.28 (7)		

表 248 母語話者による *qu'* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
C'	51.84 (34)	EST	70.13 (46)	CE	102.15 (67)	QU'	727.23 (477)	IL	224.12 (147)	Y	100.62 (66)	A	85.38 (56)
EST	24.39 (16)	JE	51.84 (34)	PARCE	88.43 (58)			ON	199.72 (131)	A	51.84 (34)	QUE	42.69 (28)
HESIT	21.34 (14)			VRAI	33.54 (22)			ILS	112.82 (74)	CE	45.74 (30)	PAS	28.97 (19)
								EST	51.84 (34)	ONT	24.39 (16)	AVAIT	21.34 (14)
								ELLE	39.64 (26)	EST	21.34 (14)		
								EN	33.54 (22)				

### 7.2.2.1 2. *vois*

*vois* は、動詞 VOIR の直説法一人称もしくは二人称単数現在形の活用形である。まず、母語話者、上級学習者、初中級学習者の間で *vois* の使用頻度に差があるのかどうかを比較する。

表 249 *vois* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
231.74 (152)	24.46 (6)	30.03 (3)	$\chi^2=59.59$ ; $p<.001$

有意差が見られたため、多重比較を行ったところ以下のようになった。

<i>vois</i>	初中級学習者 = 上級学習者 < 母語話者
-------------	-----------------------

このように、学習者のレベルを問わず使用頻度が低かった。次に、MWU レベルでの比較を行うためコンコーダンスを見てみると、L1 で *tu* が用いられることが多いことが分かる。そこで、*tu vois* という MWU について比較を行う。*tu vois* は文末で使用されて、相手の意見を求めたり内容を確認したりする機能を持ち、ディスコースマーカースとして使用されることが多い。そこで *vois* の使用頻度に対して、ディスコースマーカースとしての *tu vois* がどれくらい使用されるかを比較した。

表 250 DM としての *tu vois* の頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
DM としての <i>tu vois</i>	176.85 (116)	20.39 (5)	20.02 (1)	$\chi^2=0.32$ ; $p>.05$

*vois* の頻度に対してこのディスコースマーカースの使用頻度の割合に有意差は見られなかった。しかしそれはもともと *vois* の使用頻度が低いことに起因している。つまり、学習者によるこのディスコースマーカースの使用が低いため、*vois* の過少使用が引き起こされたともいえる。

表 251 学習者による *vois* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度
-	-	TU	20.28 (7)	VOIS	26.07 (9)	-	-

表 252 母語話者による *vois* の共起語

L3 頻度	L2 頻度	L1 頻度	KW	頻度	R1 頻度	R2 頻度	R3 頻度
-	-	TU	195.15 (128)	VOIS	231.74 (152)	-	-

### 7.2.2.1 3. là

まず là について、母語話者、上級学習者、初中級学習者の使用頻度を比較したところ、以下のようになった。

表 253 là のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
577.82 (379)	224.24 (55)	170.15 (17)	$\chi^2=69.22$ ; $p<.001$

有意差が確認されたため、3 グループのどの間に有意差があらわれたかどうか多重比較を行って調べたところ、以下のような結果となった。

là	初中級学習者 = 上級学習者 < 母語話者
----	-----------------------

したがって、学習者はレベルを問わずこの語を過少使用している。

次に、R1 の単語を見てみると、*bas* が最も多く共起されていることが分かる。*là* の後ろに *bas* が後続した場合、*là-bas* という「向こうに」を意味する語となる。そこでこの語の使用頻度を比較したところ、以下のようになった。

表 254 là-bas のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>là-bas</i>	42.69 (28)	106.01 (26)	130.12 (13)	$\chi^2=113.43$ ; $p<.001$

有意差が確認されたため、多重比較を行ったところ、初中級学習者と上級学習者は母語話者よりも多く *là-bas* という語を用いていることが分かった。

<i>là-bas</i>	初中級学習者 = 上級学習者 > 母語話者
---------------	-----------------------

このことは、学習者は母語話者よりも *là* の使用頻度数そのものは低いものの、*là-bas* という語となると、学習者の使用頻度の方が高くなるということを意味している。

また、学習者コーパスにも母語話者コーパスにも指示形容詞 *ce* が L2 の位置に見られている。指示形容詞とともに *là* が用いられることによって、強意の意味が付加される。このような例について *là* の生起頻度をもとに比較したところ、以下の通りとなった。

	母語話者	上級学習者	初中級学習者
指示形容詞 + 名詞 + là	22.87 (15)	69.31 (17)	0 (0)

初中級学習者コーパスにはこの使用が一例も見られなかったため、母語話者と上級学習者の間で有意差検定を行ったところ、有意差が見られた ( $\chi^2=51.08$ ;  $p<.001$ )。したがって、この使用については上級学習者のほうが母語話者よりも多く使用しているといえる。

一方、学習者コーパスには L1 位置に調整頻度 20 以上のコロケートが見られないが、母語話者コーパスのコロケートを見てみると、接続詞が多いことが分かる。接続詞の後ろに *là* がおかれることによっても、「そこで」という意味に強意の意味が付加される。そこで、接続詞 *et, mais, donc* との共起頻度について比較を行う。

表 255 接続詞との共起頻度

	母語話者	上級学習者	初中級学習者
<i>et là</i>	33.54 (22)	0 (0)	0 (0)
<i>mais là</i>	33.54 (22)	0 (0)	0 (0)
<i>donc là</i>	28.97 (19)	0 (0)	0 (0)

表のとおり、母語話者にはこのような使用が多く見られたのに対して、上級学習者と初中級学習者には一例も見られなかった。このことより、学習者によって使用される *là* は *là-bas* の一部として使用される傾向が高いが、母語話者は強意の意味を持つ *là* を多く使用していることが分かった。また、強意の意味を持つ *là* の過少使用が、*là* の過少使用の一因であるといえる。

表 256 学習者による *là* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
—	—	—	—	—	—	—
	CE	49.25 (17)	LÀ	208.59 (72)	BAS	104.29 (36)
	C'			C'	20.28 (7)	

表 257 母語話者による *là* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
—	—	—	—	—	—	—
	CE	22.87 (15)	EST	38.11 (25)	LÀ	577.82 (379)
	C'	21.34 (14)	ET	33.54 (22)	BAS	42.69 (28)
			MAIS	33.54 (22)	C'	35.07 (23)
			DONC	28.97 (19)	JE	32.02 (21)
					EST	36.59 (24)

#### 7.2.2.1 4. *bon*

*bon* の使用頻度について、母語話者、上級学習者、初中級学習者の頻度を比較した。

表 258 *bon* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
506.17 (332)	150.85 (37)	280.25 (28)	$\chi^2=60.86; p<.001$

有意差が見られたので、多重比較を行いグループ間の比較を行った。その結果、母語話者による使用が最も多く、学習者グループと有意差が見られた。学習者間においても上級学習者と初中級学習者の間に有意差が見られた。

<i>bon</i>	初中級学習者<上級学習者<母語話者
------------	-------------------

学習者のコロケートを見てみると、調整頻度 20 以上のものは左側にしか現れず、動詞 ÊTRE が前に現れていることより、学習者は *bon* について「よい」という形容詞的使用を好む傾向があると考えられる。一方、母語話者のコロケートの多くは右側に多く見られ、さらに R1 の位置に主語が見られることから、この後ろに文が続くと考えられる。つまり、学習者と母語話者とで使用方法が違うと考えられる。

純粹な形容詞的使用以外の *bon* について、Hansen (1998)は間投詞的使用と談話的使用に分けている。後者の場合は文中や文末におかれる。特に、文末におかれた場合は、発話の断続的な段階を終了させる役割を持ち、標準的である形容詞的使用との関係は明白である。しかし文中におかれる *bon* は、発言の真理について全てを言わないというような発言の強さを軽減する役割を持つ (Beeching 2007 :81 より引用)。本研究では、この 2 つのディスコースマーカ的使用について、母語話者、上級学習者、初中級学習者の間で比較を行う。なお、この 2 つの使用については特に区別を行わない。

表 259 DMとしての *bon* のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
DMの <i>bon</i>	422.31 (277)	28.54 (7)	20.02 (2)	$\chi^2=131.78$ ; $p<.001$

実際のデータをみてみると、文頭にもこのディスコースマーカが使用されている例が多く見られた。これは、文末におかれるディスコースマーカと同じような機能を表すが、対話者の前言について一度終了させ、自分の発話を始めるときに使用されていた。このような例も含めたディスコースマーカの使用について、*bon* 全体の生起頻度における割合を比較したところ、以下の通りとなった。

DMの <i>bon</i>	初中級学習者=上級学習者<母語話者
----------------	-------------------

したがって、このディスコースマーカの使用が少ないため、*bon* の過少使用が引き起こされたと考えられる。

表 260 学習者による *bon* の共起語

L3 頻度	L2 頻度	L1 頻度	KW 頻度	R1 頻度	R2 頻度	R3 頻度
C' 37.66 (13)	C' 43.46 (15)	TRÈS 43.46 (15)	BON 188.31 (65)	-	-	-
	EST 28.97 (10)	INT 40.56 (14)				
		EST 28.97 (10)				

表 261 母語話者による *bon* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	MAIS	99.1 (65)	BON	506.17 (332)	C'	45.74 (30)	EST	41.16 (27)	-	-
				HESIT	32.02 (21)			HESIT	33.54 (22)				
				INT	22.87 (15)			JE	21.34 (14)				

### 7.2.2.1 5. *tout*

まず、*tout* の使用頻度を母語話者、上級学習者、初中級学習者の間で比較する。

表 262 *tout* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
480.25 (315)	207.93 (51)	110.10 (11)	$\chi^2=56.29$ ; $p<.001$

有意差が見られたため 3 グループの間を比較したところ、母語話者は学習者よりも有意に多く *tout* を使用しており、学習者間の間には有意差が見られなかった。

<i>tout</i>	初中級学習者 = 上級学習者 < 母語話者
-------------	-----------------------

学習者のコロケートをしてみると、*tout* の使用の多くは *pas du tout* であり、両コーパスで現れていた。そこで、*tout* の使用頻度における *pas du tout* の使用の割合について、比較したところ、以下の通りとなった。

表 263 *pas du tout* のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>pas du tout</i>	22.87 (15)	46.35 (16)	2.90 (1)	$\chi^2=40.02$ ; $p<.001$

使用に有意差が見られたため多重比較を行ったところ、以下のような結果となった。

<i>pas du tout</i>	上級学習者 > 初中級学習者 = 母語話者
--------------------	-----------------------

したがって、上級学習者の *tout* の使用の多くは *pas du tout* の一部として用いられており、母語話者と比べてもその頻度割合が高いということが分かった。つまり、上級学習者の *pas du tout* が *tout* の現れる主なコンテキストであるといえる。

一方、母語話者のコロケートをみると、*tout à fait* (副詞句としての「完全に」や、相手への完全同意としての「まさに」) という MWU や曖昧性付加句である *et tout (ça)* (など) など口語でよく使用される MWUs が含まれているのが分かる。これらの使用について、各コーパスの頻度をカウントしたところ以下のようなになった。

表 264 *tout* との共起頻度

	母語話者	上級学習者	初中級学習者
<i>tout à fait</i> (副詞)	4.57 (3)	8.15 (2)	0 (0)
<i>tout à fait</i> (同意)	32.02 (21)	8.15 (2)	0 (0)
<i>et tout (ça)</i>	64.03 (42)	12.13 (3)	0 (0)

このとおり、初級学習者にはこれらの使用は一例も見られず、上級学習者には数例しか見られなかった。したがって、このような使用の数が学習者にあまり見られなかったことにより、過少使用が引き起こされたものと考えられる。

表 265 学習者による *tout* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	PAS	49.25 (17)	DU	49.25 (17)	TOUT	179.62 (62)	-	-	-	-	-	-

表 266 母語話者による *tout* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	PAS	25.92 (17)	ET	71.66 (47)	TOUT	480.25 (315)	ÇA	77.75 (51)	FAIT	38.11 (25)	EST	28.97 (19)
		C'	22.87 (15)	DU	36.59 (24)			LE	59.46 (39)	MONDE	30.49 (20)		
				EST	21.34 (14)			À	50.31 (33)	QUI	22.87 (15)		
								CE	32.02 (21)				

#### 7.2.2.1 6. *avoir*

動詞 AVOIR の不定詞 *avoir* の使用頻度を母語話者、上級学習者、初中級学習者の間で比較する。

表 267 *avoir* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
147.89 (97)	12.23 (3)	0 (0)	NA

初中級学習者による使用が一例も見られなかったため、母語話者と上級学習者を比較した。その結果、母語話者と上級学習者には有意差が見られた( $\chi^2=29.64$ ;  $p<.001$ )。したがって、レベルを問わず学習者はこの不定詞の使用頻度が低いということが考えられる。そこで、母語話者コーパスの L1 位置に多く見られた前置詞 *de* の省略形の *d'*と、中性代名詞の *y* との共起について概観した。母語話者コーパス内で 5 例見られた *d'avoir* という MWU は、*c'est*+形容詞+*d'avoir* という形で用いられていた。また、*y avoir* という MWU は、*il y a* という MWU に助動詞が付加されて、動詞 AVOIR の三人称単数現在形である *a* が *avoir* という不定詞に変わったというコンテキスト内で 15 例見られた。いずれのケースにせよ、学習者にはこのような使用が一例も見られなかった。



**c'est** peut être **compliqué d'avoir** une politique à long terme (FL07)

même si **c'est** très **dur** dans une université **d'avoir** l'ensemble de l'information quand il y a vingt cinq mille étudiants HESIT (FL08)

non mais il **doit** sûrement **y avoir** quelque chose c'est évident (FL18)

ça veut dire qu'il **va y avoir** deux côtés (FL19)

表 268 学習者による *avoir* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	-	-	AVOIR	8.69 (3)	-	-	-	-	-	-

表 269 学習者による *avoir* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	D'	30.49 (20)	AVOIR	147.89 (97)	UN	35.07 (23)	-	-	-	-
				Y	24.39 (16)								

### 7.2.2.1 7. *leur*

母語話者、上級学習者、初中級学習者の *leur* の使用頻度について比較したところ、以下のような結果になった。*leur* は主に「彼らに、彼女らに」という人称代名詞の三人称複数間接目的を示す場合と、「彼らの、彼女らの」という所有形容詞として使用される。

表 270 *leur* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
135.69 (89)	8.15 (2)	0 (0)	NA

初中級学習者による使用が一例も見られなかったため、母語話者と上級学習者を比較した。その結果、母語話者と上級学習者には有意差が見られた ( $\chi^2=28.78$ ;  $p<.001$ )。そこで、母語話者コーパスにおけるこれらの使用を比較したところ、以下の通りとなった。

	母語話者	上級学習者
人称代名詞	68.61 (45)	4.08 (1)
所有形容詞	65.56 (43)	4.08 (1)

*leur* の使用頻度は人称代名詞も所有形容詞もおおよそ同じくらいであった。学習者コーパスではそれぞれ 1 例ずつしか見られなかった。

人称代名詞

on **leur** a appris à être chercheur (FL03)

dès que les NP seront installés on pourra **leur** offrir des connexions HESIT (FL08)

et qu'on **leur** dit que vous avez cours à l'NP (FL12)

mais on **leur** a donné trente minutes de temps libre pour faire pour acheter des souvenirs (JA11)

所有形容詞

à la fac enfin ils ont déjà **leur** bac tout ça (FL04)

enfin l'université c'est **leur** vie quoi en gros (FL14)

et tu cherches **leur** signification et voilà (FL17)

et la famille utilisent l'anglais dans dans la dans le travail dans dans **leur** travail (JA14)

表 271 学習者による *leur* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	-	-	LEUR	5.79 (2)	-	-	-	-	-	-

表 272 母語話者による *leur* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
-	-	-	-	ON	24.39 (16)	LEUR	135.69 (89)	-	-	-	-	-	-

### 7.2.2.1 8. *même*

*même* の使用頻度について、母語話者、上級学習者、初中級学習者の頻度を比較したところ、以下のようになった。

表 273 *même* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
385.72 (253)	167.16 (41)	70.06 (7)	$\chi^2=48.14$ ; $p<.001$

有意差が見られたので、多重比較を行いグループ間の比較を行った。その結果、母語話者と学習者グループの間で使用頻度の差が見られたが、学習者間には有意差が見られなかった。

<i>même</i>	初中級学習者=上級学習者<母語話者
-------------	-------------------

次に、MWUs レベルでの比較を行う。学習者コーパスにも母語話者コーパスにも *quand même* という MWU が見られることが分かる。*quand même* の中心的な意味は、主体間の解釈を促進するものであり、Moeschler and Spengler (1981 : 110)はこの MWU について、矛盾を容認させる機能を持つものであると指摘している (Beeching 2007 : 82 より引用)。これは、日本語で言う「それでもやはり」のような意味を持つものである。しかし、このような標準的な使用での反意や譲歩の意味よりも、近年では、発言のニュアンスを弱めるために使用される例の方が多く見られる (Beeching 2007 : 82)。本研究では

この 2 つの違いは区別を行わない。 *quand même* の使用頻度について比較を行ったところ、以下の通りとなった。

表 274 *quand même* のグループ別頻度

	母語話者	上級学習者	初中級学習者
<i>quand même</i>	157.03 (103)	44.85 (11)	10.01 (1)

*même* の使用頻度について *quand même* の占める割合には有意差が見られなかった ( $\chi^2=4.62$ ;  $p>.05$ )。しかし、学習者による *quand même* の使用頻度が低いことには変わりはない。

また、母語話者コーパスのみに多く現れた MWUs は *en même temps* や *même pas* である。まず、「同時に」の意味を持つ *en même temps* について生起頻度を調べたところ、以下の通りとなった。

	母語話者	上級学習者	初中級学習者
<i>en même temps</i>	38.11 (25)	4.08 (1)	0 (0)

次に、「～さえもない」という否定の副詞 *pas* を強調する *même* の使用についての頻度を行ったところ、以下の通りとなった。

	母語話者	上級学習者	初中級学習者
<i>même pas</i>	33.54 (22)	4.08 (1)	0 (0)

このように、これらの使用については学習者コーパスにはあまり見られないことが分かった。したがって、これらの MWUs の使用頻度の低さが過少使用を引き起こした一因であると言える。

表 275 学習者による *même* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
—	—	MAIS	23.18 (8)	QUAND	34.76 (12)	MÊME	139.06 (48)	—	—	—	—	—	—

表 276 母語話者による *même* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
Y	27.44 (18)	EST	22.87 (15)	QUAND	155.51 (102)	MÊME	385.72 (253)	TEMPS	38.11 (25)	C'	24.39 (16)	EST	24.39 (16)
		A	21.34 (14)	EN	38.11 (25)			PAS	33.54 (22)				
				LA	27.44 (18)								

### 7.2.2.1 9. *sais*

*sais* の使用頻度について、母語話者、上級学習者、初中級学習者の頻度を比較したところ、以下のようになった。

表 277 *sais* のグループ別頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
510.74 (335)	220.17 (54)	230.21 (23)	$\chi^2=45.68$ ; $p<.001$

有意差が見られたので、多重比較を行いグループ間の比較を行った結果、母語話者と学習者グループの間で使用頻度の差が見られたが、学習者間には有意差が見られなかった。

<i>sais</i>	初中級学習者 = 上級学習者 < 母語話者
-------------	-----------------------

次に、MWUs レベルの比較を行う。まず、両コーパスにおける *sais* のコロケートをみると、*je* が L1 位置にあり、R1 位置には *pas* がおかれていることより、*je sais pas* (「分からない」という MWU が見られることが分かる。これは、*je ne sais pas* の否定辞 *ne* が脱落した形であり、口語でよく用いられる。また、学習者コーパスには L1 の位置に *ne* が見られることより、否定辞が保持された形である *je ne sais pas* という MWU も含まれると思われる。そこで、*je sais pas* と *je ne sais pas* についてグループ間の生起頻度を比較したところ、以下の通りとなった。

表 278 *je ne sais pas* と *je sais pas* のグループ別頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>je ne sais pas</i>	30.49 (20)	36.69 (9)	40.04 (4)	$\chi^2=10.13$ ; $p<.01$
<i>je sais pas</i>	189.05 (124)	130.47 (32)	80.07 (8)	$\chi^2=9.86$ ; $p<.01$

*sais* の使用頻度に対するこれらの MWUs の生起頻度の割合について有意差があるかどうか調べたところ、*je ne sais pas* についても *je sais pas* についても有意差が見られた。そこで多重比較を行ったところ、どちらの MWUs も上級学習者が母語話者よりもこれらの MWUs を *sais* の使用頻度に対して多い割合で使用していたことが分かった。つまり、上級学習者による *sais* 自体の生起頻度は母語話者よりも有意に少ないが、*sais* の使用は *je ne sais pas* や *je sais pas* といった MWUs に偏っていたと言える。

この *je (ne)sais pas* という MWU は英語の *I don't know* に相当するものである。英語の *I don't know* の機能について Aijmer (2004) は、特に会話において話者の流暢さを促進するための方略であると指摘している。しかし、この MWU は話し手が会話内容に全責任を負っていないということを示唆するため、学習者はこの表現を多用しているが、このことによって母語話者と比べると学習者の自信の無さが強調されているような印象を与えると続けている (ibid: 186-187)。したがって、この MWU の過剰使用は対話者に良い印象を与えない可能性がある。

一方母語話者コーパスをみると、*sais*の主語が *je* だけでなく *tu* も用いられていることが分かる。そこで、*tu sais*(「あなたは知っている」)という MWU は対話者が事の真偽を知っているという事を表すだけでなく、口語では相手の注意を引く際に用いられ、ディスコースマーカのような役割を果たす。そこで、標準的使用の *tu sais* とディスコースマーカ的使用されている *tu sais* について 3 つのコーパスを比較したところ、以下のようになった。

表 279 *tu sais* の機能の比較

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
標準	28.97 (19)	20.39 (5)	10.01 (1)	$\chi^2=1.77$ ; $p>.05$
DM	153.98 (101)	0 (0)	0 (0)	NA

このように、標準の使用における *tu sais* の使用には有意差が見られなかったが、母語話者はディスコースマーカとして *tu sais* を多く使用しているのに対して、上級学習者および初中級学習者による使用は一例も見られなかった。このディスコースマーカの使用が少ない事が、過少使用の一因となったと言える。また、ディスコースマーカ的使用でない場合は標準的使用であると見なしたが、学習者コーパスの *tu sais* のなかには不適切な使用も見られた。これは、動詞 CONNAÎTRE が使用されるべき場面において、SAVOIR が使用された例である。日本語では両方「知っている」と訳す事ができる。しかし、使用方法についてまず統語面で違いがある。主に直接目的語が後ろにおかれるときは動詞 CONNAÎTRE が使用され、動詞 SAVOIR は節を後ろにとる。また意味面でも違いがあり、動詞 CONNAÎTRE はただ単に「知っている」という意味であるのに対して、動詞 SAVOIR は深い知識があるという意味も持つ。このような違いについて学習者は混同している可能性がある。

#### エラー例

c'est FW **tu sais** (JD2)

on a visité HESIT INT NP **tu sais** HESIT c'est à NP il y a musée de NP (JA04)

**je sais je sais** le visage du visage de cette personne (JA15)

#### 標準的使用

donc HESIT comme **tu sais** je vais à NP cet été (JA09)

INT NP NP **tu sais** d'où elle vient aussi (FL17)

**tu sais** pas qui est passé avant toi HESIT (FL18)

#### ディスコースマーカ的使用

**tu sais** des fois l'article défini l'article indéfini ça change beaucoup de sens (FL03)

HESIT **tu sais** c'est la fille aux cheveux rouges là qui est un jour j'étais avec elle et NP et une autre à la cafété (FL18)

表 280 学習者による *sais* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
—	—	JE	52.15 (18)	JE	144.85 (50)	SAIS	223.07 (77)	PAS	159.34 (55)	—	—	—	—
				NE	43.46 (15)								

表 281 母語話者による *sais* の共起語

L3	頻度	L2	頻度	L1	頻度	KW	頻度	R1	頻度	R2	頻度	R3	頻度
—	—	JE	45.74 (30)	JE	286.62 (188)	SAIS	510.74 (335)	PAS	225.64 (148)	SI	32.02 (21)	—	—
		HESIT	36.59 (24)	TU	182.95 (120)			PLUS	54.89 (36)	IL	27.44 (18)		
		ENFIN	35.07 (23)					QUE	33.54 (22)	JE	22.87 (15)		
		MOI	22.87 (15)										

### 7.3. コロケーション分析から見る学習者の使用語彙特徴

本節では、タスクに基づくコーパスと自由会話コーパスのコロケーション分析で一貫して見られた学習者の使用語彙の特徴についてまとめる。学習者の特徴は、人称代名詞の使用、否定辞の保持、短縮形や俗語の過少使用、曖昧性付加句の過少使用、ディスコースマーカの過少使用に代表された。

#### 7.3.1. 人称代名詞

タスクに基づくコーパスにも自由会話コーパスでも、人称代名詞主語である *j(e)* の過剰使用が見られた。これは、主に学習者の一回の発話ターンに含まれる語数が少ないため、主語を必ず置かなくてはならないフランス語では、発話者を指す一人称代名詞主語を多く使用したものであると考えられる。しかしこの要因以外にも、母語話者が人称主語よりも非人称主語を好むことから人称代名詞主語の頻度が高くなった例も見られた。たとえば、学習者は *je dois*+動詞を使用するのに対して、母語話者は *il faut*+動詞を選択するという結果は、タスクに基づくコーパスから明らかになったことである。また、同様の傾向が *vous* にも見られた。道順を説明するタスクについて、学習者は *vous prenez* と *vous* を主語にとって「(あなたは)～に乗る」と説明するのに対して、母語話者は *il faut prendre* というように「～に乗らなければならない」というように非人称主語である *il* を用いた構文を多く用いていた。

*je* や *vous* のように過剰使用される人称代名詞があった一方で、*on* は過少使用されていた例がタスクに基づくコーパスに見られた。これは、学習者が *on* を使用する代わりに *nous* を使用していたことに起因していた。*on* には様々な機能がある。たとえば、*quelqu'un* (誰か)、*un homme* (ある人) というような曖昧な意味を持つ。また、1人以上の特定の人を示すこともある。しかし、特に口語では *nous* の意味でよく使用されている (Grevisse 1986)。母語話者の会話では *nous* よりも *on* がより多く使用されているが、特に初級学習者はこの *on* を過少使用していた。また、自由会話コーパスには第三人称複数

主語の *ils* の過少使用が見られた。この過少使用は、学習者が第三人称複数主語自体をあまり使用していなかったことに起因していた。

### 7.3.2. 否定辞

タスクに基づくコーパスでは、過剰使用語として否定辞 *ne* が抽出された。*ne* のコロケーション分析より、学習者は否定辞 *ne* を保持するが、母語話者はこの否定辞を省略する傾向にあることが分かった。また、*c'* のコロケーション分析から、タスクに基づくコーパスも自由会話コーパスからも学習者による *c'est pas* という MWU の過少使用が見られた。これは、学習者が *c'est pas* という否定辞 *ne* が省略された形の MWU よりも、否定辞を保持した *ce n'est pas* という MWU を好むことに起因している。さらに自由会話コーパス分析より、否定辞 *ne* の保持は特に初中級学習者に多く見られるということが分かった。実際に、初級学習者コーパスであるタスクに基づくコーパスにもこの傾向が見られたことより、フランス語の学習初期に否定辞 *ne* の保持が見られると結論づけることが出来るだろう。

しかし、文法的には否定辞 *ne* の保持が正しく、否定辞の省略の方が本来非文である。特に、外国語としてのフランス語教育について考えるとき、初級時から標準的なフランス語の否定形を教えるのか、もしくは話し言葉において頻度が高くより簡単な否定形から教え始めるのかについては大きな問題である。標準的なフランス語の否定形を教えるべきという立場の支持者は、まず複雑な形を習得し、そのあと単純化する方がより学習が簡単であるという仮定のもと、書き言葉に見られる非連続的な否定形をまず学習することがより簡単でより正確であると主張している。しかし、非連続的な否定は、2 つの否定形が動詞を軸に異なる場所に位置されるという点をまず考慮すべきである。次に、ゲルマン語系の否定形と否定の簡略形の位置は一致しており、ゲルマン語系の話者にとって学習が容易となるだけでなく、第一外国語としてこれらの言語を学習した者にとっても学習が容易になると考えられる (Andersen 2004: 205)。本研究では、記述の通り初級学習者に *ne* の保持が見られ、上級学習者に *ne* の脱落が見られた。否定辞 *ne* の脱落は上級者に見られたことより、否定形の習得が困難なことからこの否定辞 *ne* の省略が見られた訳ではない。つまり、上級学習者は標準的な否定形を学習した上で、否定辞の省略を行っている。実際に本研究に置ける上級学習者はフランス語圏に 1 年以上留学した経験があるため、母語話者の非標準的使用に触れることによって、否定辞 *ne* の脱落を行うようになったと考えられる。留学経験のない学習者には、否定辞の脱落が非文法的であれ母語話者の脱落頻度が高い以上、標準的な否定形を学習した後に否定辞 *ne* の省略の現象を明示的に説明する必要がある。

### 7.3.3. 短縮形・俗語

学習者による語の短縮形や俗語の過少使用は、タスクに基づくコーパスにも自由会話コーパスにも見られた現象である。たとえば、タスクに基づくコーパス分析から学習者は *football* を使用していたのに対して母語話者は *foot* を使用していたということが分かった。

また、学習者が標準的な語である *université* を使用していたのに対して、母語話者は *fac* という平俗な形を好んで使用していた。また、自由会話コーパス分析から、学習者は標準形である *oui* を使用するのに対して、母語話者は *ouais* を多く使用することが明らかになった。また、学習者のレベルが上がるにつれて平俗な形の *ouais* を使用する頻度が高くなることも分かった。このように短縮形や俗語の使用が留学経験者に多くなるという点は、Thomas (2008)による先行研究の結果を支持することになった。

語の短縮形や俗語の使用についても、否定辞 *ne* と同様にフランス語教育で考慮しなければならない点である。なぜなら、これらも非標準的なフランス語であるためである。特に、俗語については言語使用コンテクストをよく知った上でのみ使用すべきである。つまり、場面に応じて言語レベルを適応させる能力が必要である。本研究で使用したコーパスはモノログと友人間の会話であったため、平俗な語を使用しても問題ではない。しかし、初中級学習者がこれらの語を使用すると、聞き手に良い印象を与えない可能性があるため、安易に使用すべきではないと考える。

#### 7.3.4. 曖昧性付加句

タスクに基づくコーパスにも自由会話コーパスにも、曖昧性付加句の過少使用が見られた。タスクに基づくコーパスでは *à peu près*(おおよそ)という MWU が過少使用として抽出され、*et tout*(~など)という MWUs はタスクに基づくコーパスにも自由会話コーパスにも過少使用 MWU として抽出された。

Channell(1994:21)は、英語学習者の外国語らしさの要因について以下のように述べている：

It is often noticed by teachers that the English of advanced students, while grammatically, phonologically and lexically correct, may sound rather bookish and pedantic to a native speaker. This results in part from an inability to use appropriate vague expression.

(教師によってしばしば指摘される点であるが、文法的、音韻的、語彙的に正しくても、英語上級学習者の英語は、堅苦しく、もったいぶったように母語話者の耳に聞こえてしまうことがある。これは、曖昧表現を適切に使用できないことに起因している。)

実際に、ディベートのような公式な場では母語話者は明確にそして詳細に話す。非公式な場では情報の詳細さよりも話者間の接触を築いて維持することに重きが置かれるため、話者は通常明確さや正確さにこだわらず会話を行う (De Cock 2004: 236)。この現象は英語学習者に限定されず、他の言語にも適応できると言える。フランス語母語話者による会話には、曖昧性が高く、非具体的な表現も良く使用されている (Andersen 2006: 57)。特に、形式張らない会話のような対人コミュニケーションで必要とされ、直接的になりすぎてしまわないように語調を和らげるような効果もある (McCarthy and Carter 1997: 36)。つまり、間主観性の指標であり、対人レベルにおける非公式的なインタラクションにおいて重要な役割を果たすと考えられている。つまり、経験の分かち合いや社会的な親密さを相手に伝える役割をする (De Cock 2004: 236)。このように、曖昧語を使用することによって対話者へ情報を無理強いしていないという意味を示すことができるため、重要な対人



ストラテジーであると考えられている。一方、多用されすぎると怠惰でだらしない印象を与えるため、嫌厭されることもある (McCarthy and Carter 1997: 36-37)。本分析の結果、学習者は母語話者が使っている機能と同じようには、これらの語を使用していなかった。

### 7.3.5. ディスコースマーカ―

英語学習者によって使用されるディスコースマーカ―のレンジや種類に関する研究は大変限定的である (Fung and Carter 2007: 410)。これは、フランス語学習者に関しても同様のことである。

ディスコースマーカ―は分節内の言語単位や超分節の言語単位と定義でき、談話レベルでは非命題的機能や、つなぎ語の機能を幅広くカバーする。会話の段階を展開させる連結部であることを知らせ、前のコンテキストとの発言の関連を示し、話し手と聞き手とメッセージの間の相互関係を指し示す (Fung and Carter 2007: 411)。そして、ディスコースマーカ―は発話の解釈自体に影響をほとんど与えず、真理条件的な意味で表れることはない上、命題の役割も持つことはない。したがって、これらの言語指標によって命題の内容は変わらないが、聞き手の発話解釈を促し、情報を共有しているという事実を聞き手に知らせる機能を持つ (Aijmer 2004: 176-177)。

またディスコースマーカ―は、ためらい表現の役割を果たすこともある。思考をまとめるために時間を稼いだり、躊躇している間を持たせたり、そして適切な語を探したりするための方略的な単語や表現を使いながら、学習者が口語で自主的に自分の考えを表現できるように指導することは重要である。これらの語や表現は、母語話者にとっても良く使用されていることは明白で、外国語で何かを表現するときにはなおさら話者がこれらの語を使用する機会が増えると考えられることができる (Andersen 2004: 211)。Guilquin (2008)は実際の学習者による会話について、ポーズや語彙以外の手段を過剰使用するが、話し言葉の中に現れる”small words”を過少使用していると指摘している。そしてその結果、語彙以外のためらい手段を使用することで、聞き手に与える流暢さに関する印象が悪くなる。このため、学習者の口頭産出は母語話者と「異なっている」というよりは何か「欠如している」というように評価されてきており、特に上級の外国語教育カリキュラムに組み込むべきであると言われてきた。これは、ためらい表現が自然会話に不可欠であるため、これらを教育から除外するのではなく、学習者にインタラクションをより阻害しないためのためらいのテクニックを学習者に身につけさせようという考えに起因している (ibid: 140-144)。

このように、本研究の結果に見られたような方略的表現やディスコースマーカ―の過少使用は、会話を行う際のテクニックの欠如の現れである。たとえば、モノローグの会話に相当するタスクに基づくコーパス分析では、*donc voilà* や *HESIT voilà* のような、自分の発話ターンの終了を示すマーカ―が過少使用されていた。また、ペアによる自由会話のコーパス分析でも同様に、発話ターンの終了を知らせるマーカ―である *quoi* が過少使用されていたほか、前言の補足や訂正の機能を持つ *enfin* や *en fait*、発言のニュアンスを弱める機能を持つ *quand même* や *bon*、情報の共有を促進させる *tu sais*、*tu vois*、*hein* といったディスコースマーカ―が過少使用されていた。上級学習者は *en fait* を母語話者

よりも過剰使用していた例を除いて、これらのディスコースマーカ―は学習者のレベルを問わず過少使用されていた点は、見過ごすことができないだろう。これらのディスコースマーカ―が教育場面でどのように指導されているのか、また全く指導されていないのかについての調査は今後の課題とする。

本章では、特徴語が文脈の中でどのように使用されているのかについて分析を行った。その結果、学習者の使用語彙の様々な傾向が明らかになった。

学習者の使用語彙分析の最後となる次章では、特徴語に限定せず、計量的手法を用いて自動的に特徴的な語連鎖を抽出する。このことによって、より広い観点から学習者による MWUs の特徴を分析することが可能となる。

## 第八章 N-grams 分析

N-grams 分析は、学習者と母語話者の MWUs の使用をより客観的に概観することを目的としている。まず、Kjellmer (1991)の研究結果のように、学習者は MWUs レベルではなく WUs レベルで文を構成するのか、それとも De Cock et al. (1998)や De Cock (2004, 2007)の研究結果のように、学習者も MWUs レベルで発話を行っているのかどうかについて調べる。次に、学習者が繰り返し使用する N-gram を持つとすれば、一部の N-gram に依存しているのかそうでないのかを、母語話者の N-gram の使用と比較する。なお、本分析では 2 語連鎖から 5 語連鎖までを分析対象とする。

まず、各コーパスの 2 語連鎖から 5 語連鎖における N-gram の総数を求める。次に、この総数を基に調整頻度を求める。本分析では、トークンを基に 10 万 N-grams における調整頻度 100 以上、50 以上 100 未満、50 未満といった層を設け、各層ごとの割合を比較した。このことによって、一部の N-gram に使用が偏っているのかどうか分かる。

次に、WordSmith を使用し、特徴的 N-gram を抽出する。そして、特徴的 N-grams について *VocabProfil* を用いてどのような頻度層の語で構成されているのかを分析する。また、この特徴的 N-grams と特徴語に関連があるのかどうかについて分析を行う。最後に、特徴的 N-grams について質的な分析を行う。

### 8.1. タスクに基づくコーパス

#### 8.1.1. 繰り返し使用される N-grams 数の比較

WordSmith (ver.6.0)によって各コーパスにおける N-grams のタイプとトークンの総数を求めたところ、以下のような結果となった(表 282)。なお、語彙の豊かさの分析の際には固有名詞、外国語起源の単語、間投詞、ためらい語はコーパスに含まなかったが、特徴語分析同様、N-gram 分析ではこれらの語を全て含むこととする。

表 282 各コーパスの N-gram 総数

	学習者コーパス		母語話者コーパス	
	Type	Token	Type	Token
総語数	1488	28436	2474	51823
2 語連鎖数	7068	26515	14592	49812
3 語連鎖数	11861	24608	27332	47834
4 語連鎖数	14567	22711	34635	45868
5 語連鎖数	15552	20906	37554	43967

続いて、各 N-gram のトークンについて、10 万 N-grams あたりの調整頻度が 100、50、そして 20 という閾値を設け、それに値する実測値を求めたところ、以下ようになった。

表 283 10万トークンあたり生起回数 100、50 および 20 以上となるための実測数

	学習者コーパス			母語話者コーパス		
	100	50	20	100	50	20
2 語連鎖数	26.52	13.26	5.30	49.81	24.91	9.97
3 語連鎖数	24.61	12.30	4.92	47.83	23.92	9.57
4 語連鎖数	22.71	11.35	4.54	45.87	22.93	9.17
5 語連鎖数	20.91	10.45	4.18	43.97	21.98	8.79

以上の回数を基に小数第一位以下を四捨五入し、トークンの調整頻度が 100 以上のタイプ数、50 以上 100 未満のタイプ数、20 以上 50 未満のタイプ数、20 未満のタイプ数に分類を行う。トークンの閾値をもとに、各層に含まれる N-grams の種類とその頻度や割合をまとめたものが以下の表である (表 284)。

表 284 各層における N-gram の数とその割合

N-gram	調整頻度	学習者コーパス		母語話者コーパス	
		頻度	割合(%)	頻度	割合(%)
2 語連鎖	100<	174	2.46	108	0.74
	50 ≦9	180	2.55	172	1.18
	20 ≦9	509	7.20	551	3.78
	1 ≦9	6205	87.79	13761	94.31
3 語連鎖	100<	106	0.89	36	0.13
	50 ≦9	178	1.50	100	0.37
	20 ≦9	463	3.90	364	1.33
	1 ≦9	11114	93.70	26832	98.17
4 語連鎖	100<	63	0.43	9	0.03
	50 ≦9	98	0.67	46	0.13
	20 ≦9	364	2.50	213	0.61
	1 ≦9	14042	96.40	34367	99.23
5 語連鎖	100<	38	0.24	4	0.01
	50 ≦9	69	0.44	24	0.06
	20 ≦9	400	2.57	93	0.25
	1 ≦9	15045	96.74	37433	99.68

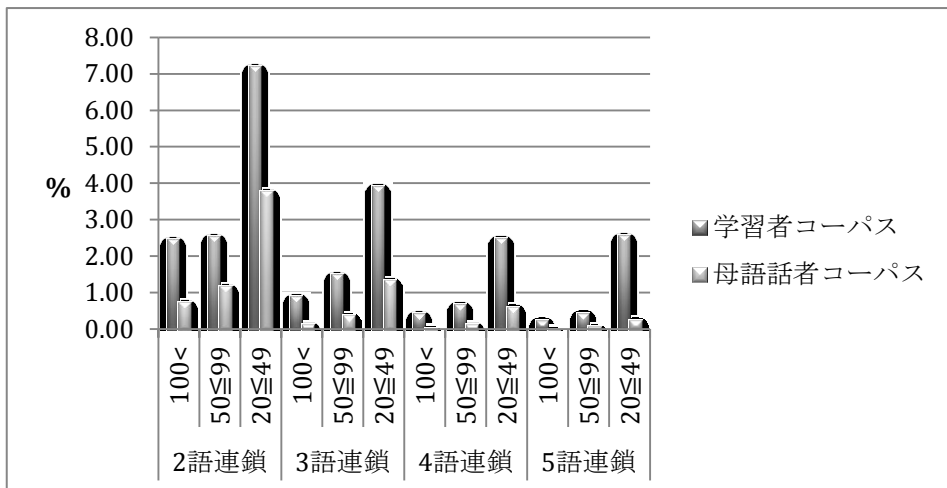


図 17 各層における N-gram のタイプ数とその割合 (タイプ調整頻度 20 以上)

上の図が示す通り、学習者はタイプの調整頻度 20 回以上の N-grams について、母語話者よりも多くの種類の N-grams を使用していることが分かる。また、カイ二乗検定を行いボンフェローニ補正後の有意水準 1.25% で有意差があるかどうかを分析したところ、全ての N-grams の全頻度層に有意差が認められた。調整頻度 20 回以上の N-grams において学習者による使用割合が全て高かったことから、これは学習者がある一部の N-grams を繰り返し使用する傾向にあるということを意味している。一方、母語話者の N-grams の使用は、一部の種類の N-grams に固まること無く、より分散しているということである。

また、トークンについてまとめたものが表 285 である。それぞれの調整頻度層の中に現れた N-grams の頻度の総計とその割合を示している。

表 285 各層における N-gram のトークン数とその割合

N-gram	調整頻度	学習者コーパス		母語話者コーパス	
		頻度総計	割合(%)	頻度総計	割合(%)
2 語連鎖	100<	11009	41.52	12290	24.67
	50≤99	3199	12.06	5917	11.88
	20≤49	3592	13.55	8207	16.48
	1≤19	8715	32.87	23398	46.97
3 語連鎖	100<	4530	18.41	2734	5.72
	50≤99	2871	11.67	3307	6.91
	20≤49	3147	12.79	5138	10.74
	1≤19	14060	57.14	36655	76.63
4 語連鎖	100<	2264	9.97	679	1.48
	50≤99	1472	6.48	1464	3.19
	20≤49	2384	10.50	2711	5.91
	1≤19	16591	73.05	41014	89.42
5 語連鎖	100<	1181	5.65	237	0.54
	50≤99	972	4.65	707	1.61
	20≤49	2163	10.35	1235	2.81
	1≤19	16590	79.36	41788	95.04

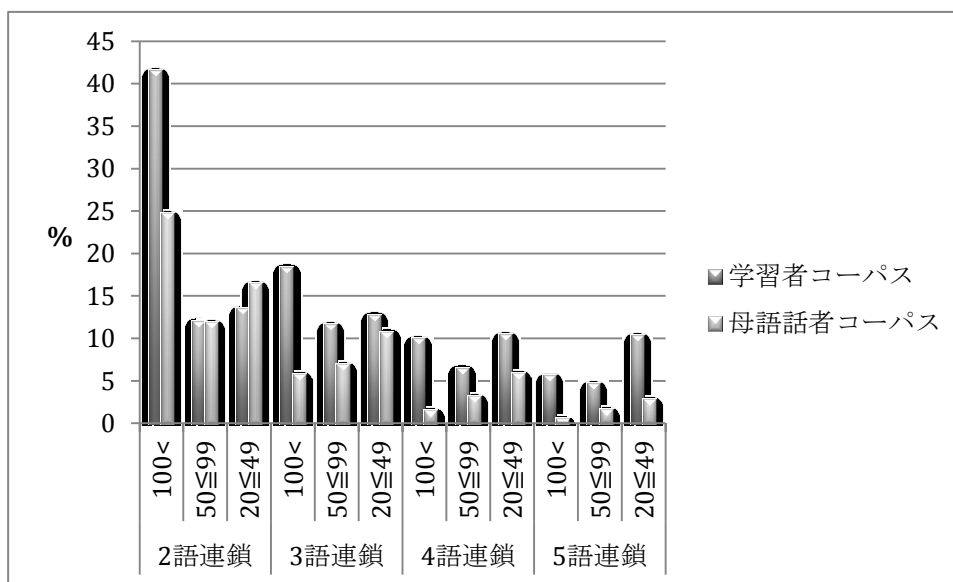


図 18 各層における N-gram のトークン数とその割合 (調整頻度 20 以上)

上の図より、2 語連鎖の調整頻度 20 回以上 50 回未満の N-grams を除いて、調整頻度 20 以上の 2 語連鎖から 5 語連鎖までの N-grams について、学習者によるトークン数が母語話者よりも多いことが分かる。カイ二乗検定より、2 語連鎖の 50 回以上 100 回未満の頻度層を除いて、ボンフェローニ補正後の有意水準 1.25% で 2 グループ間に有意差が見

られた。特に学習者は調整頻度 100 以上の N-grams を繰り返し使用していることが図よりわかる。この層についてタイプ数と比較してみると、タイプ数に対してトークン数が明らかに多いことが分かる。つまり、調整頻度 100 以上の N-grams について、学習者はごく一部の N-grams を多用していることが伺える。

N-gram のタイプ数とトークン数における母語話者と学習者の使用の比較より、学習者は母語話者よりも一部の N-grams を繰り返し使用していることが明らかになった。また、学習者によるそれらの N-grams の使用頻度は非常に高かった。つまり、Kjellmer (1991) の結果とは異なり、De Cock et al. (1998) や De Cock (2004, 2007) の研究結果を追認する結果となった。

次に、各 N-gram のタイプの頻度上位 10 位までに限定し、学習者による高頻度の N-grams の使用の偏りについて精査する。下の表は、各 N-gram の上位 10 位についてまとめたものである。なお、調整頻度は 10 万 N-grams における頻度を指し、レンジはグループ内における N-grams を利用した人数の割合を指す。なお、背景色がついているものは、両コーパスの上位 10 位に含まれていた N-grams である。

また、2 グループにおける上位の N-grams 数に有意差があるのかどうかについて調べる。そのため、同じ順位の N-grams の頻度がそれぞれどれくらいの割合を占めるのかについて、カイ二乗検定を行って有意差を検定する。そのことによって、高頻度の N-grams の数に偏りがあるのかどうか明らかになる。

表 286 各 N-gram の上位 10 位

2 語連鎖	学習者コーパス				母語話者コーパス			
	N-grams	頻度	調整頻度	レンジ	N-grams	頻度	調整頻度	レンジ
1	À NP	381	1436.92	100	C'EST	675	1355.10	100
2	JE SUIS	339	1278.52	100	J'AI	511	1025.86	100
3	J'AIME	325	1225.72	100	À NP	498	999.76	100
4	J'AI	285	1074.86	100	JE SUIS	472	947.56	100
5	DE NP	256	965.49	97.44	J'AIME	429	861.24	100
6	C'EST	214	807.09	97.44	PARCE QUE	257	515.94	97.44
7	NP NP	213	803.32	97.44	NP NP	243	487.83	100
8	PARCE QUE	207	780.69	100	S'IL	236	473.78	100
9	JE VOUDRAIS	202	761.83	100	À LA	221	443.67	94.87
10	NP À	178	671.32	92.31	NP À	206	413.55	100
3 語連鎖	N-grams	頻度	調整頻度	レンジ	N-grams	頻度	調整頻度	レンジ
1	IL Y A	95	386.05	87.18	IL VOUS PLAÎT	173	361.67	100
2	EST-CE QUE	93	377.93	97.44	S'IL VOUS	173	361.67	100
3	J'AIME LE	86	349.48	87.18	J'AIME BIEN	147	307.31	79.49
4	APPELLE NP NP	77	312.91	92.31	IL Y A	144	301.04	84.62
5	HABITE À NP	75	304.78	79.49	EST-CE QUE	124	259.23	97.44
	QU EST-CE	75	304.78	94.87				
6	IL VOUS PLAÎT	70	284.46	94.87	NP À NP	109	227.87	92.31
	S'IL VOUS	70	284.46	94.87				
7	DE NP À	69	280.40	76.92	QUE C'EST	93	194.42	92.31
8	QUATORZE HEURES	66	268.21	89.74	DE NP À	83	173.52	92.31
	QUARANTE	66	268.21	89.74				
9	PARCE QUE JE	64	260.08	79.49	LA LIGNE NP	76	158.88	100
10	JE N'AIME	63	256.01	82.05	JE N' AI	75	156.79	69.23
	L'OFFICIEL DES	63	256.01	97.44				
	OFFICIEL DES SPECTACLES	63	256.01	97.44				



4 語連鎖	学習者コーパス				母語話者コーパス			
	N-grams	頻度	調整頻度	レンジ	N-grams	頻度	調整頻度	レンジ
1	QU EST-CE QUE	73	321.43	94.87	S'IL VOUS PLAÎT	173	377.17	100
2	S'IL VOUS PLAÎT	70	308.22	94.87	L'OFFICIEL DES SPECTACLES	73	159.15	97.44
3	JE N'AIME PAS	62	273.00	82.05	DE NP À NP	67	146.07	87.18
	L'OFFICIEL DES SPECTACLES	62	273.00	97.44				
4	IL Y A BEAUCOUP	55	242.17	79.49	JE N'AI PAS	64	139.53	69.23
					PARCE QUE C'EST	64	139.53	87.18
5	À QUATORZE HEURES QUARANTE	53	233.37	79.49	DES SPECTACLES S'IL	60	130.81	94.87
					SPECTACLES S'IL VOUS	60	130.81	94.87
6	DIX SEPT HEURES TRENTE	52	228.96	87.18	JE N'AIME PAS	59	128.63	76.92
	JE N'AI PAS	52	228.96	64.1	OFFICIEL DES SPECTACLES S'	59	128.63	94.87
	SEPT HEURES TRENTE NEUF	52	228.96	87.18				
7	EST-CE QUE TU	51	224.56	84.62	NP IL FAUT PRENDRE	44	95.93	66.67
8	ELLE S'APPELLE NP	48	211.35	84.62	EST-CE QUE TU	43	93.75	69.23
9	Y A BEAUCOUP DE	45	198.14	74.36	ALLER DE NP À	41	89.39	74.36
					POUR ALLER DE NP	41	89.39	76.92
					S'IL TE PLAÎT	41	89.39	79.49
10	À DIX SEPT HEURES	42	184.93	76.92	À NP IL FAUT	40	87.21	58.97
					N'AI PAS DE	40	87.21	61.54
					PRENDRE LA LIGNE NP	40	87.21	66.67
5 語連鎖	N-grams	頻度	調整頻度	レンジ	N-grams	頻度	調整頻度	レンジ
1	DIX SEPT HEURES TRENTE NEUF	52	248.73	87.18	DES SPECTACLES S'IL VOUS	60	136.47	94.87
					SPECTACLES S'IL VOUS PLAÎT	60	136.47	94.87
2	IL Y A BEAUCOUP DE	45	215.25	74.36	OFFICIEL DES SPECTACLES S'IL	59	134.19	94.87
3	QU EST-CE QUE TU	44	210.47	84.62	L'OFFICIEL DES SPECTACLES S'	58	131.92	92.31
4	À DIX SEPT HEURES TRENTE	41	196.12	74.36	ALLER DE NP À NP	41	93.25	74.36
5	ELLE S'APPELLE NP NP	38	181.77	79.49	JE N'AI PAS DE	40	90.98	61.54
					POUR ALLER DE NP À	40	90.98	74.36
6	NP À QUATORZE HEURES QUARANTE	37	176.98	69.23	SEPT RUE DE LA NP	35	79.61	89.74
7	JE M'APPELLE NP NP	36	172.20	89.74	JE M'APPELLE NP NP	34	77.33	87.18
8	ELLE A TRENTE HUIT ANS	35	167.42	84.62	À NP IL FAUT PRENDRE	33	75.06	53.85
	TRENTE HUIT ANS ELLE EST	35	167.42	84.62	CAFÉ S'IL VOUS PLAÎT	33	75.06	84.62
					SOIXANTE SEPT RUE DE LA	33	75.06	84.62
9	A TRENTE HUIT ANS ELLE	34	162.63	82.05	UN CAFÉ S'IL VOUS	32	72.78	82.05
	CAFÉ S'IL VOUS PLAÎT	34	162.63	87.18				
	S'APPELLE NP NP ELLE	34	162.63	74.36	DIX SEPT HEURES TRENTE NEUF	31	70.51	71.79
10	SOIXANTE SEPT RUE DE LA	33	157.85	79.49	NP S'IL VOUS PLAÎT	31	70.51	64.1
	UN CAFÉ S'IL VOUS	33	157.85	84.62				

図 19 は、2 語連鎖の上位 10 位のタイプの調整頻度について、学習者と母語話者を比較したものである。

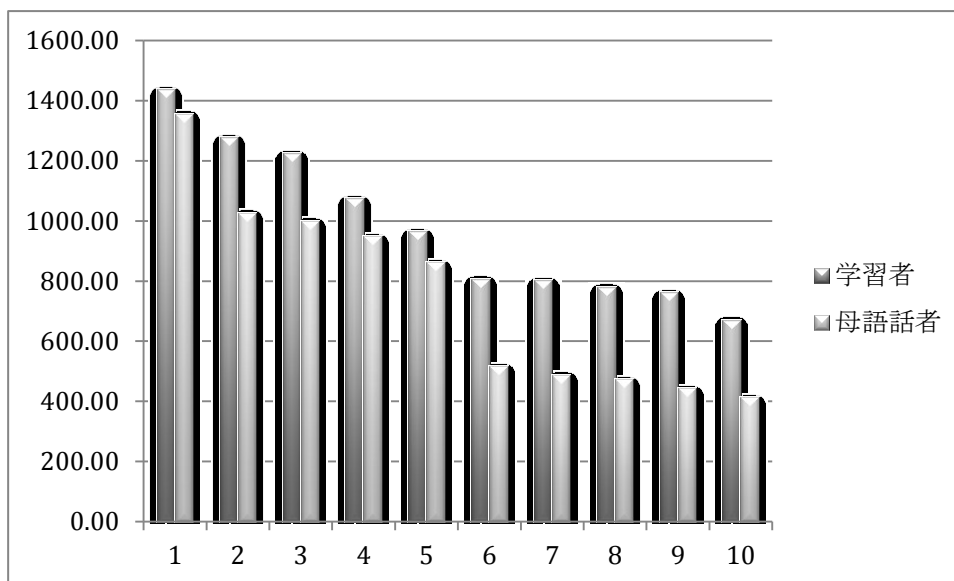


図 19 2 語連鎖上位 10 位の頻度

その結果、頻度上位 1 位から 10 位までの全てについて、学習者の 2 語連鎖数が有意に多いことが分かった。特に母語話者は、第 1 位と第 2 位の差が大きく、また第 6 位から急激に頻度が下がっているが、学習者はそれほど急激な頻度の下降は見られない。

次に、3 語連鎖におけるタイプの上位 10 位までの調整頻度について、学習者と母語話者を比較した結果、第 1 位、第 2 位、第 4 位の頻度には有意差が見られなかったが、その他の頻度には有意差が見られた(図 20)。

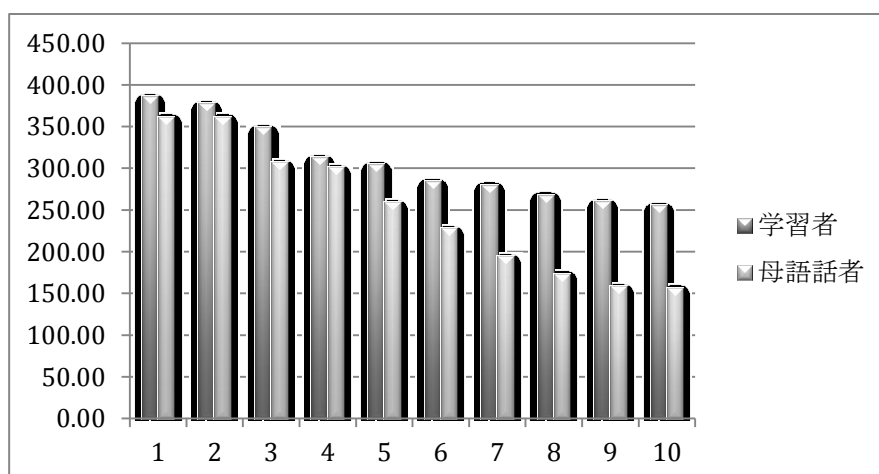


図 20 3 語連鎖上位 10 位の頻度

図 20 から分かるように、母語話者の 3 語連鎖数について第 5 位から急激に頻度が下降している。一方、学習者が使用した頻度上位の 3 語連鎖数の頻度の下降は緩やかである。つまり、学習者の方が N-grams の使用が上位に固まっていると言える。

図 21 は、4 語連鎖におけるタイプの上位 10 位までの調整頻度について、学習者と母語話者を比較したものである。

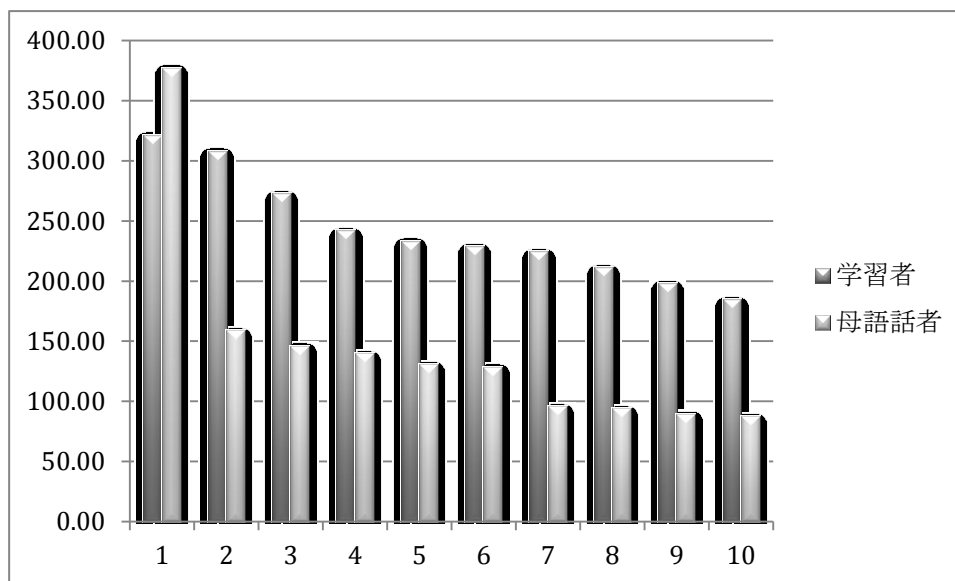


図 21 4 語連鎖上位 10 位の頻度

有意差検定を行った結果、第 1 位のみ有意差が見られなかったが、第 2 位以降には有意差が見られた。4 語連鎖はこれまでの N-grams と様子が異なる。それは、第 1 位の母語話者の使用頻度が学習者よりも高い点である。なお、この 4 語連鎖は *s'il vous plaît* 「お願いします」という N-gram であった。学習者コーパスにもこの N-gram は第二位に見られており、4 語連鎖数における *s'il vous plaît* の使用頻度の割合には有意差が見られなかった( $\chi^2=0.07$ ;  $p>.05$ )。また、母語話者による 4 語連鎖数の第 1 位から第 2 位にかけての下降は急であるが、学習者の下降は緩やかである。つまりここでもまた、学習者による一部の 4 語連鎖の使用頻度が高いことが明らかになった。

5 語連鎖については、以下の通りとなった。

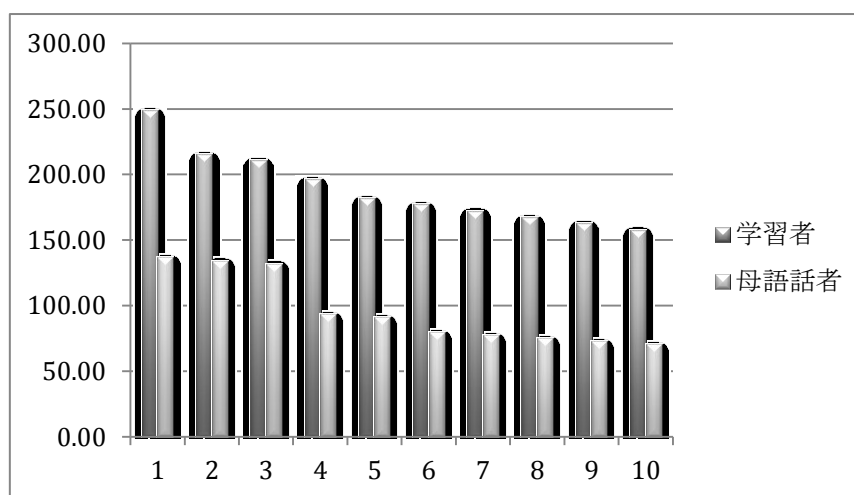


図 22 5 語連鎖上位 10 位の頻度

5 語連鎖の使用頻度の差について有意差検定を行ったところ、全てに有意差が見られた。つまり、5 語連鎖数全体に対しての学習者が使用する頻度上位の 5 語連鎖が占める割合は、母語話者よりも有意に高い。

この通り、ほとんどの頻度上位の N-grams について、学習者は母語話者よりも多く使用していることが分かった。また、ある一定の順位以降に学習者の N-grams の数は緩やかに下降しているのに対して、母語話者による N-grams は急激に減少し、そのことによって二グループ間の差が見られる傾向もある。つまり、両グループともある一定の頻度が高い N-grams に依存しているが、学習者の方が依存する N-grams の数をはるかに多いと言える。

#### 8.1.2. 特徴的 N-grams と語彙分析の関連

WordSmith Tools(ver.6.0)を使用して、2 語連鎖から 5 語連鎖について特徴的 N-grams を抽出した。なお本分析では、keyness の値について+20 以上を過剰使用 N-grams、-20 以下を過少使用 N-grams と位置付けた。本節では、これらの特徴的 N-grams と今まで分析を行ってきた語彙の豊かさ、特徴語の観点との関連について扱う。

まず、特徴的 N-grams と頻度層との関係を概観する。特徴的 N-grams として抽出されたものは、どのような頻度層の語で構成されているのかを概観する。また、過剰使用 N-grams と過少使用 N-grams にどのような特徴の違いが見られるのかについても分析を行う。

次に特徴的 N-grams について、特徴語との関係を概観する。過剰使用語を含む N-grams がどの程度を占めるのかについてを明らかにする。また過少使用 N-grams についても同様の分析を行う。このことによって、特徴的 N-grams が特徴語からの影響を大きく受けるものなのか、それとも MWUs で特徴的使用となるのかを俯瞰することができる。

##### 8.1.2.1. 特徴的 N-grams と頻度層

過剰使用として特徴的な N-grams の頻度層、学習者コーパスと母語話者コーパスの実測値と調整頻度、そして Keyness をまとめたものが表 287 である。頻度層を見てみると K1 機能語と K1 内容語が大半を占めており、K2 や K3 が含まれる MWUs は少なく、含まれていたとしてもごく限られた語であることが分かる。

表 287 特徴的 2 語連鎖と 3 語連鎖に含まれる K1 レベル以外の単語

頻度層	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
K2	gare	gare	gare	gare
	habite	habite	habite	habite
		étudie	étudie	étudie
		combien	combien	combien
		vélo	vélo	vélo
		printemps	printemps	
K3	mariée	mariée	mariée	café
				thé
				riz
Off-list		désolée	désolée	
			excusez	
			salade	salade

たとえば、K2 頻度層に分類された *gare, habite, étudie, combien, vélo, printemps* は、中頻度語とはいえども学習者になじみが深い語であるといえる。また、K3 に分類された *mariée, thé, riz* も学習の観点に鑑みると、難易度が高い語ではない。Off-list に分類された *désolée, excusez* は、話し言葉で使用される語であるため *Vocabprofil* では低頻度語に分類されているが、これも初級のうちに習得する語である。また、*salade* についても習得が難しい語ではない。

したがって、学習者が母語話者に比べて過剰使用として特徴的に使用する 2 語連鎖から 5 語連鎖までの MWUs は、高頻度語がほとんどを占めており、中頻度語や低頻度語が含まれていたとしても、学習者にとって習得が難しいと考えられる語ではない。

一方、過少使用的 MWUs については、ためらい語や固有名詞を除いて全て高頻度語のみで構成されていた。つまり、過少使用的 MWUs は語の頻度によって決定されるのではなく、むしろ高頻度語の結びつき方法によって決定されるものと考えられる。

表 288 特徴的 5 語連鎖(過剰使用)

順位	特徴的 MWUs	頻度層					NNS	調整頻度	NS	調整頻度	Keyness
1	ELLE S'APPELLE NP NP	K1_F	K1_F	K1_C	NA	NA	38	181.77	0	0.00	78.89
2	S'APPELLE NP NP ELLE	K1_F	K1_C	NA	NA	K1_F	34	162.63	2	4.55	56.88
3	APPELLE NP NP ELLE EST	K1_C	NA	NA	K1_F	K1_F	32	153.07	2	4.55	52.97
4	RUE DE LA NP NP	K1_C	K1_F	K1_F	NA	NA	25	119.58	0	0.00	51.90
5	IL Y A BEAUCOUP DE	K1_F	K1_F	K1_F	K1_C	K1_F	45	215.25	10	22.74	50.01
6	JE SUIS ALLÉE À NP	K1_F	K1_F	K1_C	K1_F	NA	31	148.28	3	6.82	46.68
7	À QUATORZE HEURES QUARANTE ET	K1_F	K1_C	K1_C	K1_C	K1_F	32	153.07	4	9.10	44.81
8	QU EST-CE QUE TU	K1_F	K1_F	K1_F	K1_F	K1_F	44	210.47	14	31.84	39.48
9	NP NP JE SUIS JAPONAISE	NA	NA	K1_F	K1_F	K1_C	18	86.10	0	0.00	37.36
10	JAPONAIS ELLE EST MARIÉE ET	K1_C	K1_F	K1_F	K3	K1_F	17	81.32	0	0.00	35.29
11	ANS ELLE EST PROFESSEUR DE	K1_C	K1_F	K1_F	K1_C	K1_F	17	81.32	0	0.00	35.29
12	NP ELLE EST JAPONAISE ELLE	NA	K1_F	K1_F	K1_C	K1_F	30	143.50	6	13.65	35.08
13	ELLE EST JAPONAISE ELLE A	K1_F	K1_F	K1_C	K1_F	K1_F	32	153.07	8	18.20	33.39
14	NP NP ELLE EST JAPONAISE	NA	NA	K1_F	K1_F	K1_C	30	143.50	7	15.92	32.50
15	QU EST-CE QUE C'	K1_F	K1_F	K1_F	K1_F	K1_F	25	119.58	4	9.10	32.12
16	ET ELLE A DEUX ENFANTS	K1_F	K1_F	K1_F	K1_C	K1_C	23	110.02	3	6.82	31.77
17	MARIÉE ET ELLE A DEUX	K3	K1_F	K1_F	K1_F	K1_C	23	110.02	3	6.82	31.77
18	PROFESSEUR DE JAPONAIS ELLE EST	K1_C	K1_F	K1_C	K1_F	K1_F	15	71.75	0	0.00	31.13
19	UNE HEURE ET DIX MINUTES	K1_F	K1_C	K1_F	K1_C	K1_C	15	71.75	0	0.00	31.13
20	JE SUIS JAPONAISE J'AI	K1_F	K1_F	K1_C	K1_F	K1_F	15	71.75	0	0.00	31.13
21	JE N' AIME PAS LE	K1_F	K1_F	K1_C	K1_F	K1_F	20	95.67	2	4.55	29.86
22	ELLE EST PROFESSEUR DE JAPONAIS	K1_F	K1_F	K1_C	K1_F	K1_C	14	66.97	0	0.00	29.06
23	NP JE SUIS JAPONAISE J'	NA	K1_F	K1_C	K1_C	K1_F	14	66.97	0	0.00	29.06
24	A TRENTE HUIT ANS ELLE	K1_F	K1_C	K1_C	K1_C	K1_F	34	162.63	12	27.29	28.27
25	EST MARIÉE ET ELLE A	K1_F	K3	K1_F	K1_F	K1_F	21	100.45	3	6.82	28.13
26	PARCE QUE IL Y A	K1_F	K1_F	K1_F	K1_F	K1_F	19	90.88	2	4.55	27.98
27	JE VOUDRAIS RESTER À LA	K1_F	K1_C	K1_C	K1_F	K1_F	13	62.18	0	0.00	26.98
28	EST PROFESSEUR DE JAPONAIS ELLE	K1_F	K1_C	K1_F	K1_C	K1_F	13	62.18	0	0.00	26.98
29	DE JAPONAIS ELLE EST MARIÉE	K1_F	K1_C	K1_F	K1_F	K3	13	62.18	0	0.00	26.98
30	VOUDRAIS RESTER À LA MAISON	K1_C	K1_C	K1_F	K1_F	K1_C	13	62.18	0	0.00	26.98
31	EST-CE QUE C'EST	K1_F	K1_F	K1_F	K1_F	K1_F	25	119.58	6	13.65	26.68
32	ELLE EST MARIÉE ET ELLE	K1_F	K1_F	K3	K1_F	K1_F	20	95.67	3	6.82	26.33
33	DIX SEPT HEURES TRENTE NEUF	K1_C	K1_C	K1_C	K1_C	K1_C	52	248.73	31	70.51	25.37
34	À LA GARE DE NP	K1_F	K1_F	K2	K1_F	NA	12	57.40	0	0.00	24.91
35	SUIS ÉTUDIANTE J'HABITE À	K1_F	K1_C	K1_F	K2	K1_F	12	57.40	0	0.00	24.91
36	QUARANTE ET ARRIVE À NP	K1_C	K1_F	K1_C	K1_F	NA	12	57.40	0	0.00	24.91
37	JE VAIS À L'UNIVERSITÉ	K1_F	K1_C	K1_F	K1_F	K1_C	19	90.88	3	6.82	24.54
38	QUE IL Y A BEAUCOUP	K1_F	K1_F	K1_F	K1_F	K1_C	15	71.75	1	2.27	24.53

順位	特徴的 MWUs	頻度層					NNS	調整頻度	NS	調整頻度	Keyness
39	HEURES QUARANTE ET ARRIVE À	K1_C	K1_C	K1_F	K1_C	K1_F	15	71.75	1	2.27	24.53
40	QUATORZE HEURES QUARANTE ET ARRIVE	K1_C	K1_C	K1_C	K1_F	K1_C	15	71.75	1	2.27	24.53
41	ET ARRIVE À NP À	K1_F	K1_C	K1_F	NA	K1_F	17	81.32	2	4.55	24.25
42	TRENTE HUIT ANS ELLE EST	K1_C	K1_C	K1_C	K1_F	K1_F	35	167.42	16	36.39	23.20
43	C'EST UNE BONNE IDÉE	K1_F	K1_F	K1_F	K1_C	K1_C	11	52.62	0	0.00	22.83
44	J'AI DIX NEUF ANS	K1_F	K1_F	K1_C	K1_C	K1_C	19	90.88	4	9.10	21.68
45	EST JAPONAISE ELLE A TRENTE	K1_F	K1_C	K1_F	K1_F	K1_C	28	133.93	11	25.02	21.34
46	NP EST SOIXANTE SEPT RUE	NA	K1_F	K1_C	K1_C	K1_C	10	47.83	0	0.00	20.75
47	D'ABORD VOUS PRENEZ LE	K1_F	K1_C	K1_F	K1_C	K1_F	10	47.83	0	0.00	20.75
48	MAIS JE NE VEUX PAS	K1_F	K1_F	K1_F	K1_C	K1_F	10	47.83	0	0.00	20.75
49	ELLE A TRENTE HUIT ANS	K1_F	K1_F	K1_C	K1_C	K1_C	35	167.42	18	40.94	20.47

表 289 特徴的 4 語連鎖(過剰使用)

273

順位	特徴的 MWUs	頻度層				NNS	調整頻度	NS	調整頻度	Keyness
1	ELLE S'APPELLE NP	K1_F	K1_F	K1_C	NA	48	211.35	1	2.18	90.77
2	JE NE VEUX PAS	K1_F	K1_F	K1_C	K1_F	29	127.69	0	0.00	60.20
3	APPELLE NP NP ELLE	K1_C	NA	NA	K1_F	34	149.71	2	4.36	56.88
4	IL Y A BEAUCOUP	K1_F	K1_F	K1_F	K1_C	55	242.17	14	30.52	56.83
5	QU EST-CE QUE	K1_F	K1_F	K1_F	K1_F	73	321.43	30	65.41	53.53
6	S'APPELLE NP NP	K1_F	K1_C	NA	NA	41	180.53	7	15.26	51.36
7	Y A BEAUCOUP DE	K1_F	K1_F	K1_C	K1_F	45	198.14	10	21.80	50.01
8	JE NE PEUX PAS	K1_F	K1_F	K1_C	K1_F	36	158.51	5	10.90	48.70
9	SUIS ALLÉE À NP	K1_F	K1_C	K1_F	NA	31	136.50	3	6.54	46.68
10	JE SUIS ALLÉE À	K1_F	K1_F	K1_C	K1_F	38	167.32	7	15.26	46.11
11	QUATORZE HEURES QUARANTE ET	K1_C	K1_C	K1_C	K1_F	36	158.51	6	13.08	45.53
12	ELLE HABITE À NP	K1_F	K2	K1_F	NA	28	123.29	2	4.36	45.18
13	DE LA NP NP	K1_F	K1_F	NA	NA	25	110.08	1	2.18	44.29
14	VOUS PRENEZ LE NP	K1_F	K1_C	K1_F	NA	21	92.47	0	0.00	43.59
15	NP NP ELLE EST	NA	NA	K1_F	K1_F	40	176.13	11	23.98	39.48
16	NP JE SUIS JAPONAISE	NA	K1_F	K1_F	K1_C	19	83.66	0	0.00	39.44
17	ELLE EST PROFESSEUR DE	K1_F	K1_F	K1_C	K1_F	19	83.66	0	0.00	39.44
18	NP ELLE EST JAPONAISE	NA	K1_F	K1_F	K1_C	34	149.71	7	15.26	39.22
19	JAPONAIS ELLE EST MARIÉE	K1_C	K1_F	K1_F	K3	18	79.26	0	0.00	37.36
20	UNE HEURE ET DIX	K1_F	K1_C	K1_F	K1_C	18	79.26	0	0.00	37.36
21	PROFESSEUR DE JAPONAIS ELLE	K1_C	K1_F	K1_C	K1_F	18	79.26	0	0.00	37.36
22	JE VIENS DE NP	K1_F	K1_C	K1_F	NA	17	74.85	0	0.00	35.29
23	NP ELLE HABITE À	NA	K1_F	K2	K1_F	17	74.85	0	0.00	35.29

順位	特徴的 MWUs	頻度層				NNS	調整頻度	NS	調整頻度	Keyness
24	ELLE EST JAPONAISE ELLE	K1_F	K1_F	K1_C	K1_F	33	145.30	8	17.44	35.03
25	MARIÉE ET ELLE A	K3	K1_F	K1_F	K1_F	24	105.68	3	6.54	33.60
26	ET ELLE A DEUX	K1_F	K1_F	K1_F	K1_C	23	101.27	3	6.54	31.77
27	JE SUIS JAPONAISE J'	K1_F	K1_F	K1_C	K1_F	15	66.05	0	0.00	31.13
28	JE VOUDRAIS ALLER AU	K1_F	K1_C	K1_C	K1_F	15	66.05	0	0.00	31.13
29	HEURE ET DIX MINUTES	K1_C	K1_F	K1_C	K1_C	15	66.05	0	0.00	31.13
30	SUIS JAPONAISE J'AI	K1_F	K1_C	K1_F	K1_F	15	66.05	0	0.00	31.13
31	D'ABORD VOUS PRENEZ	K1_F	K1_C	K1_F	K1_C	15	66.05	0	0.00	31.13
32	DE JAPONAIS ELLE EST	K1_F	K1_C	K1_F	K1_F	15	66.05	0	0.00	31.13
33	ELLE A DEUX ENFANTS	K1_F	K1_F	K1_C	K1_C	33	145.30	10	21.80	30.61
34	EST MARIÉE ET ELLE	K1_F	K3	K1_F	K1_F	22	96.87	3	6.54	29.94
35	N' AIME PAS LE	K1_F	K1_C	K1_F	K1_F	20	88.06	2	4.36	29.86
36	EST PROFESSEUR DE JAPONAIS	K1_F	K1_C	K1_F	K1_C	14	61.64	0	0.00	29.06
37	JE VOUDRAIS ALLER À	K1_F	K1_C	K1_C	K1_F	14	61.64	0	0.00	29.06
38	JE VOUDRAIS RESTER À	K1_F	K1_C	K1_C	K1_F	14	61.64	0	0.00	29.06
39	J'Étudie LE FRANÇAIS	K1_F	K2	K1_F	K1_C	14	61.64	0	0.00	29.06
40	EST JAPONAISE ELLE A	K1_F	K1_C	K1_F	K1_F	33	145.30	11	23.98	28.64
41	HEURES QUARANTE ET ARRIVE	K1_C	K1_C	K1_F	K1_C	17	74.85	1	2.18	28.44
42	PARCE QUE IL Y	K1_F	K1_F	K1_F	K1_F	19	83.66	2	4.36	27.98
43	VOUDRAIS RESTER À LA	K1_C	K1_C	K1_F	K1_F	13	57.24	0	0.00	26.98
44	JE NE COMPRENDS PAS	K1_F	K1_F	K1_C	K1_F	13	57.24	0	0.00	26.98
45	SEPT HEURES TRENTE NEUF	K1_C	K1_C	K1_C	K1_C	52	228.96	31	67.59	25.37
46	À LA GARE DE	K1_F	K1_F	K2	K1_F	12	52.84	0	0.00	24.91
47	JE MANGE DU PAIN	K1_F	K1_C	K1_F	K1_C	12	52.84	0	0.00	24.91
48	QUE IL Y A	K1_F	K1_F	K1_F	K1_F	19	83.66	3	6.54	24.54
49	VAIS À L'UNIVERSITÉ	K1_C	K1_F	K1_F	K1_C	19	83.66	3	6.54	24.54
50	PARCE QUE JE PEUX	K1_F	K1_F	K1_F	K1_C	19	83.66	3	6.54	24.54
51	JE VAIS À L'	K1_F	K1_C	K1_F	K1_F	19	83.66	3	6.54	24.54
52	TRENTE HUIT ANS ELLE	K1_C	K1_C	K1_C	K1_F	36	158.51	16	34.88	24.53
53	HUIT ANS ELLE EST	K1_C	K1_C	K1_F	K1_F	36	158.51	16	34.88	24.53
54	QUARANTE ET ARRIVE À	K1_C	K1_F	K1_C	K1_F	15	66.05	1	2.18	24.53
55	EST-CE QUE C'	K1_F	K1_F	K1_F	K1_F	25	110.08	7	15.26	24.39
56	DIX SEPT HEURES TRENTE	K1_C	K1_C	K1_C	K1_C	52	228.96	32	69.77	24.29
57	ET ARRIVE À NP	K1_F	K1_C	K1_F	NA	17	74.85	2	4.36	24.25
58	JE SUIS DÉSOLÉE MAIS	K1_F	K1_F	OFF	K1_F	23	101.27	6	13.08	23.42
59	J'AI MANGÉ DU	K1_F	K1_F	K1_C	K1_F	20	88.06	4	8.72	23.38
60	À QUATORZE HEURES QUARANTE	K1_F	K1_C	K1_C	K1_C	53	233.37	34	74.13	23.33
61	ELLE A TRENTE HUIT	K1_F	K1_F	K1_C	K1_C	37	162.92	18	39.24	23.00
62	J'HABITE À NP	K1_F	K2	K1_F	NA	38	167.32	19	41.42	22.94
63	BONJOUR C'EST NP	K1_C	K1_F	K1_F	NA	11	48.43	0	0.00	22.83
64	EST UNE BONNE IDÉE	K1_F	K1_F	K1_C	K1_C	11	48.43	0	0.00	22.83



順位	特徴的 MWUs	頻度層				NNS	調整頻度	NS	調整頻度	Keyness
65	C'EST UNE BONNE	K1_F	K1_F	K1_F	K1_C	11	48.43	0	0.00	22.83
66	LE NP DE NP	K1_F	NA	K1_F	NA	18	79.26	3	6.54	22.76
67	DE NP POUR NP	K1_F	NA	K1_F	NA	16	70.45	2	4.36	22.40
68	J'AI DIX NEUF	K1_F	K1_F	K1_C	K1_C	19	83.66	4	8.72	21.68
69	AI DIX NEUF ANS	K1_F	K1_C	K1_C	K1_C	19	83.66	4	8.72	21.68
70	ABORD VOUS PRENEZ LE	K1_C	K1_F	K1_C	K1_F	10	44.03	0	0.00	20.75
71	NP EST SOIXANTE SEPT	NA	K1_F	K1_C	K1_C	10	44.03	0	0.00	20.75
72	SPECTACLES ÇA FAIT COMBIEN	K1_C	K1_F	K1_C	K2	10	44.03	0	0.00	20.75
73	JE SUIS DE NP	K1_F	K1_F	K1_F	NA	10	44.03	0	0.00	20.75
74	FAUT UNE HEURE ET	K1_C	K1_F	K1_C	K1_F	10	44.03	0	0.00	20.75
75	MOI MAIS JE NE	K1_F	K1_F	K1_F	K1_F	10	44.03	0	0.00	20.75
76	MAIS JE NE VEUX	K1_F	K1_F	K1_F	K1_C	10	44.03	0	0.00	20.75
77	L'UNIVERSITÉ EN VÉLO	K1_F	K1_C	K1_F	K2	10	44.03	0	0.00	20.75
78	À NP IL FAIT	K1_F	NA	K1_F	K1_C	10	44.03	0	0.00	20.75
79	NP ET J'AI	NA	K1_F	K1_F	K1_F	13	57.24	1	2.18	20.65
80	LA GARE DE NP	K1_F	K2	K1_F	NA	13	57.24	1	2.18	20.65
81	J'AIME LE PRINTEMPS	K1_F	K1_C	K1_F	K2	13	57.24	1	2.18	20.65
82	CE QUE C'EST	K1_F	K1_F	K1_F	K1_F	25	110.08	9	19.62	20.46
83	JE SUIS ÉTUDIANTE J'	K1_F	K1_F	K1_C	K1_F	18	79.26	4	8.72	20.00

表 290 特徴的 3 語連鎖(過剰使用)

順位	特徴的 MWUs	頻度層			NNS	調整頻度	NS	調整頻度	Keyness
1	J'AIME LE	K1_F	K1_C	K1_F	86	349.48	17	35.54	101.20
2	JE SUIS JAPONAISE	K1_F	K1_F	K1_C	46	186.93	0	0.00	95.51
3	ELLE S'APPELLE	K1_F	K1_F	K1_C	49	199.12	1	2.09	92.81
4	JE VOUDRAIS ALLER	K1_F	K1_C	K1_C	38	154.42	0	0.00	78.89
5	ELLE HABITE À	K1_F	K2	K1_F	43	174.74	2	4.18	74.66
6	HABITE À NP	K2	K1_F	NA	75	304.78	23	48.08	69.05
7	NE VEUX PAS	K1_F	K1_C	K1_F	29	117.85	0	0.00	60.20
8	JE NE VEUX	K1_F	K1_F	K1_C	29	117.85	0	0.00	60.20
9	Y A BEAUCOUP	K1_F	K1_F	K1_C	55	223.50	14	29.27	56.83
10	NP ELLE EST	NA	K1_F	K1_F	54	219.44	14	29.27	55.21
11	PRENEZ LE NP	K1_C	K1_F	NA	26	105.66	0	0.00	53.97
12	JE NE PEUX	K1_F	K1_F	K1_C	38	154.42	5	10.45	52.35
13	S'APPELLE NP	K1_F	K1_F	NA	60	243.82	20	41.81	52.09
14	NE PEUX PAS	K1_F	K1_C	K1_F	37	150.36	5	10.45	50.52
15	A BEAUCOUP DE	K1_F	K1_C	K1_F	46	186.93	11	23.00	49.20

順位	特徴的 MWUs	頻度層			NNS	調整頻度	NS	調整頻度	Keyness
16	HEURES QUARANTE ET	K1_C	K1_C	K1_F	38	154.42	6	12.54	49.08
17	VOUS PRENEZ LE	K1_F	K1_C	K1_F	27	109.72	1	2.09	48.29
18	ALLÉE À NP	K1_C	K1_F	NA	31	125.98	3	6.27	46.68
19	SUIS ALLÉE À	K1_F	K1_C	K1_F	38	154.42	7	14.63	46.11
20	QU EST CE	K1_F	K1_F	K1_F	75	304.78	37	77.35	45.97
21	D'ABORD VOUS	K1_F	K1_C	K1_F	22	89.40	0	0.00	45.67
22	NP NP ELLE	NA	NA	K1_F	45	182.87	12	25.09	45.25
23	LA NP NP	K1_F	NA	NA	25	101.59	1	2.09	44.29
24	PROFESSEUR DE JAPONAIS	K1_C	K1_F	K1_C	21	85.34	0	0.00	43.59
25	APPELLE NP NP	K1_C	NA	NA	77	312.91	41	85.71	43.31
26	À LA GARE	K1_F	K1_F	K2	24	97.53	1	2.09	42.30
27	JAPONAIS ELLE EST	K1_C	K1_F	K1_F	20	81.27	0	0.00	41.51
28	MAIS JE NE	K1_F	K1_F	K1_F	32	130.04	5	10.45	41.49
29	ET ELLE A	K1_F	K1_F	K1_F	28	113.78	3	6.27	41.03
30	NP DE NP	NA	K1_F	NA	52	211.31	20	41.81	40.37
31	EST PROFESSEUR DE	K1_F	K1_C	K1_F	19	77.21	0	0.00	39.44
32	JE MANGE DU	K1_F	K1_C	K1_F	19	77.21	0	0.00	39.44
33	ELLE EST JAPONAISE	K1_F	K1_F	K1_C	37	150.36	9	18.82	39.20
34	JE VAIS À	K1_F	K1_C	K1_F	32	130.04	6	12.54	38.53
35	J'AIME LES	K1_F	K1_C	K1_F	28	113.78	4	8.36	37.51
36	ÇA FAIT COMBIEN	K1_F	K1_C	K2	18	73.15	0	0.00	37.36
37	DE JAPONAIS ELLE	K1_F	K1_C	K1_F	18	73.15	0	0.00	37.36
38	JE VIENS DE	K1_F	K1_C	K1_F	18	73.15	0	0.00	37.36
39	HEURE ET DIX	K1_C	K1_F	K1_C	18	73.15	0	0.00	37.36
40	NP POUR NP	NA	K1_F	NA	26	105.66	3	6.27	37.30
41	JE SUIS ALLÉE	K1_F	K1_F	K1_C	49	199.12	20	41.81	36.14
42	MARIÉE ET ELLE	K3	K1_F	K1_F	25	101.59	3	6.27	35.45
43	VIENS DE NP	K1_C	K1_F	NA	17	69.08	0	0.00	35.29
44	NP JE VOUDRAIS	NA	K1_F	K1_C	17	69.08	0	0.00	35.29
45	JE SUIS JAPONAIS	K1_F	K1_F	K1_C	17	69.08	0	0.00	35.29
46	JE VOUDRAIS RESTER	K1_F	K1_C	K1_C	17	69.08	0	0.00	35.29
47	PARCE QUE IL	K1_F	K1_F	K1_F	26	105.66	4	8.36	33.91
48	ELLE A DEUX	K1_F	K1_F	K1_C	35	142.23	11	23.00	31.67
49	JAPONAISE J'AI	K1_C	K1_F	K1_F	15	60.96	0	0.00	31.13
50	SUIS JAPONAISE J'	K1_F	K1_C	K1_F	15	60.96	0	0.00	31.13
51	VOUDRAIS ALLER AU	K1_C	K1_C	K1_F	15	60.96	0	0.00	31.13
52	ABORD VOUS PRENEZ	K1_C	K1_F	K1_C	15	60.96	0	0.00	31.13
53	ÉTUDIE LE FRANÇAIS	K2	K1_F	K1_C	15	60.96	0	0.00	31.13
54	ANS ELLE EST	K1_C	K1_F	K1_F	43	174.74	18	37.63	31.00
55	DIX NEUF ANS	K1_C	K1_C	K1_C	26	105.66	5	10.45	30.95
56	EXCUSEZ MOI MAIS	OFF	K1_F	K1_F	18	73.15	1	2.09	30.40

順位	特徴的 MWUs	頻度層			NNS	調整頻度	NS	調整頻度	Keyness
57	ET ARRIVE À	K1_F	K1_C	K1_F	24	97.53	4	8.36	30.35
58	EST JAPONAISE ELLE	K1_F	K1_C	K1_F	34	138.17	11	23.00	30.15
59	TU ES LIBRE	K1_F	K1_C	K1_C	22	89.40	3	6.27	29.94
60	VOUDRAIS ALLER À	K1_C	K1_C	K1_F	14	56.89	0	0.00	29.06
61	MANGE DU PAIN	K1_C	K1_F	K1_C	14	56.89	0	0.00	29.06
62	VOUDRAIS RESTER À	K1_C	K1_C	K1_F	14	56.89	0	0.00	29.06
63	MOI MAIS JE	K1_F	K1_F	K1_F	14	56.89	0	0.00	29.06
64	DE NP ET	K1_F	NA	K1_F	23	93.47	4	8.36	28.59
65	J'AIME L'	K1_F	K1_C	K1_F	17	69.08	1	2.09	28.44
66	QUARANTE ET ARRIVE	K1_C	K1_F	K1_C	17	69.08	1	2.09	28.44
67	HUIT ANS ELLE	K1_C	K1_C	K1_F	38	154.42	16	33.45	27.24
68	DE LA SALADE	K1_F	K1_F	OFF	13	52.83	0	0.00	26.98
69	JE NE COMPRENDS	K1_F	K1_F	K1_C	13	52.83	0	0.00	26.98
70	LE LIGNE NP	K1_F	K1_C	NA	13	52.83	0	0.00	26.98
71	NE COMPRENDS PAS	K1_F	K1_C	K1_F	13	52.83	0	0.00	26.98
72	ELLE A TRENTE	K1_F	K1_F	K1_C	40	162.55	18	37.63	26.93
73	WEEKEND JE VOUDRAIS	K1_C	K1_F	K1_C	16	65.02	1	2.09	26.48
74	À NP AVEC	K1_F	NA	K1_F	16	65.02	1	2.09	26.48
75	ET DIX MINUTES	K1_F	K1_C	K1_C	16	65.02	1	2.09	26.48
76	JAPONAISE ELLE A	K1_C	K1_F	K1_F	36	146.29	15	31.36	26.06
77	DIX SEPT HEURES	K1_C	K1_C	K1_C	54	219.44	33	68.99	25.47
78	HEURES TRENTE NEUF	K1_C	K1_C	K1_C	52	211.31	31	64.81	25.37
79	NP IL FAIT	NA	K1_F	K1_C	12	48.76	0	0.00	24.91
80	MAIS JE VOUDRAIS	K1_F	K1_F	K1_C	12	48.76	0	0.00	24.91
81	ET DEUX ENFANTS	K1_F	K1_C	K1_C	12	48.76	0	0.00	24.91
82	ET PUIS VOUS	K1_F	K1_F	K1_F	12	48.76	0	0.00	24.91
83	JE VOUDRAIS SORTIR	K1_F	K1_C	K1_C	12	48.76	0	0.00	24.91
84	VAIS À L'	K1_C	K1_F	K1_F	19	77.21	3	6.27	24.54
85	QUE IL Y	K1_F	K1_F	K1_F	19	77.21	3	6.27	24.54
86	SEPT HEURES TRENTE	K1_C	K1_C	K1_C	52	211.31	32	66.90	24.29
87	NP À DIX	NA	K1_F	K1_C	51	207.25	31	64.81	24.23
88	PARCE QUE JE	K1_F	K1_F	K1_F	64	260.08	46	96.17	23.56
89	SUIS DÉSOLÉE MAIS	K1_F	OFF	K1_F	23	93.47	6	12.54	23.42
90	AI MANGÉ DU	K1_F	K1_C	K1_F	20	81.27	4	8.36	23.38
91	NON JE N'	K1_C	K1_F	K1_F	44	178.80	25	52.26	22.85
92	JE VOUDRAIS FAIRE	K1_F	K1_C	K1_C	11	44.70	0	0.00	22.83
93	UNE BONNE IDÉE	K1_F	K1_C	K1_C	11	44.70	0	0.00	22.83
94	SON ADRESSE DE	K1_F	K1_C	K1_F	11	44.70	0	0.00	22.83
95	EST UNE BONNE	K1_F	K1_F	K1_C	11	44.70	0	0.00	22.83
96	BONJOUR C'EST	K1_C	K1_F	K1_F	11	44.70	0	0.00	22.83
97	GARE DE NP	K2	K1_F	NA	14	56.89	1	2.09	22.58

順位	特徴的 MWUs	頻度層			NNS	調整頻度	NS	調整頻度	Keyness
98	C'EST COMBIEN	K1_F	K1_F	K2	14	56.89	1	2.09	22.58
99	À L'UNIVERSITÉ	K1_F	K1_F	K1_C	60	243.82	43	89.89	22.19
100	AI DIX NEUF	K1_F	K1_F	K1_F	19	77.21	4	8.36	21.68
101	J'AI MANGÉ	K1_F	K1_F	K1_C	40	162.55	22	45.99	21.63
102	PRENDS LE NP	K1_C	K1_F	NA	17	69.08	3	6.27	21.00
103	JE BOIS DU	K1_F	K1_C	K1_F	30	121.91	13	27.18	20.94
104	UNIVERSITÉ EN VÉLO	K1_C	K1_F	K2	10	40.64	0	0.00	20.75
105	SUIS DE NP	K1_F	K1_F	NA	10	40.64	0	0.00	20.75
106	NP EST SOIXANTE	NA	K1_F	K1_C	10	40.64	0	0.00	20.75
107	SPECTACLES ÇA FAIT	K1_C	K1_F	K1_C	10	40.64	0	0.00	20.75
108	NP ET VOUS	NA	K1_F	K1_F	10	40.64	0	0.00	20.75
109	AVEC TOI CE	K1_F	K1_F	K1_F	10	40.64	0	0.00	20.75
110	ELLE EST UNE	K1_F	K1_F	K1_F	10	40.64	0	0.00	20.75
111	AIME LE PRINTEMPS	K1_C	K1_F	K2	13	52.83	1	2.09	20.65
112	LA GARE DE	K1_F	K2	K1_F	13	52.83	1	2.09	20.65
113	N' EST PAS	K1_F	K1_F	K1_F	15	60.96	2	4.18	20.57
114	J'AIME LA	K1_F	K1_C	K1_F	21	85.34	6	12.54	20.23
115	SUIS ÉTUDIANTE J'	K1_F	K1_C	K1_F	18	73.15	4	8.36	20.00

278

表 291 特徴的 2 語連鎖(過剰使用)

順位	特徴的 MWUs	頻度層		NNS	調整頻度	NS	調整頻度	Keyness
1	JE VOUDRAIS	K1_F	K1_C	202	761.83	59	118.45	192.48
2	JE NE	K1_F	K1_F	166	626.06	53	106.40	148.90
3	AIME LE	K1_F	K1_F	88	331.89	17	34.13	104.64
4	SUIS JAPONAISE	K1_F	K1_C	46	173.49	0	0.00	95.51
5	ELLE S'	K1_F	K1_F	51	192.34	3	6.02	85.34
6	DE NP	K1_F	NA	256	965.49	199	399.50	82.24
7	VOUDRAIS ALLER	K1_C	K1_C	38	143.32	0	0.00	78.89
8	HABITE À	K2	K1_F	94	354.52	34	68.26	76.76
9	NP ELLE	NA	K1_F	91	343.20	32	64.24	75.95
10	PRENEZ LE	K1_C	K1_F	38	143.32	1	2.01	70.46
11	VOUS PRENEZ	K1_F	K1_C	45	169.72	5	10.04	65.29
12	LE FRANÇAIS	K1_F	K1_C	30	113.14	0	0.00	62.28
13	D'ABORD	K1_F	K1_C	46	173.49	7	14.05	60.25
14	NE VEUX	K1_F	K1_C	29	109.37	0	0.00	60.20
15	ELLE A	K1_F	K1_F	92	346.97	45	90.34	56.93
16	PROFESSEUR DE	K1_C	K1_F	27	101.83	0	0.00	56.05

順位	特徴的 MWUs	頻度層		NNS	調整頻度	NS	調整頻度	Keyness
17	JAPONAIS ELLE	K1_C	K1_F	30	113.14	1	2.01	54.32
18	NE PEUX	K1_F	K1_C	39	147.09	5	10.04	54.18
19	NP LIGNE	NA	K1_C	26	98.06	0	0.00	53.97
20	ELLE EST	K1_F	K1_F	153	577.03	113	226.85	53.84
21	LE FILM	K1_F	K1_C	29	109.37	1	2.01	52.31
22	VEUX PAS	K1_F	K1_F	29	109.37	1	2.01	52.31
23	ET VOUS	K1_F	K1_F	28	105.60	1	2.01	50.30
24	APPELLE NP	K1_C	NA	96	362.06	55	110.42	49.38
25	MANGE DU	K1_C	K1_F	23	86.74	0	0.00	47.74
26	A BEAUCOUP	K1_F	K1_C	58	218.74	21	42.16	47.29
27	MAIS JE	K1_F	K1_F	85	320.57	46	92.35	46.90
28	JE VEUX	K1_F	K1_C	46	173.49	12	24.09	46.86
29	QUARANTE ET	K1_C	K1_F	38	143.32	7	14.05	46.11
30	ALLÉE À	K1_C	K1_F	38	143.32	7	14.05	46.11
31	QU EST	K1_F	K1_F	75	282.86	37	74.28	45.97
32	ABORD VOUS	K1_C	K1_F	22	82.97	0	0.00	45.67
33	JE SOUVENT	K1_F	K1_C	22	82.97	0	0.00	45.67
34	S'APPELLE	K1_F	K1_C	63	237.60	27	54.20	44.46
35	ANS ELLE	K1_C	K1_F	52	196.12	18	36.14	43.89
36	TU PEUX	K1_F	K1_C	48	181.03	15	30.11	43.61
37	AVEC TOI	K1_F	K1_F	38	143.32	9	18.07	40.85
38	ET ARRIVE	K1_F	K1_C	29	109.37	4	8.03	39.32
39	JE VIENS	K1_F	K1_C	27	101.83	3	6.02	39.17
40	AIME LES	K1_C	K1_F	28	105.60	4	8.03	37.51
41	LA GARE	K1_F	K2	28	105.60	4	8.03	37.51
42	VIENS DE	K1_C	K1_F	18	67.89	0	0.00	37.36
43	FAIT COMBIEN	K1_C	K2	18	67.89	0	0.00	37.36
44	NON JE	K1_F	K1_F	80	301.72	50	100.38	36.59
45	L'UNIVERSITÉ	K1_F	K1_C	76	286.63	46	92.35	36.34
46	SUIS ALLÉE	K1_F	K1_C	50	188.57	21	42.16	35.94
47	LE NP	K1_F	NA	84	316.80	55	110.42	35.90
48	SUIS JAPONAIS	K1_F	K1_C	17	64.11	0	0.00	35.29
49	VOUDRAIS RESTER	K1_C	K1_C	17	64.11	0	0.00	35.29
50	EST JAPONAISE	K1_F	K1_C	38	143.32	12	24.09	34.27
51	PUIS VOUS	K1_F	K1_F	16	60.34	0	0.00	33.21
52	EST NP	K1_F	NA	56	211.20	29	58.22	32.51
53	LE FOOTBALL	K1_F	K1_C	19	71.66	1	2.01	32.37
54	ES LIBRE	K1_F	K1_C	23	86.74	3	6.02	31.77
55	JAPONAISE ELLE	K1_C	K1_F	41	154.63	16	32.12	31.43
56	ELLE HABITE	K1_F	K2	52	196.12	26	52.20	31.39
57	LE LIGNE	K1_F	K1_C	15	56.57	0	0.00	31.13

順位	特徴的 MWUs	頻度層		NNS	調整頻度	NS	調整頻度	Keyness
58	BEAUCOUP DE	K1_C	K1_F	96	362.06	75	150.57	30.43
59	DU RIZ	K1_F	K3	20	75.43	2	4.02	29.86
60	EST TRÈS	K1_F	K1_F	64	241.37	40	80.30	29.26
61	DU PAIN	K1_F	K1_C	32	120.69	10	20.08	29.07
62	VAIS À	K1_C	K1_F	32	120.69	10	20.08	29.07
63	EN VÉLO	K1_F	K2	14	52.80	0	0.00	29.06
64	AIME L'	K1_C	K1_F	17	64.11	1	2.01	28.44
65	QUE IL	K1_F	K1_F	26	98.06	6	12.05	28.33
66	SEPT HEURES	K1_C	K1_C	57	214.97	34	68.26	27.79
67	NEUF ANS	K1_C	K1_C	27	101.83	7	14.05	27.59
68	NE COMPRENDS	K1_F	K1_C	13	49.03	0	0.00	26.98
69	COMPRENDS PAS	K1_C	K1_F	13	49.03	0	0.00	26.98
70	LA SALADE	K1_F	OFF	13	49.03	0	0.00	26.98
71	LE THÉ	K1_F	K3	13	49.03	0	0.00	26.98
72	NP AVEC	NA	K1_F	28	105.60	8	16.06	26.98
73	TRENTE NEUF	K1_C	K1_C	53	199.89	31	62.23	26.52
74	ARRIVEZ À	K1_C	K1_F	16	60.34	1	2.01	26.48
75	MOI MAIS	K1_F	K1_F	29	109.37	9	18.07	26.47
76	ET DIX	K1_F	K1_C	20	75.43	3	6.02	26.33
77	ÉTUDIE LE	K2	K1_F	18	67.89	2	4.02	26.11
78	POUR NP	K1_F	NA	39	147.09	18	36.14	25.60
79	ET ELLE	K1_F	K1_F	44	165.94	23	46.17	25.27
80	VOUDRAIS SORTIR	K1_C	K1_C	12	45.26	0	0.00	24.91
81	ET DEUX	K1_F	K1_C	12	45.26	0	0.00	24.91
82	BONNE IDÉE	K1_C	K1_C	12	45.26	0	0.00	24.91
83	HEURES QUARANTE	K1_C	K1_C	73	275.32	55	110.42	24.74
84	ALLER AU	K1_C	K1_F	57	214.97	37	74.28	24.66
85	NP NP	NA	NA	213	803.32	243	487.83	24.58
86	JAPONAISE J'	K1_C	K1_F	15	56.57	1	2.01	24.53
87	DIX SEPT	K1_C	K1_C	55	207.43	35	70.26	24.50
88	BOIS DU	K1_C	K1_F	32	120.69	13	26.10	23.69
89	À NP	K1_F	NA	381	1436.92	498	999.76	23.67
90	DIX MINUTES	K1_C	K1_C	28	105.60	10	20.08	23.07
91	AI MANGÉ	K1_F	K1_C	41	154.63	22	44.17	22.83
92	BONJOUR C'	K1_C	K1_F	11	41.49	0	0.00	22.83
93	VOUDRAIS FAIRE	K1_C	K1_C	11	41.49	0	0.00	22.83
94	GARE DE	K2	K1_F	14	52.80	1	2.01	22.58
95	AVEC PLAISIR	K1_F	K1_C	14	52.80	1	2.01	22.58
96	EST COMBIEN	K1_F	K2	14	52.80	1	2.01	22.58
97	DU THÉ	K1_F	K3	30	113.14	12	24.09	22.51
98	JE VOIS	K1_F	K1_C	25	94.29	8	16.06	22.33

順位	特徴的 MWUs	頻度層		NNS	調整頻度	NS	調整頻度	Keyness
99	JE N'	K1_F	K1_F	133	501.60	136	273.03	22.18
100	AIME LA	K1_C	K1_F	22	82.97	6	12.05	21.82
101	ÇA FAIT	K1_F	K1_C	36	135.77	18	36.14	21.73
102	OÙ EST	K1_F	K1_F	17	64.11	3	6.02	21.00
103	NP DE	NA	K1_F	63	237.60	48	96.36	20.92
104	MANGÉ DU	K1_C	K1_F	20	75.43	5	10.04	20.87
105	LIGNE À	K1_C	K1_F	10	37.71	0	0.00	20.75
106	RIZ ET	K3	K1_F	10	37.71	0	0.00	20.75
107	SPECTACLES ÇA	K1_C	K1_F	10	37.71	0	0.00	20.75
108	LE CAFÉ	K1_F	K2	10	37.71	0	0.00	20.75
109	TOI CE	K1_F	K1_F	10	37.71	0	0.00	20.75
110	AU MUSÉE	K1_F	K1_C	10	37.71	0	0.00	20.75
111	ET JE	K1_F	K1_F	80	301.72	69	138.52	20.68
112	ME SENS	K1_F	K1_C	13	49.03	1	2.01	20.65
113	LIBRE CE	K1_C	K1_F	15	56.57	2	4.02	20.57
114	AVEC MOI	K1_F	K1_F	25	94.29	9	18.07	20.46
115	À DIX	K1_F	K1_C	82	309.26	72	144.54	20.36
116	PAS ALLER	K1_F	K1_C	18	67.89	4	8.03	20.00

表 292 特徵的 4 語連鎖(過少使用)

順位	特徵的 MWUs	頻度層				NNS	調整頻度	NS	調整頻度	Keyness
1	PRENDRE LA LIGNE NP	K1_C	K1_F	K1_C	NA	1	4.40	40	87.21	-27.68

表 293 特徵的 3 語連鎖(過少使用)

順位	特徵的 MWUs	頻度層			NNS	調整頻度	NS	調整頻度	Keyness
1	J'AIME BIEN	K1_F	K1_C	K1_C	2	8.13	147	307.31	-111.68
2	J'AI PAS	K1_F	K1_F	K1_F	1	4.06	63	131.71	-46.91
3	HESIT C'EST	NA	K1_F	K1_F	1	4.06	52	108.71	-37.66
4	PRENDRE LA LIGNE	K1_C	K1_F	K1_C	1	4.06	40	83.62	-27.68
5	HESIT JE SUIS	NA	K1_F	K1_F	1	4.06	36	75.26	-24.38
6	AIME PAS FAIRE	K1_C	K1_F	K1_C	2	8.13	40	83.62	-23.07
7	POUR ALLER DE	K1_F	K1_C	K1_F	5	20.32	52	108.71	-22.00
8	PARCE QUE HESIT	K1_F	K1_F	NA	1	4.06	33	68.99	-21.93
9	UNE HEURE DIX	K1_F	K1_C	K1_C	2	8.13	38	79.44	-21.52
10	C'EST PAS	K1_F	K1_F	K1_F	1	4.06	31	64.81	-20.30



表 294 特徵的 2 語連鎖(過少使用)

順位	特徵的 MWUs	頻度層		NNS	調整頻度	NS	調整頻度	Keyness
1	AIME BIEN	K1_C	K1_C	3	11.31	150	301.13	-108.06
2	ET HESIT	K1_F	NA	4	15.09	107	214.81	-67.54
3	HESIT JE	NA	K1_F	10	37.71	133	267.00	-64.71
4	UN PEU	K1_F	K1_C	6	22.63	114	228.86	-64.61
5	ON A	K1_F	K1_F	5	18.86	96	192.72	-54.61
6	C'EST	K1_F	K1_F	214	807.09	675	1355.10	-53.88
7	HESIT C'	NA	K1_F	1	3.77	65	130.49	-48.60
8	HESIT J'	NA	K1_F	3	11.31	77	154.58	-48.04
9	NP HESIT	NA	NA	13	49.03	118	236.89	-45.54
10	QU ON	K1_F	K1_F	4	15.09	76	152.57	-43.06
11	HESIT ON	NA	K1_F	1	3.77	45	90.34	-31.82
12	C'ÉTAIT	K1_F	K1_F	21	79.20	120	240.91	-29.92
13	PRENDRE LA	K1_C	K1_F	1	3.77	42	84.32	-29.33
14	JE SAIS	K1_F	K1_C	3	11.31	52	104.39	-28.45
15	QUE HESIT	K1_F	NA	1	3.77	38	76.29	-26.03
16	QUE ÇA	K1_F	K1_F	1	3.77	38	76.29	-26.03
17	POUR ALLER	K1_F	K1_C	25	94.29	121	242.91	-24.09
18	HESIT À	NA	K1_F	1	3.77	35	70.26	-23.56
19	EN FAIT	K1_F	K1_C	3	11.31	45	90.34	-23.16
20	ET QUE	K1_F	K1_F	2	7.54	39	78.29	-22.30
21	HEURE DIX	K1_C	K1_C	2	7.54	38	76.29	-21.52
22	HESIT IL	NA	K1_F	3	11.31	42	84.32	-20.93
23	PEU PRÈS	K1_C	K1_F	1	3.77	31	62.23	-20.30

#### 8.1.2.2. 特徴的 N-grams と特徴語の関係

特徴語として抽出された語が特徴的 N-grams に含まれているかどうかについて分析を行ったところ、表 295 のようになった。なお、特徴語は太字で、灰色で色付けされている N-grams が特徴語を含む N-grams である。

この表より、特徴的 N-grams として抽出された語は、特に特徴度が高い N-grams については主に特徴語を含む N-grams であることが分かる。したがって、学習者は WUs レベルで過剰使用しているだけでなく、それらの語の前後にも決まったパターンを使用していることが分かる。

一方、過少使用語については、特徴が含まれる N-grams もあったが、そうでない N-grams も多く見られた。つまり、過少使用された N-grams は語彙レベルが問題となっている場合と、語彙レベルでは過少使用されていないが N-grams レベルで過少使用されるものが混合されているといえる。

表 295 特徴的 N-gram と特徴語の関係 (過剰使用語)

順位	特徴的 MWUs			
	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
1	ELLE S'APPELLE NP NP	ELLE S'APPELLE NP	J'AIME LE	JE VOUDRAIS
2	S'APPELLE NP NP ELLE	JE NE VEUX PAS	JE SUIS JAPONAISE	JE NE
3	APPELLE NP NP ELLE EST	APPELLE NP NP ELLE	ELLE S'APPELLE	AIME LE
4	RUE DE LA NP NP	IL Y A BEAUCOUP	JE VOUDRAIS ALLER	SUIS JAPONAISE
5	IL Y A BEAUCOUP DE	QU EST-CE QUE	ELLE HABITE À	ELLE S'
6	JE SUIS ALLÉE À NP	S'APPELLE NP NP	HABITE À NP	DE NP
7	À QUATORZE HEURES QUARANTE ET	Y A BEAUCOUP DE	NE VEUX PAS	VOUDRAIS ALLER
8	QU EST-CE QUE TU	JE NE PEUX PAS	JE NE VEUX	HABITE À
9	NP NP JE SUIS JAPONAISE	SUIS ALLÉE À NP	Y A BEAUCOUP	NP ELLE
10	JAPONAIS ELLE EST MARIÉE ET	JE SUIS ALLÉE À	NP ELLE EST	PRENEZ LE
11	ANS ELLE EST PROFESSEUR DE	QUATORZE HEURES QUARANTE ET	PRENEZ LE NP	VOUS PRENEZ
12	NP ELLE EST JAPONAISE ELLE	ELLE HABITE À NP	JE NE PEUX	LE FRANÇAIS
13	ELLE EST JAPONAISE ELLE A	DE LA NP NP	S'APPELLE NP	D'ABORD
14	NP NP ELLE EST JAPONAISE	VOUS PRENEZ LE NP	NE PEUX PAS	NE VEUX
15	QU EST-CE QUE C'	NP NP ELLE EST	A BEAUCOUP DE	ELLE A
16	ET ELLE A DEUX ENFANTS	NP JE SUIS JAPONAISE	HEURES QUARANTE ET	PROFESSEUR DE
17	MARIÉE ET ELLE A DEUX	ELLE EST PROFESSEUR DE	VOUS PRENEZ LE	JAPONAIS ELLE
18	PROFESSEUR DE JAPONAIS ELLE EST	NP ELLE EST JAPONAISE	ALLÉE À NP	NE PEUX
19	UNE HEURE ET DIX MINUTES	JAPONAIS ELLE EST MARIÉE	SUIS ALLÉE À	NP LIGNE
20	JE SUIS JAPONAISE J'AI	UNE HEURE ET DIX	QU EST CE	ELLE EST
21	JE N' AIME PAS LE	PROFESSEUR DE JAPONAIS ELLE	D'ABORD VOUS	LE FILM
22	ELLE EST PROFESSEUR DE JAPONAIS	JE VIENS DE NP	NP NP ELLE	VEUX PAS
23	NP JE SUIS JAPONAISE J'	NP ELLE HABITE À	LA NP NP	ET VOUS
24	A TRENTE HUIT ANS ELLE	ELLE EST JAPONAISE ELLE	PROFESSEUR DE JAPONAIS	APPELLE NP
25	EST MARIÉE ET ELLE A	MARIÉE ET ELLE A	APPELLE NP NP	MANGE DU
26	PARCE QUE IL Y A	ET ELLE A DEUX	À LA GARE	A BEAUCOUP
27	JE VOUDRAIS RESTER À LA	JE SUIS JAPONAISE J'	JAPONAIS ELLE EST	MAIS JE
28	EST PROFESSEUR DE JAPONAIS ELLE	JE VOUDRAIS ALLER AU	MAIS JE NE	JE VEUX
29	DE JAPONAIS ELLE EST MARIÉE	HEURE ET DIX MINUTES	ET ELLE A	QUARANTE ET

順位	特徴的 MWUs			
	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
30	<b>VOUDRAIS RESTER À LA MAISON</b>	<b>SUIS JAPONAISE J'AI</b>	<b>NP DE NP</b>	ALLÉE À
31	EST-CE QUE C'EST	<b>D'ABORD VOUS PRENEZ</b>	EST PROFESSEUR DE	QU EST
32	<b>ELLE EST MARIÉE ET ELLE</b>	<b>DE JAPONAIS ELLE EST</b>	<b>JE MANGE DU</b>	<b>ABORD VOUS</b>
33	DIX SEPT HEURES TRENTE NEUF	<b>ELLE A DEUX ENFANTS</b>	<b>ELLE EST JAPONAISE</b>	<b>JE SOUVENT</b>
34	À LA GARE DE NP	<b>EST MARIÉE ET ELLE</b>	<b>JE VAIS À</b>	S'APPELLE
35	SUIS ÉTUDIANTE J'HABITE À	N' AIME PAS LE	J'AIME LES	<b>ANS ELLE</b>
36	QUARANTE ET ARRIVE À NP	EST PROFESSEUR DE JAPONAIS	ÇA FAIT COMBIEN	<b>TU PEUX</b>
37	<b>JE VAIS À L'UNIVERSITÉ</b>	<b>JE VOUDRAIS ALLER À</b>	<b>DE JAPONAIS ELLE</b>	AVEC TOI
38	QUE IL Y A BEAUCOUP	<b>JE VOUDRAIS RESTER À</b>	<b>JE VIENS DE</b>	ET ARRIVE
39	HEURES QUARANTE ET ARRIVE À	J'Étudie LE FRANÇAIS	HEURE ET DIX	<b>JE VIENS</b>
40	QUATORZE HEURES QUARANTE ET ARRIVE	<b>EST JAPONAISE ELLE A</b>	NP POUR NP	AIME LES
41	ET ARRIVE À NP À	HEURES QUARANTE ET ARRIVE	<b>JE SUIS ALLÉE</b>	LA GARE
42	<b>TRENTE HUIT ANS ELLE EST</b>	PARCE QUE IL Y	<b>MARIÉE ET ELLE</b>	VIENS DE
43	C'EST UNE BONNE IDÉE	<b>VOUDRAIS RESTER À LA</b>	VIENS DE NP	FAIT COMBIEN
44	<b>J'AI DIX NEUF ANS</b>	<b>JE NE COMPRENDS PAS</b>	<b>NP JE VOUDRAIS</b>	<b>NON JE</b>
45	<b>EST JAPONAISE ELLE A TRENTE</b>	SEPT HEURES TRENTE NEUF	<b>JE SUIS JAPONAIS</b>	<b>L'UNIVERSITÉ</b>
46	NP EST SOIXANTE SEPT RUE	À LA GARE DE	<b>JE VOUDRAIS RESTER</b>	SUIS ALLÉE
47	<b>D'ABORD VOUS PRENEZ LE</b>	<b>JE MANGE DU PAIN</b>	PARCE QUE IL	<b>LE NP</b>
48	<b>MAIS JE NE VEUX PAS</b>	QUE IL Y A	<b>ELLE A DEUX</b>	SUIS JAPONAIS
49	<b>ELLE A TRENTE HUIT ANS</b>	<b>VAIS À L'UNIVERSITÉ</b>	<b>JAPONAISE J'AI</b>	<b>VOUDRAIS RESTER</b>
50		<b>PARCE QUE JE PEUX</b>	<b>SUIS JAPONAISE J'</b>	<b>EST JAPONAISE</b>
51		<b>JE VAIS À L'</b>	<b>VOUDRAIS ALLER AU</b>	<b>PUIS VOUS</b>
52		<b>TRENTE HUIT ANS ELLE</b>	<b>ABORD VOUS PRENEZ</b>	<b>EST NP</b>
53		<b>HUIT ANS ELLE EST</b>	Étudie LE FRANÇAIS	<b>LE FOOTBALL</b>
54		QUARANTE ET ARRIVE À	<b>ANS ELLE EST</b>	ES LIBRE
55		EST-CE QUE C'	DIX NEUF ANS	<b>JAPONAISE ELLE</b>
56		DIX SEPT HEURES TRENTE	EXCUSEZ MOI MAIS	<b>ELLE HABITE</b>
57		<b>ET ARRIVE À NP</b>	ET ARRIVE À	LE LIGNE
58		<b>JE SUIS DÉSOLÉE MAIS</b>	<b>EST JAPONAISE ELLE</b>	BEAUCOUP DE
59		J'AI MANGÉ DU	TU ES LIBRE	DU RIZ
60		À QUATORZE HEURES QUARANTE	<b>VOUDRAIS ALLER À</b>	EST TRÈS

順位	特徴的 MWUs			
	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
61		ELLE A TRENTE HUIT	MANGE DU PAIN	DU PAIN
62		J'HABITE À NP	VOUDRAIS RESTER À	VAIS À
63		BONJOUR C'EST NP	MOI MAIS JE	EN VÉLO
64		EST UNE BONNE IDÉE	DE NP ET	AIME L'
65		C'EST UNE BONNE	J'AIME L'	QUE IL
66		LE NP DE NP	QUARANTE ET ARRIVE	SEPT HEURES
67		DE NP POUR NP	HUIT ANS ELLE	NEUF ANS
68		J'AI DIX NEUF	DE LA SALADE	NE COMPRENDS
69		AI DIX NEUF ANS	JE NE COMPRENDS	COMPRENDS PAS
70		ABORD VOUS PRENEZ LE	LE LIGNE NP	LA SALADE
71		NP EST SOIXANTE SEPT	NE COMPRENDS PAS	LE THÉ
72		SPECTACLES ÇA FAIT COMBIEN	ELLE A TRENTE	NP AVEC
73		JE SUIS DE NP	WEEKEND JE VOUDRAIS	TRENTE NEUF
74		FAUT UNE HEURE ET	À NP AVEC	ARRIVEZ À
75		MOI MAIS JE NE	ET DIX MINUTES	MOI MAIS
76		MAIS JE NE VEUX	JAPONAISE ELLE A	ET DIX
77		L'UNIVERSITÉ EN VÉLO	DIX SEPT HEURES	ÉTUDIE LE
78		À NP IL FAIT	HEURES TRENTE NEUF	POUR NP
79		NP ET J'AI	NP IL FAIT	ET ELLE
80		LA GARE DE NP	MAIS JE VOUDRAIS	VOUDRAIS SORTIR
81		J'AIME LE PRINTEMPS	ET DEUX ENFANTS	ET DEUX
82		CE QUE C'EST	ET PUIS VOUS	BONNE IDÉE
83		JE SUIS ÉTUDIANTE J'	JE VOUDRAIS SORTIR	HEURES QUARANTE
84			VAIS À L'	ALLER AU
85			QUE IL Y	NP NP
86			SEPT HEURES TRENTE	JAPONAISE J'
87			NP À DIX	DIX SEPT
88			PARCE QUE JE	BOIS DU
89			SUIS DÉSOLÉE MAIS	À NP
90			AI MANGÉ DU	DIX MINUTES
91			NON JE N'	AI MANGÉ

順位	特徴的 MWUs			
	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
92			<b>JE VOUDRAIS FAIRE</b>	BONJOUR C'
93			UNE BONNE IDÉE	<b>VOUDRAIS FAIRE</b>
94			SON ADRESSE DE	GARE DE
95			EST UNE BONNE	AVEC PLAISIR
96			BONJOUR C'EST	EST COMBIEN
97			<b>GARE DE NP</b>	DU THÉ
98			C'EST COMBIEN	<b>JE VOIS</b>
99			<b>À L'UNIVERSITÉ</b>	<b>JE N'</b>
100			AI DIX NEUF	AIME LA
101			J'AI MANGÉ	ÇA FAIT
102			<b>PRENDS LE NP</b>	OÙ EST
103			<b>JE BOIS DU</b>	<b>NP DE</b>
104			<b>UNIVERSITÉ EN VÉLO</b>	MANGÉ DU
105			<b>SUIS DE NP</b>	LIGNE À
106			<b>NP EST SOIXANTE</b>	RIZ ET
107			SPECTACLES ÇA FAIT	SPECTACLES ÇA
108			<b>NP ET VOUS</b>	LE CAFÉ
109			AVEC TOI CE	TOI CE
110			<b>ELLE EST UNE</b>	AU MUSÉE
111			AIME LE PRINTEMPS	<b>ET JE</b>
112			LA GARE DE	ME SENS
113			N' EST PAS	LIBRE CE
114			J'AIME LA	AVEC MOI
115			SUIS ÉTUDIANTE J'	À DIX
116				PAS ALLER

表 296 特徴的 N-gram と特徴語の関係 (過少使用語)

順位	特徴的 MWUs			
	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
1		PRENDRE LA LIGNE NP	J'AIME BIEN	AIME BIEN
2			J'AI PAS	<b>ET HESIT</b>
3			<b>HESIT C'EST</b>	<b>HESIT JE</b>
4			PRENDRE LA LIGNE	UN PEU
5			<b>HESIT JE SUIS</b>	<b>ON A</b>
6			AIME PAS FAIRE	<b>C'EST</b>
7			POUR ALLER DE	<b>HESIT C'</b>
8			<b>PARCE QUE HESIT</b>	<b>HESIT J'</b>
9			UNE HEURE DIX	<b>NP HESIT</b>
10			<b>C'EST PAS</b>	<b>QU ON</b>
11				<b>HESIT ON</b>
12				<b>C'ÉTAIT</b>
13				PRENDRE LA
14				JE SAIS
15				<b>QUE HESIT</b>
16				QUE ÇA
17				POUR ALLER
18				<b>HESIT À</b>
19				EN FAIT
20				ET QUE
21				HEURE DIX
22				<b>HESIT IL</b>
23				PEU PRÈS

### 8.1.3. 特徴的 N-grams の質的分析

#### 8.1.3.1. 特徴的 N-grams の傾向

特徴的 N-grams の質的分析としてまず、特徴的 N-grams の傾向を調べる。2 語連鎖から 5 語連鎖までで特徴的な N-grams として抽出されたものについて、重なりがあった場合は一緒にまとめ、表にした。

過剰使用 N-grams について傾向として、過剰使用 N-grams の多くは場面と機能に分けることができた(以下の図を参照)。特に自己紹介・第三者の紹介、好き嫌い、時間・道のり、質問する、断るという場面や機能については、多くの過剰使用 N-grams が見られた。

複数の特徴的な N-grams が組み合わさって、学習者に特徴的な文や節を構成しているものが多いことが分かる。表の最も色が濃い背景色は文や節を構成する 2 語連鎖を示し、色が薄い背景色は 3 語連鎖を示している。





j025,j026														
ELLE	EST	UNE												
j025														
NP	ELLE	HABITE	À											
	ELLE	HABITE	À											
	ELLE	HABITE	À	NP										
			À	NP										
j025														
SON	ADRESSE	DE		NP										
		DE		NP	EST	SOIXANTE								
				NP	EST	SOIXANTE	SEPT							
				NP	EST	SOIXANTE	SEPT	RUE						
								RUE	DE	LA	NP	NP		
									DE	LA	NP	NP		
										LA	NP	NP		
j025														
ET	DEUX													
ET	DEUX	ENFANTS												
j024														
NP	NP	JE	SUIS	JAPONAISE										
	NP	JE	SUIS	JAPONAISE										
	NP	JE	SUIS	JAPONAISE	J'									
		JE	SUIS	JAPONAISE	J'									
		JE	SUIS	JAPONAISE	J'					AI				
			SUIS	JAPONAISE	J'					AI				
				JAPONAISE	J'					AI				
					J'					AI	DIX	NEUF		
					J'					AI	DIX	NEUF	ANS	
										AI	DIX	NEUF	ANS	
											DIX	NEUF	ANS	
												NEUF	ANS	
j024														
JE	SUIS	JAPONAIS												
	SUIS	JAPONAIS												
j024														
JE	SUIS	DE	NP											
	SUIS	DE	NP											
		DE	NP											
j024														
JE	SUIS	ÉTUDIANTE	J'											
	SUIS	ÉTUDIANTE	J'											
	SUIS	ÉTUDIANTE	J'	HABITE	À									
			J'	HABITE	À					NP				
				HABITE	À					NP				
					À					NP				
j015														
JE	VIENS													
JE	VIENS	DE												
JE	VIENS	DE	NP											
	VIENS	DE	NP											
		DE	NP											
j016														
j024	ÉTUDIE	LE	FRANÇAIS											
J'	ÉTUDIE	LE	FRANÇAIS											

図 24 好き・嫌い

j029, j042, j046						
NON	JE					
NON	JE	N'				
	JE	N'				
	JE	N'	AIME		PAS	LE
		N'	AIME		PAS	LE
j033						
JE	N'					
JE	N'	AIME	PAS		LE	
	N'	AIME	PAS		LE	
					LE	FOOTBALL
j001, j010, j011, j028, j043, j044						
J'	AIME	LE				
J'	AIME	LE	PRINTEMPS			
	AIME	LE	PRINTEMPS			
J'	AIME	LES				
	AIME	LES				
J'	AIME	L'				
	AIME	L'				
J'	AIME	LA				
	AIME	LA				

図 25 習慣

j010, j011, j036, j038			
JE		SOUVENT	
j010			
JE	VOIS		
		LE	FILM

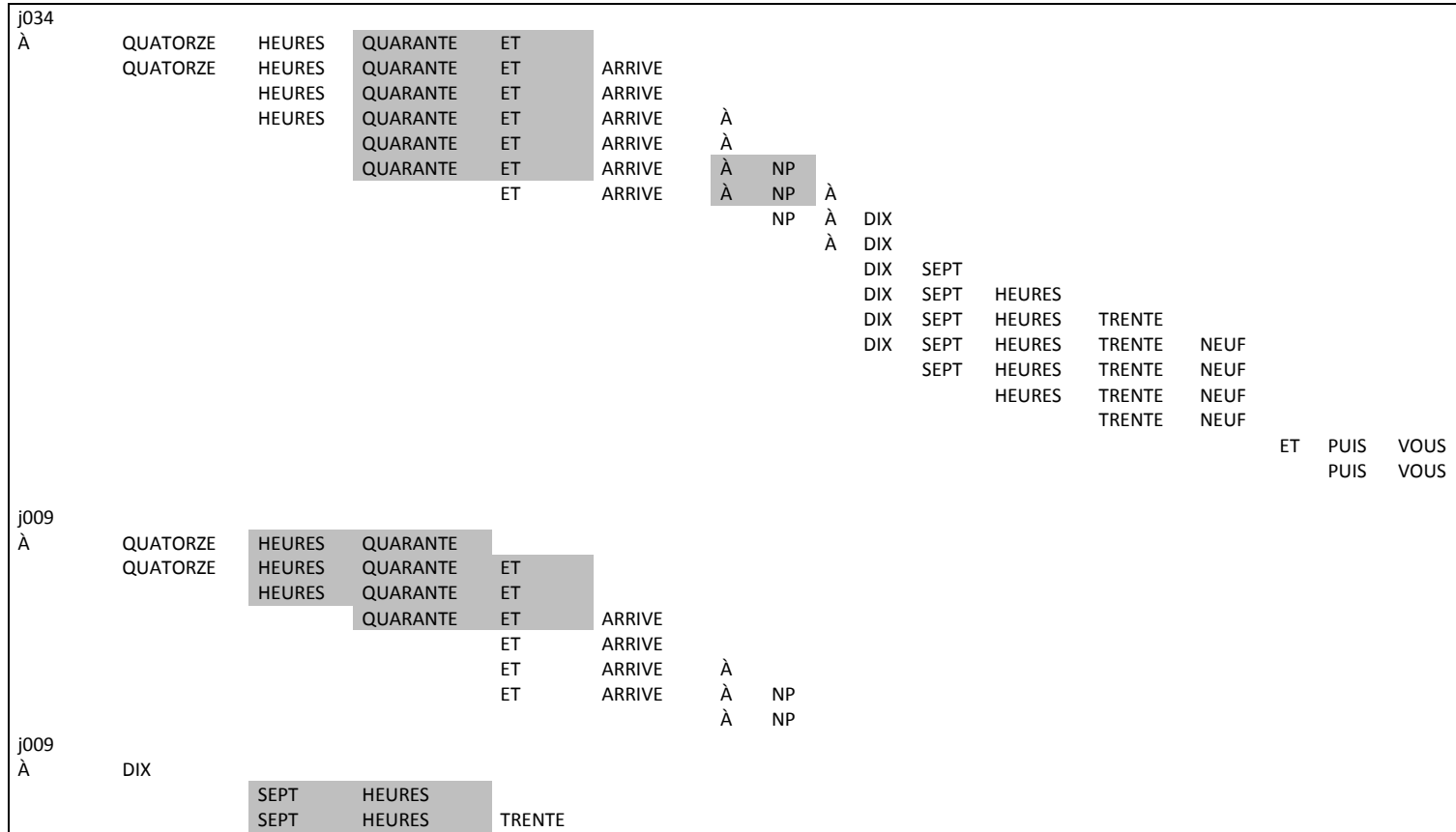
図 26 理由

j001, j028						
PARCE	QUE	IL				
PARCE	QUE	IL	Y			
PARCE	QUE	IL	Y		A	
	QUE	IL	Y		A	
	QUE	IL	Y		A	BEAUCOUP
		IL	Y		A	BEAUCOUP
		IL	Y		A	BEAUCOUP
			Y		A	BEAUCOUP
					A	BEAUCOUP
						BEAUCOUP
						DE
						DE
						DE
j001, j043						
PARCE	QUE	JE				
PARCE	QUE	JE	PEUX			

図 27 電話する

j032				
BONJOUR	C'			
BONJOUR	C'		EST	
BONJOUR	C'		EST	NP
			EST	NP

図 28 時間・道のり



j009	FAUT	UNE	HEURE	ET					
		UNE	HEURE	ET	DIX				
		UNE	HEURE	ET	DIX	MINUTES			
			HEURE	ET	DIX	MINUTES			
				ET	DIX	MINUTES			
					DIX	MINUTES			
j009	À	NP							
	À	NP	IL	FAIT					
		NP	IL	FAIT					
j008,									
j009	À	LA	GARE						
	À	LA	GARE	DE					
	À	LA	GARE	DE	NP				
		LA	GARE	DE	NP				
			GARE	DE	NP				
				DE	NP	ET			
j008,									
j035	PRENDS	LE	NP						
		LE	NP						
j009	LE	LIGNE							
	LE	LIGNE	NP						
j035	ABORD								
D'	ABORD	VOUS							
D'	ABORD	VOUS	PRENEZ						
D'	ABORD	VOUS	PRENEZ	LE					
D'	ABORD	VOUS	PRENEZ	LE					
		VOUS	PRENEZ	LE					
		VOUS	PRENEZ	LE	NP				
			PRENEZ	LE	NP				
				LE	NP	DE	NP		
					NP	DE	NP		
						DE	NP		
j034,									
j035	DE	NP							
	DE	NP	POUR	NP					
		NP	POUR	NP					

図 29 行動

j030, j031		SUIS	ALLÉE						
JE		SUIS	ALLÉE	A					
JE		SUIS	ALLÉE	À					NP
JE		SUIS	ALLÉE	À					NP
			ALLÉE	À					NP
				À					NP
j008		VAIS	À						
JE		VAIS	À	L'					UNIVERSITÉ
JE		VAIS	À	L'					UNIVERSITÉ
JE		VAIS	À	L'					UNIVERSITÉ
			À	L'					UNIVERSITÉ
				L'					UNIVERSITÉ
				L'					UNIVERSITÉ
							EN		VÉLO
							EN		VÉLO
							EN		VÉLO

図 30 質問する

j006, j032		TU	ES	LIBRE			
			ES	LIBRE			
				LIBRE			CE
j006, j032		AVEC	TOI				
		AVEC	TOI	CE			
			TOI	CE			
j006, j032		AVEC	MOI				

図 31 質問する

j023		QU'	EST				
		QU'	EST	CE			
		QU'	EST	CE	QUE		
		QU'	EST	CE	QUE	TU	
j048		SPECTACLES	ÇA	FAIT			
		SPECTACLES	ÇA	FAIT	COMBIEN		
			ÇA	FAIT	COMBIEN		
				FAIT	COMBIEN		
j048		C'	EST	COMBIEN			
			EST	COMBIEN			
j017		OÙ	EST				

図 32 誘う

j006, j032				
TU	ES	LIBRE		
	ES	LIBRE		
		LIBRE		CE
j006, j032				
AVEC	TOI			
AVEC	TOI	CE		
	TOI	CE		
j006, j032				
AVEC	MOI			

図 33 依頼する・受け入れる

j003, j004	
TU	PEUX
j004, j007	
AVEC	PLAISIR

図 34 同意する

j007				
C'	EST	UNE	BONNE	
C'	EST	UNE	BONNE	IDÉE
	EST	UNE	BONNE	IDÉE
		UNE	BONNE	IDÉE
			BONNE	IDÉE

図 35 不理解を伝える

j015, j042				
JE	NE	COMPRENDS		
JE	NE	COMPRENDS		PAS
	NE	COMPRENDS		PAS
		COMPRENDS		PAS

図 36 断る

EXCUSEZ	MOI	MAIS						
	MOI	MAIS	JE					
	MOI	MAIS	JE	NE				
		MAIS	JE	NE				
		MAIS	JE	NE	VEUX			
		MAIS	JE	NE	VEUX		PAS	
			JE	NE	VEUX		PAS	
				NE	VEUX		PAS	
					VEUX		PAS	
							PAS	ALLER

EXCUSEZ	MOI MOI MOI	MAIS MAIS MAIS MAIS	JE JE JE JE JE	NE NE NE NE NE	PEUX PEUX PEUX	PAS PAS PAS	ALLER
j005, j033	JE SUIS SUIS	DÉSOLÉE DÉSOLÉE	MAIS MAIS MAIS MAIS MAIS MAIS MAIS MAIS	JE JE JE JE JE JE JE JE	NE NE NE NE NE NE NE NE	VEUX VEUX VEUX VEUX PEUX PEUX	PAS PAS PAS PAS PAS PAS

図 37 飲食

j037	JE JE JE	MANGE MANGE MANGE	DU DU DU DU	PAIN PAIN PAIN RIZ RIZ	ET
j012	J' J'	AI AI AI	MANGÉ MANGÉ MANGÉ	DU DU DE DU RIZ RIZ	LA LA SALADE SALADE ET
j013, j038	JE	BOIS BOIS	DU DU DU LE LE	THÉ THÉ CAFÉ	



以上の図を見ると、これらの N-grams が必ずしもエラーであるとは限らないことが分かる。

次に、過少使用 N-grams をまとめたところ、ある特定の場面や機能に分けることが出来る過少使用 N-grams もあったが、過剰使用 N-grams ほど数は多くなかった。

図 38 時間・道のり

PRENDRE	LA		
PRENDRE	LA	LIGNE	
PRENDRE	LA	LIGNE	NP
POUR	ALLER		
POUR	ALLER	DE	
UNE	HEURE	DIX	
	HEURE	DIX	

図 39 好き嫌い

J'	AIME	BIEN		
	AIME	BIEN		
	AIME	PAS	FAIRE	

過少使用 N-grams は場面や機能で分類するよりも、語彙・統語・談話レベルで分類できた。

図 40 過少使用語による影響

<b>HESIT</b>				
PARCE	QUE	HESIT		
	QUE	HESIT		
	ET	HESIT		
	NP	HESIT		
		HESIT	C'	
		HESIT	C'	EST
			C'	EST
		HESIT	JE	
		HESIT	JE	SUIS
		HESIT	J'	
		HESIT	ON	
		HESIT	IL	
<b>C'</b>				
C'		EST		
C'		ÉTAIT		
<b>PEU</b>				
PEU		PRÈS		
UN		PEU		

図 41 否定辞 *ne* の脱落

C'	EST	
C'	EST	PAS
J'	AI	PAS

図 42 *qu(e)*

ET	QUE		
	QUE	ÇA	
	QU'	ON	
		ON	A

図 43 談話的機能を持つ過少使用 N-grams

JE	SAIS
EN	FAIT

次に、これまでの分析では見られなかった側面について、特徴的 N-grams の要因の一部についてより質的な分析を行う。特にタスクに基づくコーパスで見られた特徴的 N-grams の中から、WU レベルに起因すると考えられる特徴的 N-grams と MWUs レベルが影響を与えたと思われる特徴的 N-grams について分析を行う。

### 8.1.3.2. WUs レベルが影響を与えた特徴的 N-grams

WUs レベルが特徴的 N-grams の一因となったと考えられるものとして、学習者と母語話者との間に見られる人称代名詞の使用方法、そして語彙選択の違いの二点が挙げられた。

#### 8.1.3.2.1. 人称代名詞

人称代名詞の一部の過剰使用や過少使用は、特徴語分析、コロケーション分析でも指摘してきたため、これまでの分析では見られなかったより広い範囲における特徴的使用について扱う。

図 23 で示した通り、*elle* を含む N-grams の連鎖は以下のタスクで多く見られた。

j025	<p>以下の身分証明書をよく見てください。</p> <p>Nom: Tanaka          Prénom: Kazuko          Nationalité: Japonaise          Âge: 38ans          Profession: Professeur          Situation de famille: Mariée – 2 enfants          Adresse au Japon: 3 Asahi-cho Fuchu          Adresse en France: 67, rue de la Liberté 69000 Lyon          この人物を紹介してください。</p>
------	--

このタスクに対しての学習者による答えが、以下の連鎖になっている。なお、鍵カッコ内の数字は、学習者 39 名中使用した人数を示している：

elle s'appelle NP NP [32] elle est japonaise [31] elle a trente huit ans [35] elle est professeur de japonais [28] elle est mariée [22] (et) elle a deux enfants [33]

この MWUs に特徴的な点は、6 つの文がつながっており、主語と動詞が繰り返されて使用されている点である。また、カッコ内の数字からも分かる通り、学習者によるこれらの N-grams の使用レンジは高く、回答パターンが一様であったことが分かる。一方、母語話者による同タスクのモデル回答は以下の通りであった：

voici NP NP [8] elle est japonaise [9] elle a trente huit ans [22] elle est professeur [28] elle est mariée [26] (et) a deux enfants [17]

母語話者は同じタスクでも *elle* の使用回数が 2 回少なかった。また、鍵カッコ内の数字からも分かるように、以上の N-grams は最も使用された N-grams であるが、母語話者全体に占める使用割合は低いことが分かる。

また、学習者が人称代名詞を使うのに対して、母語話者は非人称の名詞を使用した例として、以下のタスクを挙げることができる。

j030	あなたが最近行った旅行について話してください。
f030	Racontez votre dernier voyage.

学習者と母語話者によって主に使用された N-grams をまとめたところ、以下の通りとなった。

表 297 タスク 030 で使用された N-grams

	学習者	母語話者
<i>je suis allé(e)</i> + 前置詞 + 場所	29	3
<i>mon dernier voyage c'était</i> + 前置詞 + 場所	0	16
<i>mon dernier voyage était</i> + 前置詞 + 場所	0	5
計	29	24

学習者は *je suis allé(e)* という N-gram を多く使用しているのに対して、母語話者は *mon dernier voyage* という N-grams を多く使用していることが分かる。つまり、学習者は「～へ行った」という行為に焦点を当てているのに対して、母語話者は「最後にした旅行は～だ」というコピュラ文を用いて、状況を説明している。この点は学習者が複合過去形を用いているのと、母語話者が半過去形を用いていることから分かる。

同様の現象として、*j'aime l(e)* の過剰使用 MWUs が説明できる。この過剰使用 MWUs は主に以下のタスクで見られた。

j001	あなたの好きな季節は何ですか？
f001	Quelle est votre saison préférée? Pourquoi?

学習者と母語話者がこのタスクで使用していた MWUs をまとめたところ、以下の通りとなった。

表 298 タスク 001 で使用された N-grams

	学習者	母語話者
<i>j'aime l'(e) + 季節</i>	28	2
<i>ma saison préférée (c') est + 季節</i>	0	23

以上の表から、学習者は「～が好きだ」という文が多く、母語話者は「好きな季節は～だ」というコピュラ文を使用する傾向にある。このタスクの場合、母語話者のタスクの指示文に *votre saison préférée* という句が見られるため、その答えとして *ma saison préférée* と回答した母語話者が多かった可能性が指摘できる。しかし、学習者に対する指示文も「あなたの好きな季節」という句が含まれており、「私の好きな季節は～だ」というコピュラ文で回答することも可能であったはずである。このような例は 1 例のみで、\**ma saison favori* という回答で始まる例であり、形容詞の性の一致がなされておらず、これは文法的に正しい使用ではなかった (JT21)。

このように、母語話者は人称代名詞を使用せず、非人称主語をとるコピュラ文を使用する傾向にある。このようなスタイルの違いが、学習者による人称代名詞の過剰使用の一因になったと考えられる。

### 8.1.3.2.2. 語彙の選択

語彙の選択として、動詞 VOULOIR と POUVOIR の選択、動詞 VOULOIR と動詞 AIMER の選択について分析を行う。

*mais je ne veux pas* という特徴的 MWU は、過剰使用 N-grams として抽出されたものの一つである。また、接続詞 *mais* を除いた *je ne veux pas* という 4 語連鎖も、特徴的 MWU として抽出されている。この MWU は学習者コーパスの以下のタスクから抽出されたものである：

j005	あなたの友達が、携帯を貸してくれるようあなたに頼んできましたが、あなたは携帯を貸したくありません。何と言って断りますか？
j027	週末にあなたがしたいこと、そうでないことについて話してください。
j033	あなたは、友人からサッカーの試合を見に行こうと誘われました。あなたはサッカーが好きではないので誘いを断りたいと思っています。理由を添えて、誘いを断ってください。

j005 と j033 は依頼や誘いを断るタスクであり、j027 は願望について答えるタスクである。過剰使用語として否定辞の *ne* が挙げられたため、否定辞なしの MWU についても分析したが、*mais je veux pas* という MWUs は一例も見られなかった。

母語話者コーパスにおいて、まず依頼や誘いを断るタスクで使用された MWUs を分析したところ、以下のようになった：

表 299 誘いを断るタスクに使用された MWUs

	日本人学習者		フランス語母語話者	
	MWUs	生起回数	MWUs	生起回数
j005	je ne peux pas	14	je (ne) peux pas	10
	je ne veux pas	3	j(e n')aime pas (trop)	6
	je ne pourrais pas	1	j(e n')ai pas envie	3
	tu ne peux pas	1	je ne vais pas pouvoir	1
j033	je ne peux pas	15	je peux pas	13
	je ne veux pas	4	je vais pas pouvoir	5
			j'ai pas envie	4
			c'est pas possible	2
			ça va pas être possible	1
			ça sera pas possible	1
			ça me dit pas	1

学習者も母語話者も依頼や誘いを断る際に使用する動詞は、動詞 **POUVOIR** の方を多く使用していることが分かる。しかし、学習者は動詞 **VOULOIR** を選択した例が見られるが、母語話者には一例も見られなかった。

### 8.1.3.3. MWUs レベルが影響を与えた特徴的 N-grams

MWUs レベルで過剰使用 N-grams となった例として、*beaucoup de* という MWU と疑問文を取り上げる。さらに、過少使用 N-grams で見られた特徴であった、否定辞 *ne* の脱落、*qu(e)*を含む N-grams、そしてディスコースマーカ―について分析を行う。また、エラーによって引き起こされた過剰使用、過少使用についてもここで取り上げる。

#### 8.1.3.3.1. *beaucoup de*

*beaucoup de* という MWU は 2 語連鎖から 5 語連鎖までの全ての過剰使用 N-grams に含まれていた(図 295)。この MWU を含む過剰使用 N-grams で最も長いものは、*parce que il y a beaucoup de* という MWU であった。まず、従属接続詞 *parce que* の *que* がエリジオンされていない点がこの MWU の過剰使用の理由として指摘できる。そこで、エリジオンされた形である *parce qu'il y a beaucoup de* という MWUs がフランス語母語話者コーパスにあるかどうか調べたところ、1 例のみしか存在しなかった。また、この MWUs には *parce que* という MWUs と *il y a* という MWUs、そして *beaucoup de* という MWUs の 3 つの意味を持つ MWUs から構成されているため、別々に頻度を比較した。さらに、この 3 つの MWUs の範囲を拡大して *parce qu(e)il y a* の MWUs、*il y a beaucoup de* の MWUs につ

いても比較を行った。なお、*il y a* には期間を表す語句を伴い「～(期間)前」という意味も表すため、その例は数に含めず、「～がある」という意味の生起頻度のみをカウントした。

表 300 *parce que il y a beaucoup de* を構成する MWUs の頻度

MWUs	学習者 (生起回数)	母語話者 (生起回数)	keyness
<i>parce qu(e)</i>	957.95 (254)	650.46 (324)	17.91
<i>il y a</i>	345.42 (85)	288.50 (138)	0.70
<i>beaucoup d'(e)</i>	403.55 (107)	168.63 (84)	33.61
<i>parce qu(e) il y a</i>	172.20 (36)	40.94 (18)	21.73
<i>il y a beaucoup de</i>	215.25 (45)	22.74(10)	50.01
<i>parce que il y a beaucoup de</i>	86.18 (13)	0 (0)	NA
<i>parce qu'il y a beaucoup de</i>	66.30 (10)	6.63 (1)	14.93

その結果、意味を持つ最も小さな 3 つの MWUs において、学習者による過剰使用が見られたのは *beaucoup d'(e)* の MWUs のみであった。また、*il y a beaucoup de* と MWUs の範囲を広げてみると、過剰使用がより際立った。

また、*parce qu(e)* も *il y a* もそれぞれこの MWU で使用される時は *Keyness* の値は高くなかったが、*parce qu(e)il y a* と MWU の長さを長くすることで、*Keyness* の値が高くなった。このことより、過剰使用の見られない MWUs でも複数組み合わせられることによって、外国語らしさを引き起こす場合があるということが示唆された。

### 8.1.3.3.2. 疑問文

疑問詞 *qu'est-ce que* に関連する MWUs として、*qu'est-ce que c'est* と *qu'est-ce que tu* という MWUs が抽出された。前者はタスク番号 j018、後者はタスク番号 j023 で主に使用された：

j018	旅先で入ったレストランのメニューの中に、聞きなれない料理がありました。あなたはお店の人に何と尋ねますか？
j023	あなたは今、フランスでカフェに來ています。 友達の Fabien が何のドリンクを頼むのか、あなたは知りません。 友達に何のドリンクを注文するのかを聞く時、何と言いますか？

学習者と母語話者がそれぞれのタスクで使用した疑問形を含む MWUs を比較したところ、以下ようになった：

表 301 タスク 018 と 023 に見られる MWUs

タスク番号	日本人学習者		フランス語母語話者	
	MWUs	生起回数	MWUs	生起回数
018	qu'est-ce que c'est	24	c'est quoi	8
	c'est comment	3	qu'est-ce que c'est	5
	quel est ce plat	2	qu'est-ce qu'il y a dedans	3
			quel est ce plat	2
			de quoi il est composé	2
			il est composé de quoi	2
023	qu'est-ce que	32	tu veux quoi	24
	tu veux quoi	1	qu'est-ce que	11
	que prends-tu	1	que veux-tu	1
	quelle boisson	1		

タスク 018 については、学習者は「EST-CE QUE」型の疑問文を好むのに対して、フランス語母語話者は疑問詞 *quoi* を用いたイントネーション型の疑問文を用いる傾向にあることが分かる。このほかにも、「この料理について説明してください」など、疑問詞を用いずに非直接的に質問するような言い回しも 12 例見られた。一方、学習者はこの種の言い回しは 3 例のみ見られただけである。

タスク 023 についても同様に、学習者はほとんど「EST-CE QUE」型を使用するのに対して、母語話者はイントネーション型が優勢である。

このことより、学習者は母語話者に比べてイントネーション型の疑問文を過少使用していることが伺える。

### 8.1.3.3.3. 否定辞 *ne* の保持と脱落

過少使用 N-grams で抽出されたもののうち、否定辞 *ne* の脱落に関する MWUs は *c'est pas* と *j'ai pas* であった。このうち、*c'est pas* については第八章 1.2.6 で既に扱ったため、ここでは *j'ai pas* について分析を行う。

*j'ai pas* と *ne* を保持した場合の *je n'ai pas* の使用頻度を比較したところ、以下の通りとなった。

表 302 否定辞 *ne* の保持と脱落

	母語話者	学習者
<i>je n'ai pas</i>	64	52
<i>j'ai pas</i>	63	1

以上の結果を基に、グループの間に有意差が見られるかどうかを検定したところ、有意差が見られた( $\chi^2=37.16$ ;  $p<.001$ )。これより、今までの分析と同様に初級学習者は否定辞 *ne* を保持する傾向にあると言える。

#### 8.1.3.3.4. *qu(e)*

*qu(e)*を含む過少使用 N-grams として、*(et)que ça* と *qu'on a* という N-grams が抽出された。*que* には、従属接続詞、従属接続詞句、関係代名詞、疑問詞、強調構文などの機能を持つため、それぞれの機能について *que ça* と *qu'on* の生起頻度を比較したところ、以下の通りとなった。

	母語話者	学習者
<i>que ça</i>		
従属接続詞	19	0
従属接続詞句	13	1
関係代名詞	1	0
疑問詞	4	0
強調構文	0	0
<i>qu'on</i>		
従属接続詞	41	0
従属接続詞句	23	0
関係代名詞	11	0
疑問詞	0	4
強調構文	1	0

以上より、圧倒的に学習者によるこれらの N-grams の生起頻度が低いことが分かる。また *qu'on* については、学習者は疑問詞として使用された N-grams しか見られなかったが、母語話者にはこのような使用は一例も見られなかった。したがって、生起頻度の差だけでなく使用方法の差も見られることが分かった。

#### 従属接続詞

félicitations HESIT j'espère **que ça** va bien se passer et plein de bonheur (FT21)

salut tu réponds pas j'aimerais qu'on se voit ce weekend (FT29)

#### 従属接続詞句

félicitations je vous souhaite plein de bonheur et pourvu **que ça** dure (FT28)

j'aime le film sur NP et HESIT fantaisie parce **que ça** m'intéresse beaucoup de regarder cinéma (JT08)

#### 関係代名詞

excusez-moi je voudrais avoir HESIT l'officiel des spectacles et surtout le prix **que ça** coûte merci (FT22)

ce weekend j'aimerais rester à la maison pour pouvoir profiter de l'appartement parce que je n' ai pas pu en profiter depuis **qu'on** a déménagé (FT17)

parce que il y a plein de choses **qu'on** peut faire (FT28)



### 疑問詞

est-ce que ça te dirait d'aller d'aller manger dans un petit restaurant sympa HESIT ce soir HESIT juste toi et moi (FT17)  
avec plaisir où est-ce qu'on se trouve (JT20)

### 強調構文

c'était comme ça qu'on s'est rencontrées (FT37)

#### 8.1.3.3.5. ディスコースマーカ―

過少使用 N-grams の中で、ディスコースマーカ―と関連すると考えられるものに *je sais* と *en fait* があった。

まず *je sais* について左側のコロケ―トを調べたところ、全 52 例について約 6 割を占めるのが *pas* であった (31 例)。*je sais pas* は「分からない」という意味を持ち、方略的役割を果たすこともある。これは、先行研究で記述の Aijmer (2004 : 187) が指摘するように、会話内容に対する責任の回避を示すことと同時に、話者の流暢さを促進するための使用である。このような方略的役割を持つ *je sais pas* の生起頻度は 21 例であった。したがって、このような機能を持つ *je sais pas* がこの *je sais* という N-gram の大半を占めていた。しかし、学習者コーパス内では *je sais* 自体の生起頻度が 3 回の上、このような機能を持つ *je sais pas* は一例も見られなかった。

### je sais pas

je suis un petit peu stressée mais j'arrive généralement à me déteindre donc je sais pas un peu perplexe mais HESIT je stresse plus avant l'examen que pendant donc voilà (FT34)  
j'aime pas HESIT j'aime pas faire la vaisselle parce que je je sais pas j'aime pas je veux bien faire le ménage mais les vaisselles j'aime pas ça j'aime pas j'aime pas travailler je suis feignant donc j'aime pas travailler (FT39)

次に *en fait* について分析を行う。この N-gram はフランス語のディスコースマーカ―の一つである。この N-gram の生起頻度はフランス語コーパス内には 45 例見られたが、学習者には 1 例しか見られなかった。さらに注目したいのは、フランス語母語話者がこの N-gram を使用したレンジであり、90.34%もの母語話者がこの N-gram を使用していた。したがって、フランス語母語話者のほとんどが使用していた *en fait* という N-gram について、学習者はほとんど使用していないということである。

### en fait

pendant les vacances en général j'aime un endroit calme donc soit la campagne surtout surtout la montagne en fait j'aime beaucoup beaucoup la montagne (FT03)  
je vis sur NP en fait depuis cinq ans HESIT depuis que j'ai quitté la maison de mes parents pour venir de faire mes études ici (FT05)  
mon dernier voyage c'était c'était jamais en fait parce que j'ai jamais vraiment beaucoup voyagé (FT16)

salut excuse-moi **en fait** j'ai oublié mon portable à la maison HESIT aujourd'hui et j'ai besoin d'appeler est-ce que tu peux me le prêter s'il te plaît (FT35)

### 8.1.3.3.6. エラー

タスク j009 で使用された *une heure et dix minutes* という MWU も過剰使用の MWU として抽出されている。これは、所要時間「1 時間 10 分」を説明するときに使用された MWU であるが、学習者コーパスに 15 例見られるのに対して、母語話者コーパスには 1 例も見られない。これは、2 つの要因が考えられる。まず、このコンテキストで接続詞 *et* を用いるのは規範から外れている。そして、*dix* の後ろに「分」を表す *minutes* を用いるのは慣例ではない。このような一種のエラーによってこの MWU の過剰使用が引き起こされたものと考えられる。

また、同様にエラーによって引き起こされた過剰使用 MWU は *je souvent* である。これは以下のタスクで使用されていた。

j010	普段どんな映画をよく見ますか？フランス語でどんな映画か説明してください。
j011	普段どんな本を読むのが好きですか？フランス語で説明してください。
j013	普段、自宅では何を飲みますか？
j036	普段どんなテレビをよく見ますか？フランス語で説明してください。
j037	あなたは普段、朝ごはんは何を食べますか？
j038	レストランでは普段、何を飲みますか？

以上の指示文からも分かる通り、これらは習慣を問うタスクである。*souvent* という副詞は、「しばしば」という意味を持つ。*souvent* の使用頻度を比較したところ、学習者は母語話者よりも有意に多くこの副詞を使用していた (学習者 58 回 vs 母語話者 63 回:  $\chi^2 = 8.28$ ;  $p > .01$ )。

MU レベルでも使用頻度の過剰使用が見られたが、習慣を述べる際にこの副詞を使用することはエラーではない。しかし、学習者コーパスには *je souvent* という MWU が 22 例見られたのに対して、母語話者コーパス内には一例も見られなかった。エラーとなった原因はこの副詞の位置である。これは、英語では一般動詞の前に副詞が用いられるが、フランス語では動詞の後に副詞が置かれる。このため、学習者は英語の影響もあり、副詞が一般動詞の前に置かれた可能性が考えられる。

フランス語母語話者：一般動詞＋副詞

je **bois souvent** du thé et du jus de fruits (FT26)

je ne **regarde** pas **souvent** la télévision un peu le journal (FT04)

je **prends** pas **souvent** de petit déjeuner (FT39)

日本人学習者：副詞＋一般動詞

je **souvent bois** un café au lait au restaurant (JT26)

je **souvent regarde** des matchs de foot à la télé (JT02)

je **souvent mange** du pain mais de temps en temps je manger du riz (JT15)

## 8.2. 自由会話 コーパス

### 8.2.1. 繰り返し使用される N-grams 数の比較

タスクに基づくコーパスと同様、N-grams のタイプとトークンの総数を求めたところ、以下のような結果となった(表 303)。

表 303 各コーパスの N-gram 総数

	学習者コーパス		母語話者コーパス	
	Type	Token	Type	Token
総語数	2064	34518	4463	65591
2 語連鎖数	11173	30806	23047	60805
3 語連鎖数	19107	27855	44754	56283
4 語連鎖数	22094	25337	47064	52096
5 語連鎖数	22029	23160	46733	48187

続いて、各 N-gram のトークンについて、10 万 N-gram あたりの調整頻度が 100、50、そして 20 という閾値を設け、それに値する実測値を求めたところ、以下のようになった。

表 304 10 万 N-gram あたり生起回数 100、50 および 20 以上となるための実測数

	学習者コーパス			母語話者コーパス		
	100	50	20	100	50	20
2 語連鎖数	30.81	15.40	6.17	60.81	30.40	12.16
3 語連鎖数	27.86	13.93	5.57	56.28	28.14	11.26
4 語連鎖数	25.34	12.67	5.07	52.10	26.05	10.42
5 語連鎖数	23.16	11.01	4.41	48.19	24.09	9.64

以上の回数を基に小数第一位以下を四捨五入し、トークンの調整頻度が 100 以上のもの、50 以上 100 未満のもの、20 以上 50 未満のもの、20 未満のものに分ける。トークンの閾値をもとに、各層に含まれるタイプ数と割合をまとめたものが以下の表である。

表 305 各層における N-gram のタイプ数とその割合

N-gram	調整頻度	学習者コーパス		母語話者コーパス	
		頻度	割合	頻度	割合
2 語連鎖	100<	107	0.96	84	0.36
	50≤99	150	1.34	120	0.52
	20≤49	519	4.65	459	1.99
	1≤19	10397	93.05	22384	97.12
3 語連鎖	100<	40	0.21	17	0.04
	50≤99	59	0.31	33	0.07
	20≤49	273	1.43	192	0.43
	1≤19	18735	98.05	44512	99.46
4 語連鎖	100<	9	0.04	3	0.01
	50≤99	22	0.10	6	0.01
	20≤49	150	0.68	63	0.13
	1≤19	21913	99.18	46992	99.85
5 語連鎖	100<	5	0.02	0	0.00
	50≤99	4	0.02	0	0.00
	20≤49	73	0.33	15	0.03
	1≤19	21947	99.63	46718	99.97

下の図が示す通り、学習者は調整頻度 20 回以上現れたタイプの N-grams について、母語話者よりも多くの種類の N-grams を使用していることが分かる。5 語連鎖については、生起頻度が 0 回の頻度層もあったため、5 語連鎖以外のカイ二乗検定を行ったところ、全ての N-grams において使用頻度の割合に有意差が見られた。そこで、ボンフェローニ補正後 1.25%水準で多重比較を行ったところ、全ての頻度層間に有意差が見られた。これは、頻度 100 以上、50 以上 100 未満、20 以上 50 未満の全ての頻度層において、学習者が使用した N-grams のタイプ数が母語話者よりも高いことを意味している。

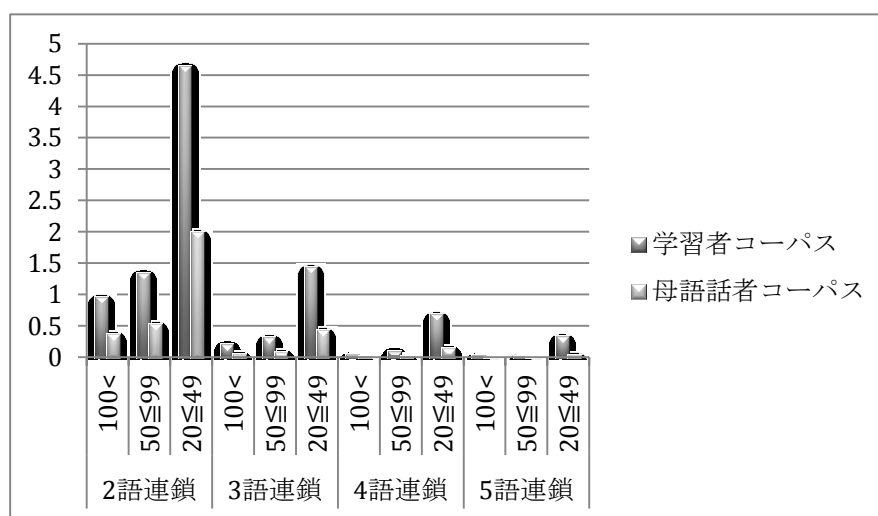


図 44 各層における N-gram のタイプ数とその割合 (調整頻度 20 以上)

また、調整頻度層で現れたトークン数とその割合についてまとめたものが表 306 である。

表 306 各層における N-gram のトークン数とその割合

N-gram	調整頻度	学習者コーパス		母語話者コーパス	
		頻度総計	割合	頻度総計	割合
2 語連鎖	100<	8432	27.37	11246	18.50
	50≤99	3026	9.82	4956	8.15
	20≤49	4489	14.57	8363	13.75
	1≤19	14859	48.23	36240	59.60
3 語連鎖	100<	2085	7.49	1837	3.26
	50≤99	1035	3.72	1242	2.21
	20≤49	2170	7.79	3102	5.51
	1≤19	22565	81.01	50102	89.02
4 語連鎖	100<	410	1.62	200	0.38
	50≤99	387	1.53	249	0.48
	20≤49	989	3.90	909	1.74
	1≤19	23551	92.95	50738	97.39
5 語連鎖	100<	143	0.62	0	0.00
	50≤99	53	0.23	0	0.00
	20≤49	407	1.76	218	0.45
	1≤19	22557	97.40	47969	99.55

図 45 より、調整頻度 20 以上の 2 語連鎖から 5 語連鎖までの N-grams について、全て学習者によるトークン数が母語話者よりも多いことが分かる。カイ二乗検定を行ったところ、1.25%水準で調整頻度 20 以上 50 未満の 2 語連鎖のトークン数に有意差が見られなかったが、その他については学習者と母語話者に有意差が見られた。特に学習者は母語話者と比べて調整頻度 100 以上の 2 語連鎖の割合が突出していることが分かる。

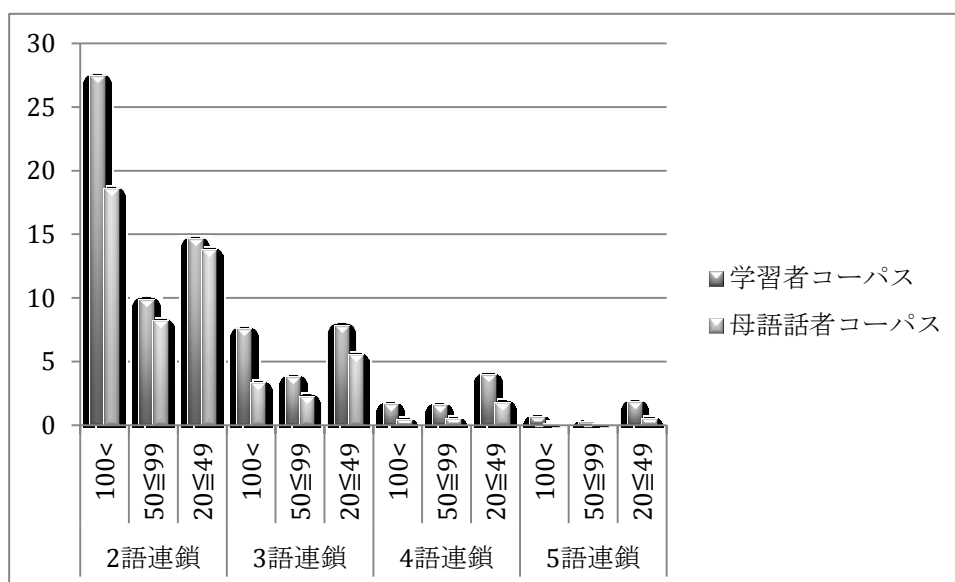


図 45 各層における N-gram のトークン数とその割合 (調整頻度 20 以上)

次に、各 N-gram の頻度上位 10 位までに限定し、高頻度で使用された N-grams について比較を行う。このことによって、頻度が最も高かった N-grams について、学習者が母語話者よりも有意に多く使用しているという以上までの結果の裏付けを行う事が出来る。表 307 は、各 N-gram の上位 10 位についてまとめたものである。なお、調整頻度は 10 万 N-grams における頻度を指し、レンジはグループ内における N-grams を利用した人数の割合を指す。

表 307 各 N-gram の上位 10 位

2 語連鎖	学習者コーパス				母語話者コーパス			
	N-grams	頻度	調整頻度	レンジ	N-grams	頻度	調整頻度	レンジ
1	C'EST	1018	3304.55	100.00	C'EST	1505	2475.13	100.00
2	J'AI	392	1272.48	92.11	IL Y	599	985.12	100.00
3	IL Y	273	886.19	81.58	Y A	452	743.36	100.00
4	Y A	234	759.59	76.32	J'AI	434	713.76	95.00
5	C'ÉTAIT	223	723.88	92.11	BC BC	322	529.56	60.00
6	OUI OUI	220	714.15	63.16	C'ÉTAIT	281	462.13	100.00
7	À NP	179	581.06	84.21	PARCE QUE	242	397.99	95.00
8	INT OUI	166	538.86	68.42	EN FAIT	217	356.88	100.00
9	DE NP	141	457.70	76.32	EST PAS	188	309.19	100.00
10	JE SUIS	139	451.21	92.11	JE SAIS	188	309.19	80.00
3 語連鎖	N-grams	頻度	調整頻度	レンジ	N-grams	頻度	調整頻度	レンジ
1	IL Y A	228	818.52	76.32	IL Y A	433	769.33	100.00
2	C'EST C'	128	459.52	63.16	C'EST PAS	144	255.85	90.00
3	EST C'EST	118	423.62	57.89	C'EST VRAI	139	246.97	75.00
	OUI C'EST	118	423.62	57.89				
4	OUI OUI OUI	80	287.20	44.74	JE SAIS PAS	124	220.32	75.00
5	C'EST TRÈS	65	233.35	52.63	HESIT C'EST	107	190.11	95.00
6	EST-CE QUE	64	229.76	63.16	QUE C'EST	92	163.46	80.00
7	C'EST ÇA	63	226.17	34.21	C'EST UN	88	156.35	75.00
	C'EST PAS	63	226.17	47.37				
8	JE SUIS ALLÉE	53	190.27	44.74	BC BC BC	87	154.58	55.00
					IL Y AVAIT	87	154.58	75.00
9	INT C'EST	52	186.68	73.68	MAIS C'EST	79	140.36	90.00
10	NP C'EST	51	183.09	50.00	C'EST C'	72	127.92	80.00
4 語連鎖	N-grams	頻度	調整頻度	レンジ	N-grams	頻度	調整頻度	レンジ
1	C'EST C'EST	118	465.72	57.89	C'EST C'EST	71	136.29	80.00
2	EST-CE QUE TU	46	181.55	42.11	IL Y A DES	68	130.53	65.00
3	OUI C'EST ÇA	43	169.71	23.68	C'EST VRAI QUE	61	117.09	55.00
4	J'AI J'AI	38	149.98	36.84	IL Y A PAS	47	90.22	85.00
	QU EST-CE QUE	38	149.98	44.74	QU IL Y A	47	90.22	85.00
5	IL Y A IL	34	134.19	36.84	IL Y A UN	46	88.30	60.00
	Y A IL Y	34	134.19	36.84				
6	A IL Y A	31	122.35	34.21	IL Y EN A	45	86.38	60.00
7	IL Y A BEAUCOUP	28	110.51	36.84	BC BC BC BC	37	71.02	35.00
8	CE QUE TU AS	23	90.78	31.58	C'EST UN PEU	27	51.83	60.00
	OUI C'EST VRAI	23	90.78	23.68				
	Y A BEAUCOUP DE	23	90.78	31.58				
9	INT OUI C'EST	22	86.83	23.68	HESIT IL Y A	24	46.07	65.00
10	C'EST UN PEU	21	82.88	31.58	QU EST-CE QUE	23	44.15	45.00
	JE PENSE QUE C'	21	82.88	23.68				



5 語連鎖	学習者コーパス				母語話者コーパス			
	N-grams	頻度	調整頻度	レンジ	N-grams	頻度	調整頻度	レンジ
1	IL Y A IL Y	34	146.80	36.84	IL Y A IL Y	22	45.66	55.00
2	QU EST-CE QUE TU	32	138.17	31.58	Y A IL Y A	21	43.58	50.00
3	Y A IL Y A	31	133.85	34.21	C'EST VRAI QUE C'	18	37.35	35.00
					IL Y A PAS DE	18	37.35	50.00
4	EST-CE QUE TU AS	23	99.31	31.58	EST VRAI QUE C'EST	16	33.20	35.00
	IL Y A BEAUCOUP DE	23	99.31	31.58	QU EST-CE QUE TU	16	33.20	35.00
5	JE PENSE QUE C'EST	18	77.72	21.05	MAIS C'EST VRAI QUE	15	31.13	35.00
6	C'EST C'EST TRÈS	12	51.81	23.68	BC BC BC BC BC	14	29.05	15.00
	JE SUIS ALLÉE EN NP	12	51.81	15.79				
7	JE SUIS ALLÉE À NP	11	47.50	18.42	C'EST À DIRE QUE	12	24.90	30.00
					IL Y A DES GENS	12	24.90	40.00
					IL Y EN A QUI	12	24.90	30.00
8	C'EST C'EST C'	9	38.86	15.79	BUREAU DE LA VIE ÉTUDIANTE	11	22.83	15.00
	C'EST C'EST UN	9	38.86	15.79	QU IL Y A DES	11	22.83	45.00
	ET QU EST-CE QUE	9	38.86	15.79				
	IL Y A PAS DE	9	38.86	10.53				
9	EST C'EST C'EST	8	34.54	15.79	C'EST POUR ÇA QUE	10	20.75	35.00
	OUI OUI C'EST ÇA	8	34.54	15.79	HESIT C'EST VRAI QUE	10	20.75	35.00
	OUI OUI OUI C'EST	8	34.54	10.53				
	PARCE QUE IL Y A	8	34.54	10.53				
	QUE C'EST C'EST	8	34.54	13.16				
10	C'EST C'EST LA	7	30.22	15.79	C'EST VRAI QU IL	8	16.60	40.00
	C'EST C'EST LE	7	30.22	18.42	EN MÊME TEMPS C'EST	8	16.60	25.00
	C'EST POUR ÇA QUE	7	30.22	13.16	IL Y A QUAND MÊME	8	16.60	40.00
	CE QUE TU AS MANGÉ	7	30.22	13.16				
	DONC C'EST C'EST	7	30.22	13.16				
	INT OUI C'EST ÇA	7	30.22	10.53				
	OUI C'EST C'EST	7	30.22	15.79				
	OUI J'AI J'AI	7	30.22	13.16				
	PARCE QU IL Y A	7	30.22	15.79				

図 46 は、2 語連鎖の上位 10 位の調整頻度について、学習者と母語話者を比較したものである。第 4 位の頻度には有意差が見られなかったが、それ以外の頻度上位の 2 語連鎖については、学習者の方が有意に多くこれらの 2 語連鎖を使用していた。特に注目すべきはそれぞれ第 1 位の頻度の高さである。これは、どちらのグループとも *c'est* という MWU であった。また、学習者は母語話者よりもこの頻度が有意に高いことより、学習者の発話にはこの MWU の使用が特に顕著であるといえる。

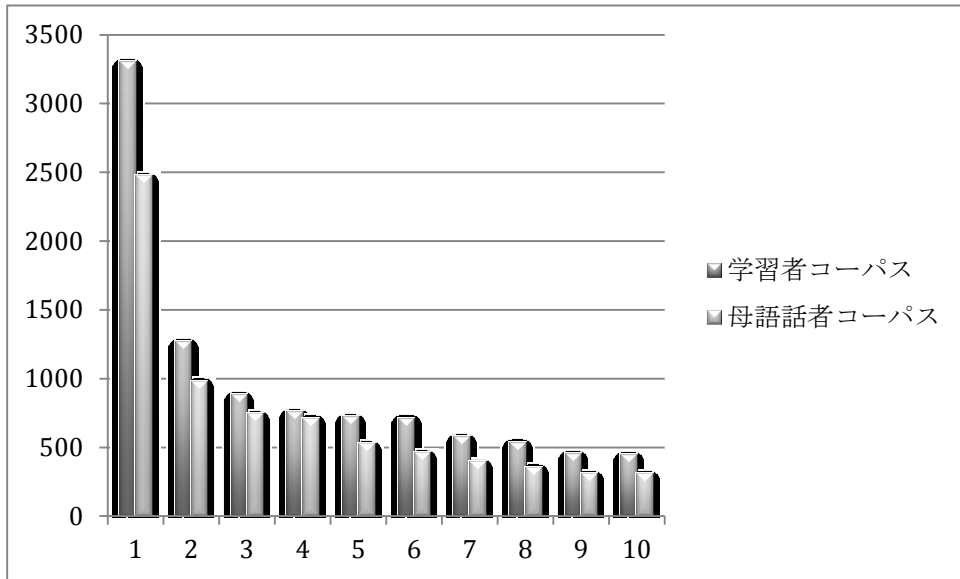


図 46 2 語連鎖上位 10 位の頻度

続いて、3 語連鎖の上位 10 位についてまとめたものが以下の図である(図 47)。第 1 位については、両コーパスとも *ilya* という MWU であったが、有意差は見られなかった。

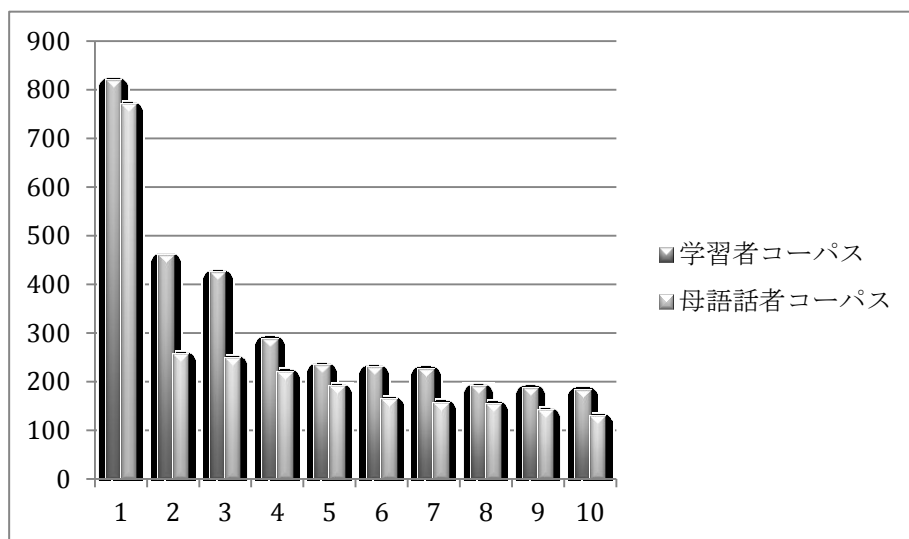


図 47 3 語連鎖上位 10 位の頻度

その他の順位に関しては、第 2、3、6、7、10 位については有意差が見られた。つまり、学習者の方が頻度の高い 3 語連鎖により依存しているといえる。また、学習者も母語話者も第 1 位から第 2 位にかけて頻度が大幅に下がっているが、特に母語話者にその傾向が顕著であった。

次は 4 語連鎖についてまとめた表である。

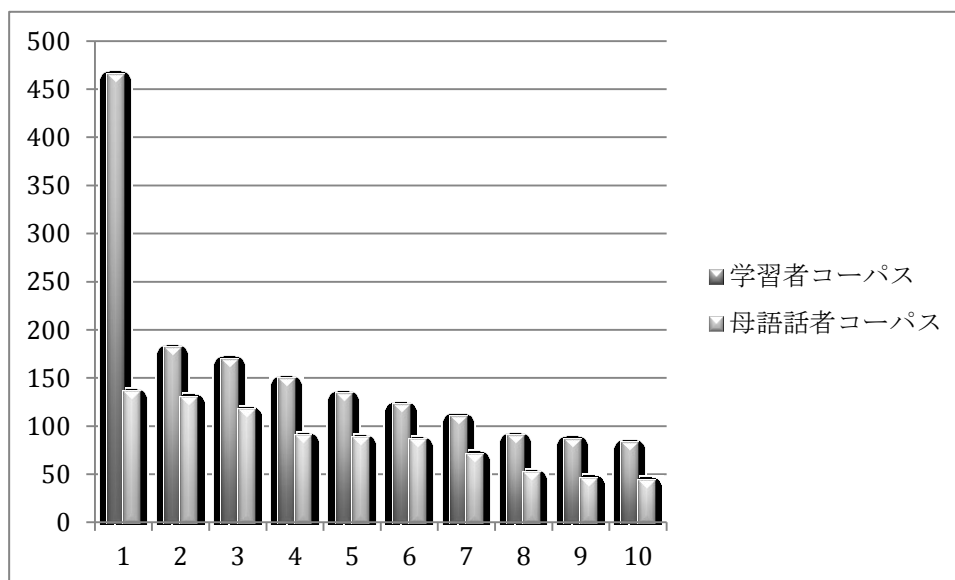


図 48 4 語連鎖上位 10 位の頻度

なお、両グループとも第一位の 4 語連鎖は *c'est c'est* で共通しており、これは *c'est* の繰り返しである。表から分かる通り、学習者はこの MWU の使用が極めて高く、有意差が見られた。一方、それ以外の 4 語連鎖に有意差が見られたのは第 4、9、10 位だけである。したがって、学習者は 4 語連鎖については *c'est c'est* に大きく依存していると言える。

5 語連鎖に関しては以下のような結果となった。有意差検定の結果、第 1 位から第 5 位までは有意差が見られたが、第 6 位以降は有意差が見られなかった。

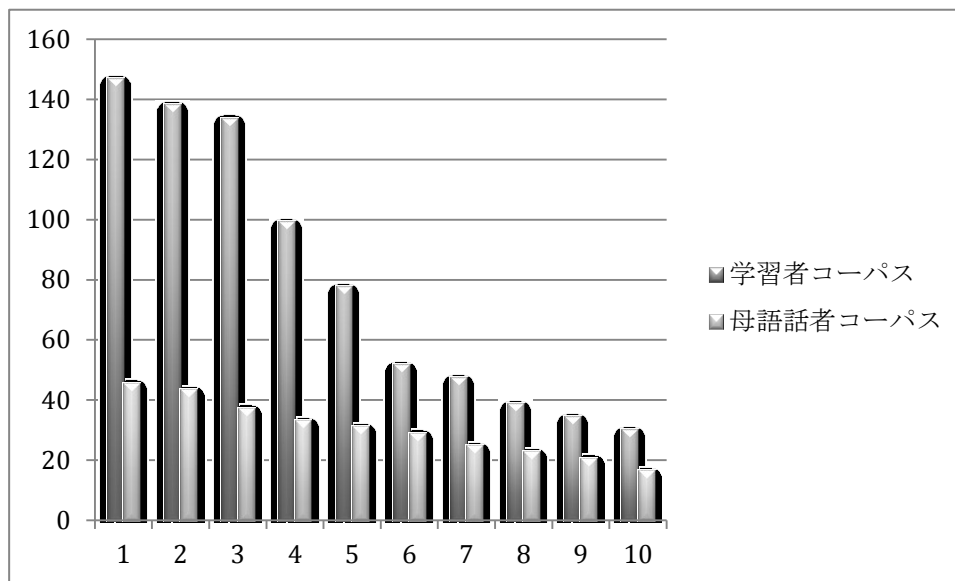


図 49 5 語連鎖上位 10 位の頻度

以上の点を総合すると、頻度の最上位である 10 位までの N-grams について学習者と母語話者で有意差が見られるものが多かったが、有意差がないものもあった。しかし、最上位に限定すると、学習者の方が一部の N-grams を繰り返し使用しているという事が分かった。つまり、Kjellmer (1991:124)が学習者言語について「既成のレンガではなく、自前のレンガを使っている」との仮説を棄却した De Cock et al. (1998)、De Cock (2004, 2007)の見解を支持する結果となった。

#### 8.2.2. 特徴的 N-grams と語彙分析の関連

タスクに基づくコーパスの分析と同様、WordSmith Tools(ver.6.0)を用いて抽出した特徴的 N-grams について、語彙の豊かさ、特徴語、コロケーションといったこれまで行ってきた語彙分析の観点との関連を調査する。

##### 8.2.2.1. 特徴的 N-grams と頻度層

過剰使用として特徴的な 5 語連鎖の頻度層、学習者コーパスと母語話者コーパスの実測値と調整頻度、そして Keyness をまとめたものが表 308 である。

頻度層を見てみると K1 機能語と K1 内容語が大半を占めていた。K2 に含まれていたのは *vin* と *visité* のみであり、K3 に含まれていたのは *dommage* だけであった。*vin* も *visité* もトピックに依存する語であると考えられる。また、*dommage* は *c'est dommage* という MWU 内で使用された以外の例は一度も見られなかった。なお、この MWU は「残念だ」という意味を持ち、低頻度語である *dommage* を含むが口語で特徴的な表現であり、MWU レベルでは難しい表現ではない。

表 308 特徴的 2 語連鎖と 3 語連鎖に含まれる K1 レベル以外の単語

頻度層	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
K2			vin	visité
K3			visité	
Off-list			dommage	

一方、過少使用的 MWUs については、ためらい語や固有名詞を除いてほぼ高頻度語のみで構成されていた。また、唯一高頻度語以外に分類されたのは Off-list の *ouais* であるが、これは K1 内容語である *oui* のより非形式的な表現であり、機能は *oui* とは変わらない。したがって、過少使用的 MWUs は語の頻度によって決定されるのではなく、むしろ高頻度語がどのように結びつくかによって決定されるものであると考えられる。

表 309 特徴的 5 語連鎖(過剰使用)

順位	特徴的 MWUs	頻度層					NNS	調整頻度	NS	調整頻度	Keyness
1	EST-CE QUE TU AS	K1_F	K1_F	K1_F	K1_F	K1_F	23	99.309	2	4.15	36.74
2	JE SUIS ALLÉE EN NP	K1_F	K1_F	K1_C	K1_F	NA	12	51.813	0	0.00	25.56
3	IL Y A BEAUCOUP DE	K1_F	K1_F	K1_F	K1_C	K1_F	23	99.309	6	12.45	24.49
4	JE PENSE QUE C'EST	K1_F	K1_C	K1_F	K1_F	K1_F	18	77.720	4	8.30	20.86
5	QU EST-CE QUE TU	K1_F	K1_F	K1_F	K1_F	K1_F	32	138.169	16	33.20	20.58

表 310 特徴的 4 語連鎖(過剰使用)

順位	特徴的 MWUs	頻度層				NNS	調整頻度	NS	調整頻度	Keyness
1	C'EST C'EST	K1_F	K1_F	K1_F	K1_F	118	465.72	71	136.29	61.26
2	OUI C'EST ÇA	K1_C	K1_F	K1_F	K1_F	43	169.71	8	15.36	54.05
3	J'AI J'AI	K1_F	K1_F	K1_F	K1_F	38	149.98	12	23.03	35.98
4	EST-CE QUE TU	K1_F	K1_F	K1_F	K1_F	46	181.55	19	36.47	35.51
5	CE QUE TU AS	K1_F	K1_F	K1_F	K1_F	23	90.78	3	5.76	32.93
6	PEUT ÊTRE C'EST	K1_C	K1_F	K1_F	K1_F	14	55.26	0	0.00	29.82
7	JE SUIS ALLÉE EN	K1_F	K1_F	K1_C	K1_F	14	55.26	0	0.00	29.82
8	IL Y A BEAUCOUP	K1_F	K1_F	K1_F	K1_C	28	110.51	8	15.36	28.27
9	JE PENSE QUE C'	K1_F	K1_C	K1_F	K1_F	21	82.88	4	7.68	26.13
10	SUIS ALLÉE EN NP	K1_F	K1_C	K1_F	NA	12	47.36	0	0.00	25.56
11	Y A BEAUCOUP DE	K1_F	K1_F	K1_C	K1_F	23	90.78	6	11.52	24.49
12	OUI OUI C'EST	K1_C	K1_C	K1_F	K1_F	20	78.94	4	7.68	24.35
13	OUI C'EST VRAI	K1_C	K1_F	K1_F	K1_C	23	90.78	7	13.44	22.31
14	ET QU EST CE	K1_F	K1_F	K1_F	K1_F	10	39.47	0	0.00	21.30
15	JE NE PEUX PAS	K1_F	K1_F	K1_C	K1_F	10	39.47	0	0.00	21.30
16	N'EST CE PAS	K1_F	K1_F	K1_F	K1_F	10	39.47	0	0.00	21.30
17	PENSE QUE C'EST	K1_C	K1_F	K1_F	K1_F	18	71.04	4	7.68	20.86
18	INT OUI C'EST	NA	K1_C	K1_F	K1_F	22	86.83	7	13.44	20.72

表 311 特徴的 3 語連鎖(過剰使用)

順位	特徴的 MWUs	頻度層			NNS	調整頻度	NS	調整頻度	Keyness
1	OUI C'EST	K1_C	K1_F	K1_F	118	423.62	43	76.40	100.96
2	J'AI VISITÉ	K1_F	K1_F	K2	37	132.83	0	0.00	78.82
3	C'EST C'	K1_F	K1_F	K1_F	128	459.52	72	127.92	72.26
4	OUI OUI OUI	K1_C	K1_C	K1_C	80	287.20	30	53.30	66.90
5	C'EST TRÈS	K1_F	K1_F	K1_F	65	233.35	19	33.76	64.73
6	EST C'EST	K1_F	K1_F	K1_F	118	423.62	71	126.15	61.26
7	JE SUIS ALLÉE	K1_F	K1_F	K1_C	53	190.27	13	23.10	58.41
8	J'AI J'	K1_F	K1_F	K1_F	49	175.91	14	24.87	49.48
9	C'ÉTAIT TRÈS	K1_F	K1_F	K1_F	32	114.88	4	7.11	46.43
10	INT C'EST	NA	K1_F	K1_F	52	186.68	19	33.76	44.36
11	TU ES ALLÉE	K1_F	K1_F	K1_C	17	61.03	0	0.00	36.21
12	AI J'AI	K1_F	K1_F	K1_F	38	136.42	12	21.32	35.98
13	PEUT ÊTRE C'	K1_C	K1_F	K1_F	16	57.44	0	0.00	34.08
14	ON EST ALLÉ	K1_F	K1_F	K1_C	16	57.44	0	0.00	34.08
15	INT D'ACCORD	NA	K1_F	K1_C	43	154.37	18	31.98	32.80
16	EST-CE QUE	K1_F	K1_F	K1_F	64	229.76	41	72.85	30.52
17	SUIS ALLÉE EN	K1_F	K1_C	K1_F	14	50.26	0	0.00	29.82
18	ÊTRE C'EST	K1_F	K1_F	K1_F	14	50.26	0	0.00	29.82
19	JE PENSE QUE	K1_F	K1_C	K1_F	46	165.14	23	40.86	29.59
20	ALLÉE EN NP	K1_C	K1_F	NA	17	61.03	1	1.78	29.33
21	QUAND J'ÉTAIS	K1_F	K1_F	K1_F	21	75.39	3	5.33	29.18
22	A BEAUCOUP DE	K1_F	K1_C	K1_F	27	96.93	7	12.44	28.85
23	Y A BEAUCOUP	K1_F	K1_F	K1_C	28	100.52	8	14.21	28.27
24	TU AS VISITÉ	K1_F	K1_F	K2	13	46.67	0	0.00	27.69
25	PAS BEAUCOUP DE	K1_F	K1_C	K1_F	20	71.80	3	5.33	27.32
26	PENSE QUE C'	K1_C	K1_F	K1_F	21	75.39	4	7.11	26.13
27	CE QUE TU	K1_F	K1_F	K1_F	46	165.14	26	46.20	25.78
28	NP DE NP	NA	K1_F	NA	19	68.21	3	5.33	25.48
29	OUI J'AI	K1_C	K1_F	K1_F	19	68.21	3	5.33	25.48
30	OUI OUI C'	K1_C	K1_C	K1_F	20	71.80	4	7.11	24.35
31	IL EST IL	K1_F	K1_F	K1_F	11	39.49	0	0.00	23.43
32	INT OUI C'	NA	K1_C	K1_F	23	82.57	7	12.44	22.31
33	C'EST DOMMAGE	K1_F	K1_F	K3	17	61.03	3	5.33	21.84
34	NE PEUX PAS	K1_F	K1_C	K1_F	10	35.90	0	0.00	21.30
35	ALLÉ À NP	K1_C	K1_F	NA	10	35.90	0	0.00	21.30
36	JE JE VAIS	K1_F	K1_F	K1_C	10	35.90	0	0.00	21.30
37	EST CE PAS	K1_F	K1_F	K1_F	10	35.90	0	0.00	21.30
38	ET QU EST	K1_F	K1_F	K1_F	10	35.90	0	0.00	21.30
39	JE NE JE	K1_F	K1_F	K1_F	10	35.90	0	0.00	21.30

表 312 特徵的 2 語連鎖(過剩使用)

順位	特徵的 MWUs	頻度層		NNS	調整頻度	NS	調整頻度	Keyness
1	OUI OUI	K1_C	K1_C	220	714.15	96	157.88	162.16
2	DE NP	K1_F	NA	141	457.70	43	70.72	136.82
3	À NP	K1_F	NA	179	581.06	75	123.35	136.73
4	OUI C'	K1_C	K1_F	129	418.75	48	78.94	108.62
5	BEAUCOUP DE	K1_C	K1_F	98	318.12	25	41.12	105.78
6	EN NP	K1_F	NA	135	438.23	58	95.39	100.78
7	JE JE	K1_F	K1_F	86	279.17	21	34.54	95.04
8	INT OUI	NA	NA	166	538.86	96	157.88	90.67
9	AI VISITÉ	K1_F	K2	39	126.60	0	0.00	83.08
10	EST TRÈS	K1_F	K1_F	80	259.69	25	41.12	76.33
11	EST C'	K1_F	K1_F	128	415.50	73	120.06	71.07
12	ÉTAIT TRÈS	K1_F	K1_F	42	136.34	5	8.22	61.84
13	J'AI	K1_F	K1_F	392	1272.48	434	713.76	59.37
14	MAIS MAIS	K1_F	K1_F	44	142.83	7	11.51	58.86
15	SUIS ALLÉE	K1_F	K1_C	56	181.78	15	24.67	58.77
16	COMMENT DIRE	K1_C	K1_C	42	136.34	6	9.87	58.37
17	AU NP	K1_F	NA	71	230.47	31	50.98	52.18
18	LE FRANÇAIS	K1_F	K1_C	36	116.86	5	8.22	50.51
19	AI J'	K1_F	K1_F	49	159.06	14	23.02	49.48
20	JE SUIS	K1_F	K1_F	139	451.21	113	185.84	45.03
21	JE PENSE	K1_F	K1_C	90	292.15	57	93.74	43.62
22	D'ABORD	K1_F	K1_C	23	74.66	1	1.64	41.52
23	INT C'	NA	K1_F	59	191.52	27	44.40	41.49
24	PAS BEAUCOUP	K1_F	K1_C	34	110.37	7	11.51	40.86
25	MES AMIS	K1_F	K1_C	19	61.68	0	0.00	40.47
26	INT FW	NA	NA	19	61.68	0	0.00	40.47
27	INT D'	NA	K1_F	48	155.81	18	29.60	40.12
28	C'EST	K1_F	K1_F	1018	3304.55	1505	2475.13	38.56
29	FW ET	NA	K1_F	18	58.43	0	0.00	38.34
30	NP ET	NA	K1_F	102	331.10	77	126.63	37.76
31	EST CE	K1_F	K1_F	90	292.15	63	103.61	37.68
32	EST ALLÉ	K1_F	K1_C	21	68.17	1	1.64	37.44
33	ALLÉE EN	K1_C	K1_F	21	68.17	1	1.64	37.44
34	ET ET	K1_F	K1_F	59	191.52	30	49.34	37.29
35	TU TU	K1_F	K1_F	52	168.80	24	39.47	36.27
36	ES ALLÉE	K1_F	K1_C	17	55.18	0	0.00	36.21
37	LE VIN	K1_F	K2	16	51.94	0	0.00	34.08
38	BC FW	NA	NA	16	51.94	0	0.00	34.08
39	ÊTRE C'	K1_F	K1_F	16	51.94	0	0.00	34.08



順位	特徴的 MWUs	頻度層		NNS	調整頻度	NS	調整頻度	Keyness	
40	NP NP	NA	NA	83	269.43	60	98.68		33.02
41	INT BC	NA	NA	21	68.17	2	3.29		32.83
42	JE VOUDRAIS	K1_F	K1_C	24	77.91	4	6.58		31.53
43	INT JE	NA	K1_F	33	107.12	11	18.09		30.11
44	JE N'	K1_F	K1_F	29	94.14	8	13.16		29.90
45	ALLÉ À	K1_C	K1_F	14	45.45	0	0.00		29.82
46	VISITÉ NP	K2	NA	14	45.45	0	0.00		29.82
47	PENSE QUE	K1_C	K1_F	46	149.32	23	37.83		29.59
48	AS VISITÉ	K1_F	K2	13	42.20	0	0.00		27.69
49	APRÈS APRÈS	K1_F	K1_F	13	42.20	0	0.00		27.69
50	OUI JE	K1_C	K1_F	37	120.11	16	26.31		27.42
51	SIX ANS	K1_C	K1_C	16	51.94	1	1.64		27.32
52	LA LANGUE	K1_F	K1_C	21	68.17	4	6.58		26.13
53	PRÈS DE	K1_F	K1_F	12	38.95	0	0.00		25.56
54	INT NP	NA	NA	19	61.68	3	4.93		25.48
55	TU AIMES	K1_F	K1_C	15	48.69	1	1.64		25.31
56	DE FW	K1_F	NA	15	48.69	1	1.64		25.31
57	SUIS ALLÉ	K1_F	K1_C	17	55.18	2	3.29		25.11
58	FW FW	NA	NA	17	55.18	2	3.29		25.11
59	A BEAUCOUP	K1_F	K1_C	33	107.12	14	23.02		24.88
60	OUI J'	K1_C	K1_F	27	87.65	9	14.80		24.63
61	INT INT	NA	NA	30	97.38	12	19.74		23.79
62	LE LE	K1_F	K1_F	49	159.06	31	50.98		23.77
63	MA COPINE	K1_F	K1_C	11	35.71	0	0.00		23.43
64	NP MAIS	NA	K1_F	34	110.37	16	26.31		23.26
65	TU CONNAIS	K1_F	K1_C	16	51.94	2	3.29		23.21
66	ET APRÈS	K1_F	K1_F	38	123.35	20	32.89		23.13
67	DONC J'	K1_F	K1_F	27	87.65	10	16.45		22.78
68	TRÈS TRÈS	K1_F	K1_F	27	87.65	10	16.45		22.78
69	MOI AUSSI	K1_F	K1_C	31	100.63	14	23.02		22.07
70	OUI ET	K1_C	K1_F	32	103.88	15	24.67		21.98
71	DE DE	K1_F	K1_F	78	253.20	68	111.83		21.93
72	J'AIME	K1_F	K1_C	55	178.54	40	65.78		21.65
73	MAIS JE	K1_F	K1_F	59	191.52	45	74.01		21.44
74	AUSSI J'	K1_C	K1_F	13	42.20	1	1.64		21.33
75	FRANÇAIS ET	K1_C	K1_F	13	42.20	1	1.64		21.33
76	LE LA	K1_F	K1_F	15	48.69	2	3.29		21.32
77	EN FRANÇAIS	K1_F	K1_C	15	48.69	2	3.29		21.32
78	TRÈS BON	K1_F	K1_C	15	48.69	2	3.29		21.32
79	AS MANGÉ	K1_F	K1_C	10	32.46	0	0.00		21.30
80	CE PAS	K1_F	K1_F	10	32.46	0	0.00		21.30
81	NE JE	K1_F	K1_F	10	32.46	0	0.00		21.30

順位	特徴的 MWUs	頻度層		NNS	調整頻度	NS	調整頻度	Keyness
82	NON NON	K1_C	K1_C	49	159.06	34	55.92	20.78
83	AUX NP	K1_F	NA	22	71.41	7	11.51	20.72
84	PAR EXEMPLE	K1_F	K1_C	48	155.81	33	54.27	20.65
85	C'ÉTAIT	K1_F	K1_F	223	723.88	281	462.13	20.62
86	ALLÉE À	K1_C	K1_F	19	61.68	5	8.22	20.13
87	JE J'	K1_F	K1_F	24	77.91	9	14.80	20.05
88	LE BUS	K1_F	K1_C	16	51.94	3	4.93	20.04

表 313 特徴的 4 語連鎖(過少使用)

順位	特徴的 MWUs	頻度層				NNS	調整頻度	NS	調整頻度	Keyness
1	C'EST VRAI QUE	K1_F	K1_F	K1_C	K1_F	3	11.84	61	117.09	-33.77
2	IL Y EN A	K1_F	K1_F	K1_F	K1_F	2	7.89	45	86.38	-25.78

324

表 314 特徴的 3 語連鎖(過少使用)

順位	特徴的 MWUs	頻度層			NNS	調整頻度	NS	調整頻度	Keyness
1	EST VRAI QUE	K1_F	K1_C	K1_F	3	10.77	61	108.38	-33.770
2	Y EN A	K1_F	K1_F	K1_F	2	7.18	46	81.73	-26.540

表 315 特徵的 2 語連鎖(過少使用)

順位	特徵的 MWUs	頻度層		NNS	調整頻度	NS	調整頻度	Keyness
1	BC OUAIS	NA	OFF	1	3.25	122	200.64	-93.76
2	QU ON	K1_F	K1_C	6	19.48	131	215.44	-74.36
3	TU VOIS	K1_F	K1_C	7	22.72	128	210.51	-68.16
4	TU SAIS	K1_F	K1_C	6	19.48	120	197.35	-66.07
5	ILS ONT	K1_F	K1_F	7	22.72	94	154.59	-43.56
6	NON MAIS	K1_C	K1_F	2	6.49	67	110.19	-42.83
7	QUE ÇA	K1_F	K1_F	4	12.98	78	128.28	-42.54
8	MAIS BON	K1_F	K1_C	2	6.49	65	106.90	-41.26
9	QU ILS	K1_F	K1_F	5	16.23	74	121.70	-35.97
10	QUAND MÊME	K1_F	K1_C	12	38.95	102	167.75	-35.12
11	VRAI QUE	K1_C	K1_F	3	9.74	62	101.97	-34.52
12	QUI SONT	K1_F	K1_F	2	6.49	56	92.10	-34.23
13	EN A	K1_F	K1_F	2	6.49	56	92.10	-34.23
14	CE QUI	K1_F	K1_F	3	9.74	61	100.32	-33.77
15	EN FAIT	K1_F	K1_C	50	162.31	217	356.88	-32.55
16	HESIT OUAIS	NA	OFF	9	29.22	84	138.15	-31.09
17	TOUT ÇA	K1_F	K1_F	2	6.49	51	83.87	-30.37
18	QU IL	K1_F	K1_F	28	90.89	147	241.76	-30.10
19	QUI EST	K1_F	K1_F	14	45.45	90	148.01	-23.77
20	DANS LES	K1_F	K1_F	2	6.49	42	69.07	-23.51
21	ENFIN JE	K1_C	K1_F	1	3.25	36	59.21	-23.38
22	DE FAIRE	K1_F	K1_C	1	3.25	36	59.21	-23.38
23	JE SAIS	K1_F	K1_C	50	162.31	188	309.19	-20.81
24	EST QUE	K1_F	K1_F	2	6.49	38	62.49	-20.52
25	ET TOUT	K1_F	K1_F	4	12.98	47	77.30	-20.23

#### 8.2.2.2. 特徴的 N-grams と特徴語の関係

特徴語として抽出された語が特徴的 MWUs に含まれているかどうかについて分析を行ったところ、表 316 のようになった。なお特徴語は太字で、灰色で色付けされている。

MWUs が特徴語を含む MWUs である。表から分かる通り、過剰使用 N-grams はほとんどが過剰使用語を含むものであった。過少使用語についても同様のことが指摘できる。つまり、特徴的 N-grams の大部分は特徴語によって構成されていた。

表 316 特徴的 N-gram と特徴語の関係 (過剰使用語)

順位	特徴的 MWUs			
	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
1	EST-CE QUE TU AS	C'EST C'EST	OUI C'EST	OUI OUI
2	JE SUIS ALLÉE EN NP	OUI C'EST ÇA	J'AI VISITÉ	DE NP
3	IL Y A BEAUCOUP DE	IL Y A IL	C'EST C'	À NP
4	JE PENSE QUE C'EST	Y A IL Y	OUI OUI OUI	OUI C'
5	QU EST-CE QUE TU	A IL Y A	C'EST TRÈS	BEAUCOUP DE
6		OUI C'EST VRAI	EST C'EST	EN NP
7		INT OUI C'EST	JE SUIS ALLÉE	JE JE
8		OUI OUI C'EST	J'AI J'	INT OUI
9		JE PENSE QUE C'	C'ÉTAIT TRÈS	AI VISITÉ
10		CE QUE TU AS	INT C'EST	EST TRÈS
11		PENSE QUE C'EST	TU ES ALLÉE	EST C'
12		PEUT ÊTRE C'EST	AI J'AI	ÉTAIT TRÈS
13		INT OUI OUI OUI	PEUT ÊTRE C'	J'AI
14		TU AS TU AS	ON EST ALLÉ	MAIS MAIS
15		SUIS ALLÉE EN NP	INT D'ACCORD	SUIS ALLÉE
16		OUI OUI OUI OUI	EST-CE QUE	COMMENT DIRE
17			SUIS ALLÉE EN	AU NP
18			ÊTRE C'EST	LE FRANÇAIS
19			JE PENSE QUE	AI J'
20			ALLÉE EN NP	JE SUIS
21			QUAND J'ÉTAIS	JE PENSE
22			A BEAUCOUP DE	D'ABORD
23			Y A BEAUCOUP	INT C'
24			TU AS VISITÉ	PAS BEAUCOUP
25			PAS BEAUCOUP DE	MES AMIS
26			PENSE QUE C'	INT FW
27			CE QUE TU	INT D'
28			NP DE NP	C'EST
29			OUI J'AI	FW ET

順位	特徴的 MWUs			
	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
30			OUI OUI C'	NP ET
31			IL EST IL	EST CE
32			INT OUI C'	EST ALLÉ
33			C'EST DOMMAGE	ALLÉE EN
34			NE PEUX PAS	ET ET
35			ALLÉ À NP	TU TU
36			JE JE VAIS	ES ALLÉE
37			EST CE PAS	LE VIN
38			ET QU EST	BC FW
39			JE NE JE	ÊTRE C'
40				NP NP
41				INT BC
42				JE VOUDRAIS
43				INT JE
44				JE N'
45				ALLÉ À
46				VISITÉ NP
47				PENSE QUE
48				AS VISITÉ
49				APRÈS APRÈS
50				OUI JE
51				SIX ANS
52				LA LANGUE
53				PRÈS DE
54				INT NP
55				TU AIMES
56				DE FW
57				SUIS ALLÉ
58				FW FW
59				A BEAUCOUP
60				OUI J'

順位	特徴的 MWUs			
	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
61				<b>INT INT</b>
62				LE LE
63				MA COPINE
64				<b>NP MAIS</b>
65				TU CONNAIS
66				ET APRÈS
67				<b>DONC J'</b>
68				<b>TRÈS TRÈS</b>
69				MOI AUSSI
70				<b>OUI ET</b>
71				DE DE
72				<b>J'AIME</b>
73				<b>MAIS JE</b>
74				<b>AUSSI J'</b>
75				<b>FRANÇAIS ET</b>
76				LE LA
77				<b>EN FRANÇAIS</b>
78				<b>TRÈS BON</b>
79				AS MANGÉ
80				CE PAS
81				<b>NE JE</b>
82				NON NON
83				<b>AUX NP</b>
84				PAR EXEMPLE
85				<b>C'ÉTAIT</b>
86				<b>ALLÉE À</b>
87				<b>JE J'</b>
88				LE BUS

表 317 特徴的 N-gram と特徴語の関係 (過少使用語)

順位	特徴的 MWUs			
	5 語連鎖	4 語連鎖	3 語連鎖	2 語連鎖
1		C'EST VRAI QUE	EST VRAI QUE	BC OUAIS
2		IL Y EN A	Y EN A	QU ON
3				TU VOIS
4				TU SAIS
5				ILS ONT
6				NON MAIS
7				QUE ÇA
8				MAIS BON
9				QU ILS
10				QUAND MÊME
11				VRAI QUE
12				QUI SONT
13				EN A
14				CE QUI
15				EN FAIT
16				HESIT OUAIS
17				TOUT ÇA
18				QU IL
19				QUI EST
20				DANS LES
21				ENFIN JE
22				DE FAIRE
23				JE SAIS
24				EST QUE
25				ET TOUT



### 8.2.3. 特徴的 N-grams の質的分析

#### 8.2.3.1. 特徴的 N-grams の傾向

特徴的 N-grams の質的分析としてまず、特徴的 N-grams の傾向を調べる。2 語連鎖から 5 語連鎖までで特徴的な N-grams として抽出されたものについて、重なりがあった場合は一緒にまとめ、表にした。

過剰使用 N-grams の傾向としては、特徴語が MWUs レベルまでに影響を及ぼしていることである。主に過剰使用語が目立つ N-grams をまとめたものが図 50 である。また、同じ語の繰り返しや言い直しなども過剰使用 N-grams に見られる傾向であった。その他には、否定辞の *ne* の保持、EST-CE QUE 型の疑問文などの発話スタイルの違いによる過剰使用も抽出された。談話レベルの特徴としては、発話の順序を明確に示す順接関係を示す N-grams や、方略的機能を持つものなども過剰使用として見られた。

全体的に複数の特徴的な N-grams が組み合わせあって、学習者に特徴的な文や節を構成しているものが多いことが分かる。表の最も色が濃い背景色は文や節を構成する 2 語連鎖を示し、色が薄い背景色は 3 語連鎖を示している。

図 50 過剰使用語を含む特徴的 N-grams

<b><u>ALLÉE</u></b>				
JE	SUIS	ALLÉE		
JE	SUIS	ALLÉE	EN	
JE	SUIS	ALLÉE	EN	NP
	SUIS	ALLÉE	EN	NP
		ALLÉE	EN	NP
			EN	NP
TU	ES	ALLÉE		
	ES	ALLÉE		
		ALLÉE	À	
<b><u>ALLÉ</u></b>				
JE	SUIS			
	SUIS	ALLÉ		
ON	EST	ALLÉ		
	EST	ALLÉ		
		ALLÉ	À	
		ALLÉ	À	NP
			À	NP
<b><u>VISITÉ</u></b>				
J'	AI			
J'	AI	VISITÉ		
	AI	VISITÉ		
		VISITÉ	NP	
TU	AS	VISITÉ		
	AS	VISITÉ		
		VISITÉ	NP	
<b><u>BEAUCOUP</u></b>				
IL	Y	A	BEAUCOUP	
IL	Y	A	BEAUCOUP	DE
	Y	A	BEAUCOUP	DE
		A	BEAUCOUP	DE
			BEAUCOUP	DE
PAS	BEAUCOUP			
PAS	BEAUCOUP	DE		
	BEAUCOUP	DE		
<b><u>INT, C'</u></b>				
INT	C'			
INT	C'	EST		
	C'	EST		
<b><u>INT</u></b>				
INT	D'			
INT	D'	ACCORD		
<b><u>C'</u></b>				
C'	EST	C'		
C'	EST	C'	EST	
	EST	C'	EST	
		C'	EST	
PEUT	ÊTRE	C'		
PEUT	ÊTRE	C'	EST	
	ÊTRE	C'	EST	
		C'	EST	

<b><u>JE</u></b>				
JE	PENSE			
JE	PENSE	QUE		
JE	PENSE	QUE	C'	
JE	PENSE	QUE	C'	EST
			C'	EST
			C'	EST
JE	VOUDRAIS			
<b><u>J'</u></b>				
QUAND	J'	ÉTAIS		
<b><u>OUI, J</u></b>				
OUI	J'			
OUI	J'	AI		
	J'	AI		
<b><u>OUI, JE</u></b>				
OUI	JE			
<b><u>TRÈS</u></b>				
C'	EST			
C'	EST	TRÈS		
	EST	TRÈS		
		TRÈS	BON	
C'	ÉTAIT			
C'	ÉTAIT	TRÈS		
	ÉTAIT	TRÈS		
		TRÈS	BON	

図 51 同じ語(句)の繰り返し

OUI	OUI			
OUI	OUI	OUI		
NON	NON			
MAIS	MAIS			
ET	ET			
TU	TU			
APRÈS	APRÈS			
LE	LE			
TRÈS	TRÈS			
DE	DE			
J'	AI			
J'	AI	J'		
J'	AI	J'	AI	
	AI	J'	AI	
		J'	AI	
C'	EST	C'		
C'	EST	C'		EST
	EST	C'		EST
		C'		EST
JE	JE			
JE	JE	VAIS		

図 52 言い直し

LE	LA
JE	J

図 53 否定辞  $n(e)$  の保持

JE	NE	PEUX	PAS
	NE	PEUX	PAS
JE	N'		

図 54 疑問文

ET	QU	EST			
ET	QU	EST	CE		
	QU	EST	CE	QUE	TU
		EST	CE	QUE	TU
		EST	CE	QUE	TU
			CE	QUE	TU
					AS
					AS

図 55 順接関係を示す N-grams

D'	ABORD
ET	APRÈS

図 56 方略的機能

COMMENT	DIRE			
PAR	EXEMPLE			
OUI	OUI	C'		
OUI	OUI	C'	EST	
	OUI	C'	EST	ÇA
	OUI	C'	EST	VRAI
INT	OUI			
INT	OUI	C'		
INT	OUI	C'	EST	
	OUI	C'	EST	
	OUI	C'	EST	ÇA
	OUI	C'	EST	VRAI
MOI	AUSSI			
	AUSSI	J'		

図 57 その他表現

C'	EST	
C'	EST	DOMMAGE
TU	CONNAIS	
J'	AIME	
TU	AIMES	
PRÈS	DE	

過少使用 N-grams は、過少使用語による影響と見られる N-grams、中性代名詞 *en* が含まれる N-grams、*qu(e)*や *qui* が含まれる N-grams、方略的機能を持つ N-grams、談話的機能を持つ N-grams などに特徴づけられた。

図 58 過少使用語を含む N-grams

<b>OUAIS</b>	
BC	OUAIS
HESIT	OUAIS
<b>ILS</b>	
ILS	ONT
<b>ENFIN</b>	
ENFIN	JE

図 59 中性代名詞 *en*

IL	Y	EN	A
	Y	EN	A
		EN	A

図 60 *qu(e)*, *qui*

QU'	ON
QU'	ILS
QU'	IL
QUI	SONT
QUI	EST
CE	QUI
QUE	ÇA

図 61 事実を示す N-grams

C'	EST	VRAI	QUE
	EST	VRAI	QUE
		VRAI	QUE

図 62 談話的機能

TU	VOIS
TU	SAIS
MAIS	BON
QUAND	MÊME
EN	FAIT
JE	SAIS
ET	TOUT
	TOUT ÇA

次に、これらの特徴的 N-grams について、語彙、統語、談話の観点からより質的な分析を行っていく。

### 8.2.3.2. 語彙レベルが影響を与えた特徴的 N-grams

語彙レベルが影響を与えた特徴的 N-grams として、過剰使用 N-grams では語や句の繰り返し、過少使用では中性代名詞の *en* に特徴づけられる。

#### 8.2.3.2.1. 語(句)の繰り返し

学習者によって使用されていた繰り返しを含む N-grams については、先行研究の De Cock (1998, 2004, 2007)で指摘されているが、これはフランス語を母語とする上級英語学習者の話し言葉コーパスに基づく研究の分析結果であった。本研究で使用した日本語母語話者による初級から上級学習者のコーパスにも、語や句の繰り返しを含む過剰使用 N-grams が多く見られた。これらの特徴は、句であっても一音節ということであり (*c'est c'est* [se se], *j'ai j'ai* [ʒɛ ʒɛ])、語レベルの繰り返しでも多くて二音節(*après après* [apʁɛ])のものであった。

母語話者、上級学習者、初級学習者の間でこれらの繰り返しについて、特に繰り返しが2回行われたものについて比較を行ったところ、以下の通りとなった。*oui, très, j'ai, je* の繰り返しについては特徴語分析で扱ったが、ここでも扱うものとする。まず、語の繰り返しについて扱う。カッコ内の数字は、10万の 2grams に対する調整頻度である。

表 318 語の繰り返しによって構成されている 2grams

	母語話者	上級学習者	初中級学習者
<i>oui oui</i>	67.43 (41)	191.49 (43)	346.31 (29)
<i>non non</i>	36.18 (22)	49.04 (11)	47.77 (4)
<i>mais mais</i>	8.22 (5)	129.28 (29)	95.33 (8)
<i>et et</i>	23.02 (14)	133.74 (30)	238.83 (20)
<i>tu tu</i>	24.67 (15)	106.99 (24)	238.83 (20)
<i>après après</i>	0.00 (0)	17.83 (4)	11.94 (1)
<i>le le</i>	41.12 (25)	124.82 (28)	131.86 (11)
<i>très très</i>	16.45 (10)	49.04 (11)	107.48 (9)
<i>de de</i>	75.65 (46)	182.77 (41)	47.77 (4)
<i>je je</i>	23.54 (21)	187.23 (42)	382.14 (32)
合計	327.28 (199)	1172.43 (263)	1647.96 (138)

以上のうち、*mais mais*、*après après*、*je je* については、上級学習者 > 初中級学習者 > 母語話者の順であり、*de de* については、上級学習者 > 母語話者 > 初中級学習者の順であったが、それら以外は全て初中級学習者 > 上級学習者 > 母語話者であった。また、特徴的 N-grams として抽出された繰り返しの語を含む N-grams の合計も、初中級学習者 > 上級学習者 > 母語話者の順となった。これより、語の繰り返しの多くは初中級学習者に顕著であることが分かった。

次に、*j'ai j'ai* と *c'est c'est* の句の連続について、各グループの生起頻度と、10万 4grams に対する調整頻度をまとめたものが以下である。

表 319 語の繰り返しによって構成されている 4grams

	母語話者	上級学習者	初中級学習者
<i>j'ai j'ai</i>	23.03 (12)	115.24 (22)	114.09 (9)
<i>c'est c'est</i>	103.65 (54)	450.47 (86)	240.15 (15)
合計	126.69 (66)	565.71 (108)	384.25 (24)

以上から、句の連続については、上級学習者>初中級学習者>母語話者の順であることが明らかになった。これより、いずれも語や句の繰り返しは学習者に特徴付けられていると言える。

#### 8.2.3.2.2. 中性代名詞 *en*

学習者による過少使用 N-grams の中に、*il y en a* という中性代名詞 *en* を含む N-gram が見られた。*en* には前置詞としての意味を持つ *en* もあるため、語レベルではこの中性代名詞が特殊語として抽出されなかったと考えられる。この中性代名詞が含まれている *il y en a* と、中性代名詞を含まない *il y a* の生起頻度を比較したところ、以下の通りとなった。

表 320 中性代名詞 *en* を含む N-grams

	母語話者	上級学習者	初中級学習者
<i>il y en a</i>	45	2	0
<i>il y a</i>	433	150	64

初中級学習者には中性代名詞 *en* を含む形である *il y en a* は見られなかった。そこで、母語話者と上級学習者の間で有意差検定を行ったところ、有意差が見られた ( $\chi^2=10.96$ ;  $p<.001$ )。したがって、中性代名詞は学習者によって使用されにくいものであると考えることができる。

#### 8.2.3.3. 統語レベルが影響を与えた特徴的 N-grams

統語レベルが影響を与えたと考えられる特徴的 N-grams について、過剰使用 N-grams では否定辞 *ne* の保持と、疑問文 EST-CE QUE 型の過剰使用が挙げられ、過少使用 N-grams では *qu(e)* の使用が挙げられる。

##### 8.2.3.3.1. 否定辞 *ne* の保持と脱落

*je ne peux pas* という N-grams はタスクに基づくコーパスでも過剰使用 N-gram として抽出されたものであるが、自由会話コーパスについても過剰使用 N-gram とされた。タスクに基づくコーパスでは、誘いを断るタスクでこの N-gram が見られたが、自由会話コーパスのコンコーダンスを調べたところそのような用法で使用されておらず、元の意味通り不可能なことを述べる際に使用されていた。この生起頻度は学習者コーパスで 10 例見ら



れたが、母語話者コーパスには一例も見られなかった。また、下の例の中で JA14 と JA15 はコンテキストより時制の選択が間違っていると考えられる。なぜなら、本来は半過去形で用いられなければならないからである。

**学習者による *je ne peux pas* の使用例**

ce tableau est très beau mais **je ne peux pas** comprendre pourquoi c'est ce tableau est aimé dans le monde entière (JD07)

je je dois faire petit boulot donc **je ne peux pas** aller à l'étranger (JD19)

**je ne peux pas** regarder tous les tous les tableaux (JA14)

j'ai dû rentrer rentrer chez moi HESIT au lendemain donc **je ne peux pas** écouter le concert (JA15)

一方、否定辞 *ne* が脱落した形である *je peux pas* の生起頻度を調べたところ、母語話者コーパス、学習者コーパスともに 4 例のみであった。学習者コーパスに見られたこの *ne* の脱落はすべて上級学習者によって使用されていた。

***je peux pas* の使用例**

mais ça **je peux pas** l'accepter quoi (FL08)

parce que si si je suis pas là **je peux pas** voir (FL09)

j'habitais avec des des copains donc **je peux pas** bien comparer la la la différence (JA06)

je peux partir là ici mais **je peux pas** revenir sans récépissé (JA07)

次に、過剰使用 N-grams に見られた *je n'* について分析を行う。これは、主語 *je* のうしろに否定辞 *ne* が来た形であるが、否定辞の後ろが母音か無音の *h* で始まる語であるためにエリジオンした形である。そこで、学習者コーパス内で *je n* の後ろに続く動詞を調べたところ *ai* が最も多く後ろに続き、*je n'ai pas* という否定文を構成していた例が 14 例見られた。また、母語話者と学習者について否定辞 *ne* の脱落した形である *j'ai pas* という生起頻度を調べた。*ne* の脱落と保持について有意差検定を行ったところ、有意差が見られた ( $\chi^2=25.45$ ;  $p<.001$ )。

表 321 否定辞 *ne* の保持と脱落

	母語話者	上級学習者	初中級学習者
<i>je n'ai pas</i>	4	6	8
<i>j'ai pas</i>	45	18	2

そこで多重比較を行ったところ、以下のようになった。

<i>ne</i> の保持	初中級学習者 > 上級学習者 = 母語話者
---------------	-----------------------

これより、特に初中級学習者に否定辞 *ne* の保持が見られることが分かり、これまでの結果を追認するものであった。

### 8.2.3.3.2. 疑問文

*qu'est-ce que tu as* という MWU が過剰使用 N-gram として抽出され、学習者コーパスには 17 例見られた。この右側の語を見てみると、動詞の過去分詞形が多く見られるため、過去の行動に対する質問文としてこの MWU が使用されていたと考えられる。*qu'est-ce que tu as* を含む MWU の数を比較したところ、以下のようになった。

表 322 *qu'est-ce que tu as* の頻度

	母語話者	上級学習者	初中級学習者
<i>qu'est-ce que tu as</i>	2	1	17

これより、この MWU は特に初中級学習者によって多く使用されていることが分かる。

さらに MWU に含まれる語数を限定して、疑問文で使用されている *est-ce que* と *qu'est-ce que* の使用頻度数について調べたところ、以下の通りとなった。

表 323 *est-ce qu(e)* と *qu'est-ce qu(e)* の頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>est-ce qu(e)</i>	6	6	17	$\chi^2=94.13$ ; $p<.001$
<i>qu'est-ce qu(e)</i>	30	5	35	$\chi^2=167.66$ ; $p<.001$

*est-ce qu(e)* も *qu'est-ce qu(e)* も、有意差が見られたため多重比較を行ったところ、初中級学習者が他の 2 グループよりも有意に多くこれらの MWUs を使用していることが分かった。以上の結果より、初中級学習者が最も EST-CE QUE 型の疑問文を使用していると言える。また、この他にも、疑問詞+EST-CE QUE の疑問文が初中級学習者には 7 例見られたのに対して、母語話者には 3 例しか見られず、上級学習者には 1 例も見られなかった。初級者が EST-CE QUE 型の疑問文を使用する傾向にあるというこの結果も、これまでの分析結果を追認するものであった。

### 8.2.3.3.3. *qu(e)*, *qui*

まず、*qu(e)* を含む特徴的 N-grams として、*qu'on*, *qu'il*, *qu'ils*, *que ça* が抽出された。タスクに基づくコーパス分析で指摘した通り、*qu(e)* には、従属接続詞、従属接続詞句、関係代名詞、疑問詞、強調構文などの機能を持つため、それぞれの機能について頻度を調べたところ、以下の通りとなった。

表 324 *qu(e)* を含む複文の頻度

	母語話者	上級学習者	初中級学習者
<i>qu'on</i>			
従属接続詞	56	0	0
従属接続詞句	26	3	0
関係代名詞	43	3	0
疑問詞	0	0	0
強調構文	4	0	0
<i>qu'il</i>			
従属接続詞	71	4	4
従属接続詞句	35	5	4
関係代名詞	34	3	0
疑問詞	6	2	6
強調構文	4	0	0
<i>qu'ils</i>			
従属接続詞	34	0	2
従属接続詞句	9	2	0
関係代名詞	27	1	0
疑問詞	3	0	0
強調構文	4	0	0
<i>que ça</i>			
従属接続詞	37	1	0
従属接続詞句	19	1	0
関係代名詞	8	0	0
疑問詞	0	1	0
強調構文	0	1	0

以上の N-grams について全体的に母語話者は *qu(e)* を従属接続詞として使用していることが分かる。また、従属接続詞句の一部や関係詞としての使用も多かった。一方、学習者の生起頻度を見てみると、母語話者と比べていずれの生起頻度もかなり低いことが分かる。特にその傾向は初中級学習者に顕著である。しかし、従属接続詞や従属接続詞句そして関係詞の使用頻度の低さのみが問題であるのではない。なぜなら、*ça* や *ils* も過少使用語であるため、*qu(e)* 以外の影響もあると考えられるからである。それでも、複文を構成するこれらの句の過少使用は注目に値する。

Andersen (2004)は、話し言葉は並列によって特徴付けられ、反対に書き言葉は従属節によって特徴付けられることは良く知られているが、ある種の従属節は口語でも頻繁に見られると指摘している。たとえば、*parce que* によって挿入された原因節は、副詞的従属節の中で最も頻度の高いものの一つである (ibid: 196)。また、Véronique (2008: 627)は、外国語としてのフランス語学習において、従属節はかなり後の段階で学習されるが、*parce que* のような従属節は早めに習得される。その後、関係節や補足説が学習されるという傾向が見られると述べている。そこで、この従属接続詞句を形成する *parce qu'on*, *parce qu'il*, *parce qu'ils*, *parce que ça* の使用頻度について比較を行ったところ、以下の通りとなった。同時に、従属接続詞句の頻度数に対する割合を示す。

表 325 *parce que* を含む N-grams の頻度

	母語話者	上級学習者	初中級学習者
<i>parce qu'on</i>	17 (65.38%)	2 (66.67%)	0
<i>parce qu'il</i>	19 (54.29%)	5 (100%)	4 (100%)
<i>parce qu'ils</i>	6 (66.67%)	2 (100%)	0
<i>parce que ça</i>	14 (73.68%)	1 (100%)	0

以上の結果から、*parce qu(e)* という従属接続詞句の使用頻度について学習者は母語話者よりも低かったが、従属接続詞句の使用は *parce qu(e)* に集中していることが分かった。

次に、*qui* について比較を行う。*qui* を含む *qui* は、関係代名詞、強調構文の一部、疑問詞の機能を持つため、これらの生起頻度について比較を行ったところ、以下の通りとなった。

表 326 *qui* を含む N-grams の頻度

	母語話者	上級学習者	初中級学習者
<i>qui sont</i>			
関係代名詞	49	2	0
疑問詞	0	0	0
強調構文	7	0	0
<i>qui est</i>			
関係代名詞	83	9	3
疑問詞	3	0	2
強調構文	4	0	0

以上のとおり、母語話者は関係代名詞としての *qui* を多く使用しているのが分かる。一方、Treffers-Daller (2009: 85)が指摘するように、フランス語学習者は関係詞を有意に過少使用していた。

これより、学習者は話し言葉において複文を使用する頻度が圧倒的に少ないということが言える。

#### 8.2.3.4. 談話レベルが影響を与えた特徴的 N-grams

談話レベルが影響を与えたと考えられる特徴的 N-grams のうち、過剰使用 N-grams について順接関係を示す N-grams と方略的使用の N-grams が見られた。また、過剰使用 N-grams と過少使用 N-grams で同意を示す N-grams が見られた。そして、過少使用 N-grams に特徴的であったのがディスコースマーカークの N-grams である。

##### 8.2.3.4.1. 順接関係を表す語

自由会話時に学習者はある一部の順接関係を表す N-grams を過剰使用することが分かった。それは、*d'abord* と *et après* という MWUs である。この結果は先行研究の Ishikawa (2007)の分析結果と同じく、学習者の発話が「affirmative-centered」であるということを追認するものであった。*d'abord* については、第八章コロケーション分析のタスクに基づくコーパスでも過剰使用が認められた(8.1.1.10)。ここでは、はじめに各グループによる *d'abord* の使用頻度について比較を行う。結果は、以下の通りであった。

表 327 *d'abord* の頻度

母語話者	上級学習者	初中級学習者
1	21	2

このように、この使用は特に上級学習者に多く見られるものであった。

次に、学習者による *d'abord* の使用について文脈を見てみると、上級学習者によって使用のエラーが見られることが分かった。上級学習者はこの *d'abord* という MWU について「最初に」のような順序を示す接続詞として使用しているのではなく、「はじめに、はじめは」というような *au début* や *au commencement* のような意味合いで使用している例が半数以上の 11 例見られた。

c'est très bien parce que **d'abord** j'hésitais j'hésitais sur sur la site INT à habiter (JA04)

c'est c'est facile à devient devenir un ami **d'abord** mais après je sais pas la la relation ne dure pas trop (JA06)

このようなエラーも *d'abord* の過剰使用の一因であると考えられるが、このような使用を除いてもグループ間に有意差が見られた ( $\chi^2=25.45$ ;  $p<.001$ )。多重比較の結果も、学習者グループが有意に多く使用しているという点で、変化はなかった。

次に、*et après* という MWU の頻度を比較したところ、以下の通りとなった。

表 328 *et après* の頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
20	34	4	$\chi^2=36.83$ ; $p<.001$

有意差が見られたため多重比較を行った結果、上級学習者と母語話者の使用頻度に有意差が見られた。したがって、*et après* は上級学習者に特徴的な MWU であると言える。

以上の結果より、*d'abord* や *et après* のような順接関係を明示的に示す N-grams は上級学習者によって特によく使用されることが分かった。

#### 8.2.3.4.2. 方略的使用の N-grams

*par exemple*(「たとえば」)は、前言に対して具体例を提示する際に使用される。

表 329 *par exemple* の頻度

母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
33	33	15	$\chi^2=24.57$ ; $p<.001$

多重比較を行ったところ、学習者間には有意差が見られなかったが、学習者グループと母語話者グループには有意差が見られ、学習者は母語話者よりも多くこの MWU を使用していることが分かった。*par exemple* という MWU は方略的使用と言える。なぜなら、具体例を示すことで前言の補足を行うことが出来るからである。使用例を以下に示した。

et on a mangé beaucoup **par exemple** la glace (JD03)

j'ai vu j'ai vu beaucoup de danses **par exemple** NP et danse traditionnelle de NP (JD08)

il cherche un travail ailleurs dans le monde entier donc s'il travaille **par exemple** aux NP oui peut être je vais (JA02)

il y a pleins de camions de publicité **par exemple** il y a la la compagnie de NP (JA07)

mais je te trouve que dans le cours **par exemple** dans les séminaires de master HESIT où j'ai assisté HESIT les étudiants étaient très sérieuses (JA12)

ouais le contenu des cours HESIT ils te disent grosso modo HESIT ce qu'on va faire quoi **par exemple** en linguistique je sais qu'on aura à faire (FL05)

et on s'est rendu compte **par exemple** quand j'ai fait des recherches sur internet (FL08)

lui NP il connaît **par exemple** quand on veut se s'inscrire en deux cursus (FL12)

次に、方略的使用の *comment dire* についての生起頻度を比較する。ここでの方略的使用とは、言葉に詰まった際などに「なんとというか」という意味で使用される *comment dire* のことを指す。この点については、七章 7.2.1.20. で扱ったため、例を示すのみとする。

c'était trop difficile et **comment dire comment dire** (JD19)

elle se demandait pourquoi aussi et elle **comment dire** elle était pas fâchée elle était pas en colère mais HESIT elle était énervée (JA01)

mais mais **comment dire** la stratégie la stratégie de l'enseignement c'est différent selon selon la nationalité (JA04)

elle est HESIT **comment dire** (FL01)

**comment dire** des des aspects de la vie étudiante que les (FL11)

次に、過剰使用 N-grams に同意を示す INT/*oui oui c'est ça*, INT/*oui oui c'est vrai, moi aussi j'* という N-grams について比較を行う。これらの N-grams の中から、同意を示す役割をもつ *c'est ça* (「そうだ」), *c'est vrai* (「本当だ」), *moi aussi* (「私も」) について生起頻度を比較したところ、以下の通りとなった。

表 330 同意を示す N-grams の頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>c'est ça</i>	35	57	5	$\chi^2=61.51$ ; $p<.001$
<i>c'est vrai</i>	53	31	5	$\chi^2=5.48$ ; $p>.05$
<i>moi aussi</i>	14	19	12	$\chi^2=29.35$ ; $p<.001$

*c'est vrai* については有意差が見られなかったため、それ以外の 2 つの N-grams について多重比較を行ったところ、以下のようになった。

<i>c'est ça</i>	上級学習者 > 初中級学習者 = 母語話者
<i>moi aussi</i>	初中級学習者 = 上級学習者 > 母語話者

いずれの場合も、上級学習者は母語話者よりもこれらの N-grams を有意に多く使用していた。そして、初中級学習者については *moi aussi* の使用頻度に母語話者との有意差が見られた。次に、*c'est ça* と *c'est vrai* について、その前に *oui* が先行する場合の生起頻度

を調べた。*oui* が先行することで、二重に同意の意思を示すこととなる。同時に、同意を示すの *c'est ça* と *c'est vrai* の総数との割合も示す。

表 331 *c'est ça* と *c'est vrai* を含む N-grams の頻度

	母語話者	上級学習者	初中級学習者	$\chi^2$ 統計量、有意確率
<i>oui c'est ça</i>	9 (25.71%)	39 (68.42%)	4 (80.00%)	$\chi^2=72.82$ ; $p<.001$
<i>oui c'est vrai</i>	7 (13.21%)	22 (70.97%)	1 (20.00%)	$\chi^2=38.29$ ; $p<.001$

まず割合を見てみると、*oui c'est ça* については、初中級学習者 > 上級学習者 > 母語話者のとなり、学習者が *c'est ça* を使用する際、多くの場合 *oui* を伴って使用することが分かった。*c'est vrai* の場合、母語話者と比較して上級学習者は高い割合で *oui* を伴い使用していることが分かった。次に、それぞれ有意差が見られたため多重比較を行ったところ、以下の通りとなった。

<i>oui c'est ça</i>	上級学習者 > 初中級学習者 = 母語話者
<i>oui c'est vrai</i>	上級学習者 > 初中級学習者 = 母語話者

どちらとも、上級学習者が他の 2 グループより有意にこれらの N-grams を多く使用していることが分かる。

以上を総合すると、特に上級学習者は聞き手に対して、同意の意思を明示的に示す傾向があることが分かった。

#### 8.2.3.4.3. ディスコースマーカー

ディスコースマーカーとしての機能を持つ過剰使用 N-gram として *n'est-ce pas* が抽出された。この N-gram は相手の同意を求めたり、情報の共有を促進させたりする際に使用される MWU である。この生起頻度を比較したところ、以下の通りとなった。

表 332 *n'est-ce pas* の頻度

	母語話者	上級学習者	初中級学習者
<i>n'est-ce pas</i>	0	5	5

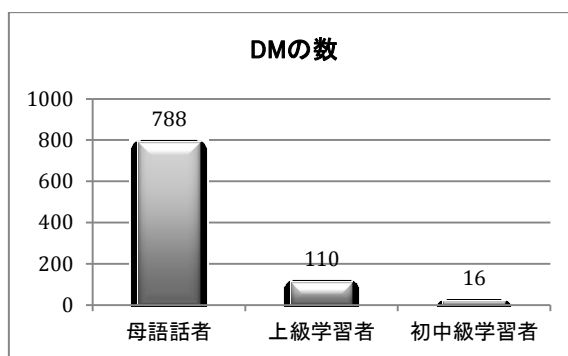
このように、この MWU を母語話者は一回も使用しておらず、なお学習者のみに使用が見られた。学習者間には有意差は見られなかった。この MWU の過剰使用の一因として、母語話者はこれと同じ機能をもつ別の語あるいは MWUs を使用していることが考えられる。たとえば、同じ機能を持つ語である *hein* は学習者によって過少使用されていたことは第八章 8.2.2.2. で記述の通りである。また、*tu vois* や *tu sais* といった MWUs についても第八章で既述の通りである上、過少使用 N-grams としても抽出されている。

ディスコースマーカーとしての機能を持つ過少使用 N-grams として、*tu vois*, *tu sais*, *mais bon*, *quand même*, *en fait*, *je sais (pas)*, *et tout ça* が抽出された。第八章で全て述べた通り、前言の補足や訂正の機能を持つ *en fait*、発言のニュアンスを緩和する機能を持つ

*quand même* や *bon*、情報の曖昧性を保つ役割をする *et tout (ça)*、情報の共有を促進させる *tu sais* や *tu vois*、そして会話内容に対する責任回避の姿勢を示す *je sais pas* といったディスコースマーカ―が過少使用されていた<sup>55</sup>。以上のディスコースマーカ―の生起頻度についてまとめたものが以下である。

表 333 DM の頻度

	母語話者	上級学習者	初中級学習者
<i>tu vois</i>	116	5	1
<i>tu sais</i>	101	0	0
<i>mais bon</i>	65	2	0
<i>quand même</i>	103	11	1
<i>en fait</i>	217	48	2
<i>je (ne) sais pas</i>	144	41	12
<i>et tout (ça)</i>	42	3	0
合計	788	110	16



表や図から、学習者はこれらのディスコースマーカ―の使用頻度が母語話者と比較して圧倒的に低いことが分かる。母語話者コーパスは上級学習者コーパスの約 1.90 倍、初中級学習者コーパスの 6.57 倍であるが、コーパスの規模を考慮に入れても使用頻度差は歴然である。

<sup>55</sup> それぞれ、詳細については以下を参照のこと：*en fait* (8.2.2.9), *tu vois* (8.2.2.12), *bon* (8.2.2.14), *et tout (ça)* (8.2.2.15), *quand même* (8.2.2.18), *tu sais* (8.2.2.19), *je sais pas* (8.2.2.19)



## 第九章 結論

本論文は、日本人フランス語学習者がフランス語を口頭産出する際に、母語話者の口頭産出と比較して何がどのように異なるのについてを使用語彙の観点から考察し、記述することを目的としてきた。そして、2種類の母語話者コーパスと2種類の学習者コーパスを用いて、使用語彙について「語彙の豊かさ」、「特徴語」、「コロケーション」、「N-grams」という4つの観点から数量的そして質的な分析を行い、記述した。「9.1. 本研究のまとめ」では、本論文で行った分析結果や考察を振り返り、まとめる。続く「9.2. 今後の課題」では、本研究の問題点と本研究の成果が今後どのように発展されるべきかについて今後の展望を述べる。

### 9.1. 本研究のまとめ

本節では、「語彙の豊かさの分析」、「特徴語分析」、「コロケーション分析」、「N-grams 分析」について、「タスクに基づくコーパス」の分析結果と「自由会話コーパス」の分析結果をまとめる。

#### 9.1.1. 語彙の豊かさの分析

語彙の豊かさの分析ではまず、1位から1000位までの機能語(K1 機能語)、1位から1000位までの内容語(K1 内容語)、1001位から2000位までの語(K2)、2001位から3000位までの語(K3)、それ以下の頻度の語(Off-list)の各頻度層におけるトークンとタイプの比率について、学習者コーパスと母語話者コーパスを比較した。次に、トークン数をタイプ数の平方根で割るギロー値についても2つのコーパスの比較を行った。その後、ギロー値と頻度層を変数として、ピアソンの積率相関係数を求めた。最後にギロー値を用いて主成分分析を行った。

##### 9.1.1.1. タスクに基づくコーパス

トークンとタイプについて各頻度層の比率を母語話者と学習者で比較したところ、K1 内容語のトークン比率、K1 機能語のタイプ比率、Off-list のトークン比率とタイプ比率について有意差が見られたが、トークンとタイプの比率を比較しても、頻度層 1000-3000位に母語話者と学習者の間で有意差が見られなかったのは、疑問の残る結果となった。そこで、各頻度層のタイプ数をトークンの平方根で割るギロー値を使用したところ、K1 機能語以外のすべての頻度層に有意差が見られた。次に、学習者と母語話者それぞれ、ギロー値と頻度層を変数としてピアソンの積率相関係数を求めたところ、学習者についてはK1 内容語と Off-list、K3 と Off-list の間に中程度の相関が見られた。また、母語話者についてはK1 機能語と K2 そして Off-list の間で中程度の負の相関が見られ、K1 内容語、K2、K3、Off-list については、それぞれの間で中程度から高い相関が見られた。つまり、K1 機能語のギロー値の値のみが異なる性質を持ち、K1 機能語のギロー値が低いか、そして他

の頻度層のギロー値が高いかで、ある程度学習者の言語能力を測れることが分かった。最後にギロー値を用いて主成分分析を行い、第 1 主成分得点をもとに学習者と母語話者の語彙の豊かさを点数化した。その結果、例外は見られたが総じて母語話者グループが学習者グループよりも得点が高かった。

#### 9.1.1.2. 自由会話 コーパス

タスクに基づくコーパスと同様の分析を行った。まず、タイプとトークンの頻度層の比率について、母語話者と学習者の間で有意差があるか検定を行ったところ、K1 機能語のトークンとタイプの比率、K1 内容語のタイプ比率、K2 のタイプ比率、K3 のタイプ比率、Off-list のタイプとトークンの比率について、有意差が見られた。したがって、タイプの比率を測定することによって語彙の豊かさを明らかにすることができることが分かった。次に、ギロー値を用いて母語話者と学習者の間で有意差があるか検定を行ったところ、全ての頻度層間に有意差が見られた。K1 機能語については、ギロー値は学習者の方が有意に高く、その他の頻度層については有意に値が低かった。次にギロー値と頻度層間の相関関係を調べたところ、学習者については K1 内容語から Off-list までについては中程度から高い相関が見られ、K1 機能語とその他の頻度層については中程度の負の相関が多く見られた。母語話者コーパスについては、全ての頻度層間のギロー値について高い相関が見られた。K1 機能語は高い負の相関を示しており、K1 機能語の特異性が伺えた。つまり、タスクに基づくコーパスの分析結果と同様に、K1 機能語のギロー値の値が低く、その他の頻度層のギロー値が高いかどうかで、学習者の言語能力をある程度測定することが出来るということが分かった。最後に主成分分析を用いて母語話者と学習者の第 1 主成分得点を求めたところ、母語話者の第 1 主成分得点は学習者よりもかなり高く、学習者グループの最高得点でも母語話者グループの平均に満たしていなかった。また、学習者グループの間には大きな得点差が見られたためグループ分けを行い、上位グループと下位グループの特徴を調べたところ、上位グループは留学経験を持ち修士以上に在籍する学習者であり、下位グループは留学経験がなく、学部生であった。これより、言語能力が高い学習者は母語話者の語彙の豊かさに近いということが分かった。

#### 9.1.2. 特徴語分析

特徴語分析ではまず、対数尤度比を用いて特徴語を算出した。過剰使用語は  $G2 \geq 50$ 、過少使用語は  $G2 \leq -50$  とした。次に、特徴語と頻度層との関係を分析した。最後に特徴語を品詞別に分類を行った。

##### 9.1.2.1. タスクに基づくコーパス

過剰使用語として *voudrais, elle, NP, ne, je, prenez, japonaise, vous, peux, abord, football, université*、過少使用語として *HESIT, on, donc, peu, bien, c', vraiment, tout* が抽出された。これらの特徴語についてまず、*VocabProfil* を用いて特徴語が頻度層に属するかについて分析を行った。その結果、過剰使用語も過少使用語も全て高頻度語(K1 機能語か K1 内容語)

であり、フランス語の頻度の上位 1000 位以内のものであることが分かった。最後に、特徴語を品詞別に分類したところ、過剰使用語に特徴付けられた品詞は人称代名詞(*elle, je, vous*)や固有名詞を含む名詞であることが分かった。一方、過少使用語に特徴付けられた品詞は副詞であり、多義語も多く見られた。

#### 9.1.2.2. 自由会話コーパス

タスクに基づくコーパスと同様の基準で過剰使用語と過少使用語を抽出したところ、過剰使用語は INT, *oui, NP, FW, très, visité, beaucoup, BC, allée, français, japonais, allé, pendant, est, c', musique, j', difficile, je, comment* であり、過少使用語は *enfin, hein, qui, quoi, les, ça, des, ils, fait, ouais, qu', vois, là, bon, tout avoir, leur, même, sais* であった。これらの特徴語について、どの頻度層に属するのかを調べたところ、過剰使用語は K2 の *visité* 以外は全て K1 機能語か K1 内容語であった。過少使用語は、Off-list の *ouais* 以外は全て K1 機能語と K1 内容語であった。しかし、*ouais* は *oui* の非形式的な形であり、話し言葉では高頻度で現れる。したがって特徴語のほとんどが頻度 1000 位までの高頻度語であった。次に、特徴語と品詞の関係を調べたところ、過剰使用語は代名詞主語(*c', je, j'*)、形容詞(*français, japonais*) が多く見られた。過少使用語は、タスクに基づくコーパスと同様に、多義語が多く見られた。また、過剰使用語には、*ah, oh* などの一音節で日本語でも長音母音として発せられる間投詞(INT)が含まれていたが、過少使用語には *quoi, hein* と行ったフランス語に特有の間投詞が多く見られた。

#### 9.1.3. コロケーション分析

コロケーション分析では、特徴語を中心語として調整頻度 20 以上の左右 3 語のコロケートをまとめた。そして、学習者と母語話者によって特徴語がどのように使用されているかについて比較を行った。グループ間の比較は、カイ二乗分析や対数尤度比を用いた共起強度を元に計量的観点から行った。また有意差が見られる場合には、どのような点に起因しているのかについて質的な考察を行った。なお、自由会話コーパスに基づく分析では、語彙の豊かさの分析で分けられた上位グループ(上級学習者グループ)とそれ以外の初中級学習者グループ間の比較も行った。最後に、タスクに基づくコーパス分析と自由会話コーパス分析とで一貫して見られた学習者による特徴的使用についてまとめた。ここでは結果の一例を挙げることにする。

##### 9.1.3.1. タスクに基づくコーパス

過剰使用語とのコロケーション分析を行ったところ、まず学習者と母語話者とで様々な WUs レベルの違いが見られた。学習者の特徴的な使用としてまず *elle* や *je* といった人称代名詞主語の過剰使用が明らかになった。これは、学習者の発話ターン内における使用語彙数が少ないため、主語を必ず伴わなければならないフランス語では必然的に主語の割合が高くなることに起因していると考えられる。しかしその他の原因として、学習者が人称

代名詞主語を使用する際に母語話者は非人称主語を使用している例なども確認できた。その他として、否定辞 *ne* を学習者は保持する傾向にあることが分かった。さらに、依頼表現において法の選択に違いが見られた。学習者は直説法を用いるのに対して母語話者はより丁寧度が上がる条件法を使用していた。語のスタイルの選択についても、2 グループ間に違いが見られた。学習者は形式的な語を使用するのに対して、母語話者は語の省略 (football⇔foot) や俗語の使用 (université⇔fac) といった非形式的な語を用いていた。

また MWUs レベルでは、意味・統語的選択について違いが見られた。母語話者は主語+VOULOIR の後ろに名詞句を用いて「～をください」という意味の中で VOULOIR を使用しするのに対して、学習者は主語+VOULOIR の後に動詞句を使用して「～がしたい」という意味内で VOULOIR を使用していた。なお、母語話者が同じ意味の文を生成する際には AIMER BIEN という MWU を使用していた。

過少使用語とのコロケーション分析では、人称代名詞主語の選択の差が見られた。母語話者は *on* を使用するのに対して、学習者は *nous* を使用していた。また、*donc voilà* のようなディスコースマーカの過少使用や、*à peu près* や *et tout* などの曖昧性付加句の過少使用が明らかになった。

#### 9.1.3.2. 自由会話コーパス

自由会話コーパスにおける過剰使用語とのコロケーション分析を行なったところ、学習者の発話は語や句の繰り返しに特徴付けられていることが分かった。また、発話スタイルの違いとして、学習者は形式的な使用である *oui* を使用し、母語話者は非形式的な *ouais* を多く使用するという傾向が挙げられた。

レベルの違いによる特徴の差も多く見られた。たとえば、初中級学習者は EST-CE QUE 型の疑問文を有意に多く使用するのに対して、上級学習者は母語話者と同様にイントネーション型の疑問文を使用する傾向にあった。また否定辞についても、初中級学習者は上級学習者や母語話者とは異なり、否定辞 *ne* を保持する傾向にあることが分かった。コミュニケーション方略としての MWUs の使用では、初級学習者は日本語や英語などのフランス語以外の語を使用していた。これは、FW の頻度だけでなく、*en japonais* の使用頻度が高かったことから説明できた。一方、上級学習者は単語がうまく出てこない場合や言葉につまった際に *comment dire* という MWU を使用して、沈黙を回避していた。

過少使用語とのコロケーション分析では、関係代名詞を含む文や強調構文などの複文の使用がレベルを問わず学習者によって過少使用されていた。また、*et tout (ça)* のような曖昧性付加句の過少使用も学習者グループに見られた。そして、ディスコースマーカの過少使用も学習者のレベルを問わず顕著であった。前言の補足や訂正機能を持つ *enfin, en fait*、発話の終了を示す *quoi*、発話のニュアンスを弱める *quand même, bon*、そして対話者に働きかける *hein, tu vois, tu sais* などのディスコースマーカを、学習者はほとんど用いていなかった。

#### 9.1.4. N-grams 分析

N-grams 分析ではまず、学習者は一部の N-grams を繰り返し使用しているのかについて数量的分析を行った。次に特徴的 N-grams を抽出し、これらの N-grams を構成する語の頻度層との関係や、特徴語との関係を明らかにした。最後に、特徴的 N-grams の質的な分析を行った。

##### 9.1.4.1. タスクに基づくコーパス

一部の N-grams が繰り返し使用されているのかについて、調整頻度 20 回以上の 2 語連鎖から 5 語連鎖までのタイプ数とトークン数を調べたところ、学習者はいずれも母語話者よりもこれらの数が有意に多かった。さらに、頻度上位に焦点を当てるため、頻度上位 10 位までのトークン数を比較したところ、そのほとんどについて母語話者よりも学習者の方がこれらの N-grams に依存していることが分かった。つまり、学習者は一部の N-grams を繰り返し使用していた。

次に、過剰使用 N-grams と過少使用 N-grams といった特徴的 N-grams について、それらの構成する語の頻度層について分析を行ったところ、過剰使用 N-grams については高頻度語(K1 機能語と K1 内容語)がほとんどを占めており、中頻度語(K2)や低頻度語(K3、Off-list)があったとしても、学習に困難をきたすものではないことが分かった。過少使用 N-grams についてはその全ての構成語が高頻度語であった。また、特徴的 N-grams と特徴語の関係を調べたところ、過剰使用 N-grams についてはその多くが特徴語を含むものであったが、過少使用 N-grams は過剰使用 N-grams ほど特徴語の影響を受けていないことが分かった。つまり、過少使用 N-grams は語彙の組み合わせによって引き起こされたものであった。そして、特徴的 N-grams の分類を行ったところ、過剰使用 N-grams は「自己紹介・第三者紹介」、「好き嫌い」、「時間・道のり」、「質問する」、「断る」といった場面と機能によって分類できる N-grams が多かったが、過少使用 N-grams は場面や機能よりも、語彙、統語、談話レベルによって分類できた。最後にこれらの特徴的 N-grams の要因について考察を行ったところ、WUs レベルとして *elle* や *je* などの人称代名詞主語の過剰使用、VOULOIR と POUVOIR の選択といった動詞の選択の差が一因として挙げられた。また、MWUs として *beaucoup de* の過剰使用、疑問文における型の選択の差、否定辞 *ne* の保持と脱落、*que* を含む複文の過少使用、*en fait* や *je sais pas* といったディスコースマーカの過少使用、副詞 *souvent* の位置のエラーが一因であることが分かった。

##### 9.1.4.2. 自由会話 コーパス

N-grams の一部が繰り返し使用されているのかについて分析を行ったところ、タスクに基づくコーパスと同じ結果となった。つまり、調整頻度 20 回以上の 2 語連鎖から 5 語連鎖のタイプ数もトークン数も学習者コーパスが母語話者コーパスを有意に上回っていた。次に、2 語連鎖から 5 語連鎖までの上位 10 位に絞ってより頻度の高い N-grams のトークン数を比較したところ、有意差が見られなかったものもあったが、最上位については学習者

が母語話者の調整頻度を有意に上回っていた。つまり、自由会話でも学習者は一部の N-grams を繰り返し使用していることが分かった。

次に特徴的 N-grams について、それらの構成する語の頻度層について分析を行ったところ、過剰使用語についてはほとんどが高頻度語であり、その他も *vin* や *visité*(いずれも K2)といったトピックに依存する語と、書き言葉において WUs レベルでは頻度が低い話し言葉において MWUs レベルで多く使用される *dommage*(K3)のみであった。過剰使用 N-grams については *oui* の非形式的表現である *ouais*(Off-list)以外全て高頻度語で構成されていた。そして、特徴的 N-grams を構成する語と特徴語の関連を調べたところ、そのほとんどが特徴語を含むものであった。最後に特徴的 N-grams の要因について考察を行ったところ、語彙レベルとして過剰使用語や過剰使用語の影響以外にも、語や句の繰り返しの多用、中性代名詞 *en* の過剰使用が原因の一部であることが分かった。また、統語レベルでは否定辞 *ne* の保持と脱落、疑問文における型の選択の差、*qui* や *que* を含む複文の過剰使用が一因と分かった。そして、談話レベルでは *d'abord* や *et après* のような順接関係を示す語の過剰使用、*par exemple* や *comment dire* などの方略的 N-grams の過剰使用、*c'est ça*、*c'est vrai*、*moi aussi* などの同意表現の過剰使用、そして *tu vois*、*tu sais*、*mais bon*、*quand même*、*en fait*、*je (ne) sais pas*、*et tout (ça)*といったディスコースマーカ―の過剰使用が特徴的 N-grams の一因であることが分かった。

#### 9.1.5. リサーチクエスチョンに対する回答

以上の結果をふまえて、リサーチクエスチョンに対する回答を簡潔にまとめると、以下の通りになる。

##### 1. 語彙の豊かさの分析

1) 母語話者は学習者よりも低頻度語の語彙の豊かさの指標が高いか。  
語彙の豊かさの指標として、各頻度層におけるギロー値が最も信頼性が高いものであるということが示唆できた。そして、この指標は特異性を持つ K1 機能語を除くすべての頻度層で母語話者のほうが学習者よりも高かった。つまり母語話者は、低頻度語だけでなく、K1 機能語以外のすべての頻度層で、学習者よりも語彙が豊かであるといえる。

2) 言語能力の高い学習者は、母語話者の語彙の豊かさに近い指標を持つか(自由会話コーパス)。

留学経験を持ち修士以上に在籍する学習者が、母語話者の語彙の豊かさに近い指標を持っていた。つまり、言語能力が高い学習者は母語話者の語彙の豊かさに近いということが分かった。

##### 2. 特徴語分析

1) 学習者の過剰使用語にはどのような特徴があるのか。

リサーチクエスチョン 4)にまとめる。

2) 学習者の過剰使用語にはどのような特徴があるのか。

リサーチクエスチョン 4)にまとめる。

3) 学習者の特徴語に、外国語運用能力の差が関係するのか(自由会話コーパス)。

初中級学習者は全体的に過少使用語として抽出された語の使用頻度が低だけでなく、これらの語を用いている人の割合も少ない。一方、上級学習者による過少使用語の使用頻度は母語話者と比べて低いものの、これらの語を少なくとも1度は用いている人の割合は初中級学習者よりも明らかに高かった。

4) タスクに基づくコーパスや自由会話に基づくコーパスなどのデータの差に関わらない一貫した特徴語が存在するのか。

コーパスの種類に関わらず、過剰使用語と過少使用語のほとんどは頻度 1000 位以内の高頻度語であることが分かった。また、過剰使用語には代名詞が多く含まれており、過少使用語は多義語が多いという点も共通していた。

### 3. コロケーション分析

1) 学習者の過剰使用語を含むコロケーションにはどのような特徴があるか。

リサーチクエスチョン 4)にまとめる。

2) 学習者の過少使用語を含むコロケーションにはどのような特徴があるか。

リサーチクエスチョン 4)にまとめる。

3) 学習者のコロケーションの差に、外国語運用能力の差が関係するのか(自由会話コーパス)。

上級者は、初中級学習者よりも非形式的な語・表現を使用する傾向にあり、この傾向は母語話者の発話スタイルに近かった。

4) タスクに基づくコーパスや自由会話に基づくコーパスなどのデータの差に関わらない一貫したコロケーションが存在するのか。

コーパスの種類に関わらず、学習者は人称代名詞 *j(e)*を含むコロケーションを過剰使用していた。一方、曖昧性付加句やディスコースマーカは、コーパスの種類やレベルに関わらず、過少使用コロケーションとして多く抽出された。

### 4. N-Grams 分析

1) 学習者は母語話者と比較して、一部の N-grams を繰り返し使用するか。

調整頻度 20 回以上の 2 語連鎖から 5 語連鎖までの N-grams について、タイプ数とトークン数を比較したところ、学習者は母語話者よりもこれらの数が有意に多かった。さらに、上位 10 位までの N-grams のトークン数を比較したところ、学習者の発話は頻度最上位の N-grams に依存していることが分かった。つまり、学習者は一部の N-grams を繰り返し使用しているといえる。

2) 特徴的 N-grams はどのような頻度層の語で構成されているか。

過剰使用 N-grams も過少使用 N-grams もそのほとんどが高頻度語(K1 機能語と K1 内容語)で構成されていた。

3) 特徴的 N-grams と特徴語にはどのような関係があるのか。

過剰使用 N-grams も過少使用 N-grams も特徴語の影響を多く受けていたが、過剰使用 N-grams にその傾向が顕著に見られた。

4) 特徴的 N-grams にはどのような傾向があるのか。

リサーチクエスチョン 5)にまとめる。

5) タスクに基づくコーパスや自由会話に基づくコーパスなどのデータの差に関わらない一貫した N-gram の特徴が存在するのか。

学習者は、一部の N-grams を繰り返し使用していた。特徴的 N-grams を構成する語は、そのほとんどが高頻度語である点も一貫した特徴であった。また、過少使用 N-grams として、ディスコースマーカ―の役割をもつ N-grams が共通して抽出された。

## 9.2. 今後の課題

本論文では、学習者コーパスを用いて学習者の口頭産出における使用語彙についての記述を行ってきたが、今後の課題として 3 点挙げられる。まず、分析面に関する点である。本研究は、学習者の言語使用のごく限られた側面のみを分析したものである。使用語彙研究に限定しても、本研究で扱うことが出来なかった点も多く存在する。たとえば、コーパス研究ではコーパスに現れなかった語や表現については、可能性でしか語る事が出来ないため、アンケートなど別の手法を用いて追従研究を行わなければならない。本研究は、コーパスのみに基づいた研究であるため、コーパスに含まれていない使用の側面については、もちろん分析されていない。また、分析を行ったコーパス自体も対象者の多くが東京外国語大学の学生であり、全ての日本人フランス語学習者を代表するデータとは言えないという点も付け加えなければならない。また、コーパスのサイズも小さく一般化するにはより多くのデータが必要となる。特に母語話者に近い超上級レベルの学習者データが不足している。

二点目は、本研究による分析結果や使用したコーパスの適用方法についてである。実際に英語教育の分野では、学習者コーパスの分析結果を基に教材開発が行われている。また、コーパス自体を教材として使用する方法が提案され、実行されている。本研究はフランス語学習者の使用語彙を記述するにとどまっているため、分析結果をどのように教育現場へ還元するかについては、今後の課題とする。

二点目と関連するが、三点目はコーパスを規範として考えるかどうかについての議論である。本分析結果にも見られたように、フランス語母語話者の実際の言語使用には、フランス語の規範から外れたものや、平俗な語や表現が多く見られた。Andersen (2004)は、これらを非フランス語母語話者が使用すると、外国語アクセントの有無に関わらず下品な印象を与える可能性があるため、これらの非標準的な使用は批判的に教育に取り入れるべきだと主張している。しかし、コーパスに見られる言語使用を規範と考えるのであれば、母語話者が行うような言語使用を学習者は学ぶべきである。Thomas (2008)も主張するように、非標準的な語や表現を学習することを促進すべきか、あるいは標準的なフランス語に限定して指導すべきかについては今後慎重に議論を行わなければならないと考える。

Mark(1998 :78)は、言語教育研究において主に行われてきたものとして、以下の 3 点を挙げている(Granger 2002 より引用) : 1)目標言語の記述 ; 2)動機、学習スタイル、ニーズ、学習態度などの学習者の性質の解明 ; 3)タスク、シラバス、カリキュラムなどの作成。しかし、学習者言語についてはほとんど注目を集めてこなかった。そして Mark (ibid)は、学習者言語を分析することによって、以上の 3 点についてさらに解明することが出来ると続けている。学習者コーパスに基づく本研究はまさにこの学習者言語に焦点を当てたも



のである。特に、日本人フランス語学習者による口頭産出について、体系的な分析はこれまで行われてきてこなかった。そのため、本研究によって新たに明らかになった点は、少なからずフランス語教育研究に示唆を与えるものであると言える。

## 参考文献

### 外国語の文献

- Aarts, J. 1998. "Introduction". In S. Johansson. and S. Oksefjell. (ed.) *Corpora and Cross-Linguistic Research*. Amsterdam-Atlanta. GA.
- Adolphs, S. and Schmitt, N. 2003. "Lexical Coverage of Spoken Discourse". *Applied Linguistics* 24 (4). pp. 425-438.
- , 2004. "Vocabulary Coverage according to Spoken Discourse Context". In P. Bogaards. and B. Laufer. (eds.). *Vocabulary in a Second Language- Selection, Acquisition, and Testing*. pp. 39-49.
- Altenberg, B. 1998. "On the Phraseology of Spoken English: the Evidence of Recurrent Word-Combinations". In A.P. Cowie. (ed.). *Phraseology. Theory, Analysis, and Applications*. Oxford: Oxford University Press. pp. 101-122.
- Altenberg, B. and Granger, S. 2001. "The Grammatical and Lexical Patterning of MAKE in Native and Non-native Student Writing". *Applied Linguistics* 22(2). pp. 173-194.
- Andersen, H. L. 2004. "Comment Utiliser les Connaissances sur le Français Parlé dans l'Enseignement du Français Langue Etrangère". In. H.L. Andersen and C. Thomsen. *Sept Approches à un Corpus. Analyses du Français Parlé*. Genève: Lang. pp. 185-212.
- , 2006. "L'Application du Cadre Européen Commun de Référence dans les Nouveaux Programmes de Français en Danemark". *Revue Française de Linguistique Appliquée* 11 (1). pp. 45-60.
- Anderson, R. C., and Freebody, P. 1983. "Reading Comprehension and the Assessment and Acquisition of Word Knowledge". In B. Hutton (ed.). *Advances in Reading/Language Research: a Research Annual*. Greenwich, CT: JAI Press. pp. 231-256.
- Aijmer, K. 2004. "Pragmatic Markers in Spoken Interlanguage". *Nordic Journal of English Studies*. pp. 173-190.
- Baker, P. 2004. "Querying Keywords: Questions of Difference, Frequency and Sense in Keywords Analysis". *Journal of English Linguistics* 32(4). pp. 346-359.
- Bauer, L. and Nation, I. S. P. 1993. "Word Families". *International Journal of Lexicography*. 6 (3). pp. 253-279.
- De Beaugrande, R. 2001. "Learge Corpora, Small Corpora, and the Learning of Language". In H. Ghadessy., A. Henry. and R. L. Roseberry. (eds.). *Small Corpus Studies and ELT- Theory and Practice*. Amsterdam: John Benjamins. pp. 3-28.
- Beeching, K. 1997. "French for Specific Purposes: the Case for Spoken Corpora". *Applied Linguistics* 18(3). pp. 374-394.

- Bert, M., Bruxelles, S., Étienne, C., Mondada, L. Teston, S., Traverso, V., Jouin-Chardon, É. and Lascar, J. 2010. "Grands Corpus et Linguistique Outillée pour l'Etude du Français en Interaction (plateforme CLAPI et corpus CIEL)". *Pratiques: Linguistique, Littérature, Didactique* 147/148. pp. 17–34.
- Biber, D. 2009. "A Corpus-Driven Approach to Formulaic Language in English: Multi-Word Patterns in Speech and Writing". *International Journal of Corpus Linguistics* 14(3). pp. 275-311.
- Biber, D. and Conrad, S. 1999. "Lexical Bundles in Conversation and Academic Prose". in H. Hasselgard and S. Oksefjell (eds.) *Out of Corpora: Studies in Honor of stir Johansson*. Amsterdam: Rodopi. pp. 181-189.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. 1999. *Longman grammar of spoken and written English*. Essex: Pearson Education.
- Biber, D. and Reppen, R. 2002. "What does Frequency Have to Do with Grammar Teaching?". *Studies in Second Language Acquisition* 24. pp. 199-208.
- Biber, D., Conrad, S. and Cortes, V. 2003. "Lexical Bundles in Speech and Writing: an Initial Taxonomy". In. A. Wilson, P. Rayson and T. McEnery (eds.) *Corpus Linguistics by the Lune: a Festschrift for Geoffrey Leech*. Frankfurt: Peter Lang. pp. 71–93.
- Biber, D., Conrad, S., and Cortes, V. 2004. "If You Look at...: Lexical Bundles in University Teaching and Textbooks". *Applied Linguistic* 25(3). pp. 371-405.
- Blanche-Benveniste C. 2005. "Le Corpus Français Parlé du GARS". *Tradizione e Innovazione. Il parlato : teoria, corpora, linguistica dei corpora. Atti del VI Convegno SILFI*. IN. E. Burr. (dir.), Florence: Franco Cesrai Editore. pp. 57-76.
- Bod, R., Hay, J., Jannedy S. (éds). 2003. *Probabilistic Linguistics*. Cambridge-London: The MIT Press. pp. 1-10.
- Bogaards, P. and Laufer, B. 2004. "Introduction". In P. Bogaards. and B. Laufer. (eds.). *Vocabulary in a Second Language- Selection, Acquisition, and Testing*. pp. vii-xiv.
- Bolinger, D. 1976. "Meaning and Memory". *Forum Linguisticum* 1. pp. 1-14.
- Breyton, M. and Fraisse, P. 1959. "Comparaison entre les Langues Oral et Ecrit". *l'Année Psychologique* 59. pp. 61-71.
- Bruxelles, S., L. Mondada , A.-C. Simon and V. Traverso. 2009. "Présentation du numéro". In. S. Bruxelles., L. Mondada., A.-C. Simon, and V. Traverso. (éds),. "Grands Corpus de Français Parlé : Bilan Historique et Perspectives de Recherche", Cahiers deLinguistique de Louvain 33(2). pp. 1-14.
- Bulté, B., Housen, A., Pierrard, M. and van Daele, S. 2008. "Investigating Lexical Proficiency Development over Time- the Case of Dutch Speaking Learners of French in Brussels". *French Language Studies* 18. pp. 277-298.
- Burns, A. and Seidlhofer, B. 2002. "Speaking and Pronunciation". In. N. Schmitt. (ed.). *An Introduction to Applied Linguistics*. London: Arnold. pp. 211-232.

- Cappeau, P. and Seijido, M. 2005. Les Corpus Oraux en Français (inventaire 2005 v. 1.0).  
([http://www.dglflf.culture.gouv.fr/recherche/corpus\\_parole/Presentation\\_Inventaire.pdf](http://www.dglflf.culture.gouv.fr/recherche/corpus_parole/Presentation_Inventaire.pdf))
- Carroll, J. B. 1964. *Language and Thought*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Cavalla C., Crozier E., Dumarest D. and Richou C. 2009. *le Vocabulaire en Classe de Langue*. Paris: Clé International.
- Čermák, F. 2009. “Spoken Corpora Design: Their Constitutive Parameters”. *International Journal of Corpus Linguistics* 14(1). pp. 113-123.
- Channell, J. 1994. *Vague Language*. Oxford: Oxford University Press.
- Chomsky, N. 1957. *Syntactic structures*. Hague: Mouton.
- De Cock, S. 2004. “Preferred Sequences of Words in NS and NNS speech”. *Belgian Journal of English Language and Literatures* 2004. pp. 225-246.
- , 2007. “Routinized Building Blocks in Native Speaker and Learner Speech : Clausal Sequences in the Spotlight”. In M. Campoy. and C. Luzón (eds.) *Spoken Corpora in Applied Linguistics*. pp. 217-233.
- De Cock, S., Granger, S., Leech, G. and McEnery, T. 1998. An Automated Approach to the Phrasicon of EFL Learners. In. S. Granger. (ed.). *Learner English on Computer*. London and New York: Longman. pp. 67-79.
- Cowie, A. P. 1998. *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.
- Crossley, S. A., Salsbury, T., McNamara, D. S. and Jarvis, S. 2011. “Predicting Lexical Proficiency in Language Learner Texts Using Computational Indices”. *Language Testing*. pp.561-580.
- Cobb, T. and Horst, M. 2004. “Is There Room for an Academic Word List in French?”. In P. Bogaards. and B. Laufer, B. (eds.) *Vocabulary in a Second Language – Selection, Acquisition, and Testing*. Amsterdam: John Benjamins. pp. 15-38.
- Conrad, S. 2000. “Will Corpus Linguistic Revolutionize Grammar Teaching in the 21st Century?”. *TESOL Quarterly* 34(3). pp. 548-560.
- Cori, M. and David, S. 2008. “Les Corpus Fondent-ils une Nouvelle Linguistique? “. *Langage* 171. pp. 111-129.
- Dalbera, J. P. 2002. “Le Corpus entre Données, Analyse et Théorie”. *Corpus et Recherche linguistiques* 1. pp. 89-104.
- Daller, H., van Hout, R. and Treffers-Daller, J. 2003. “Lexical Richness in the Spontaneous Speech of Bilinguals”. *Applied Linguistics* 24 (2). pp. 197-222.
- David, A. 2008. “A Developmental Perspective on Productive Lexical Knowledge in L2 Oral Interlanguage”. *Journal of French Language Studies* 18(3). pp. 315-331.
- Debrock, M., Flament-Boistrancourt, D. and Gevaert, R. 1999. “Le Manque de 'Naturel' des Interactions Verbales du Non-Francophone en Français. Analyse

- de Quelques Aspects à partir du Corpus LANCOM”. *Faits de Langue* 13. pp. 46-56.
- Detey, S., Durand, J. Laks, B. and Lyche, C. (eds.) 2010. *Les Variétés du Français Parlé dans l’Espace Francophone. Ressources pour l’Enseignement*. Paris: Ophrys.
- Dewaele, J. M. 1993. “Extraversion et Richesse Lexicale dans Deux Styles d’Interlangue Française”. *International Review of Applied Linguistics* 100, pp. 87–105.
- Dister, A., Francard, M., Hambye, P. and Simon, A.C. 2009. “Du Corpus à la Banque de Données. Du Son, des Textes et des Métadonnées. L’Evolution de Banque de Données Textuelles Orales VALIBEL (1989-2009)”. *Cahiers de Linguistique* 33(2). pp. 113-129.
- Ellis, N. 1997. “Vocabulary Acquisition: Word Structure, Collocation, Word-Class, and Meaning”. In N. Schmitt. and M. McCarthy (eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press. pp. 122-139.
- Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford. Oxford University Press.
- Equipe DELIC, 2004. “Présentation du Corpus de Référence du Français Parlé”. *Recherches sur le Français parlé* 18. pp. 11-42.
- Faucett, L. and Maki, I. 1932. *A Study of English-word Values Statistically Determined from the Latest Extensive Word Counts*. Tokyo: Matsumura Sanshodo.
- Faucett, L., H. E. Palmer, M. West, and E. L. Thorndike. 1936. *Interim Report on Vocabulary Selection*. London: PS King.
- Firth, J.R. 1957. “Modes of Meaning”. In J. R, Firth. *Papers in Linguistics 1934-1951*. London: Oxford University Press. pp. 190-215.
- Flamment-Boistrancourt, D. 2001. “Pragmatique et Approche Communicative: la Contribution du Corpus LANCOM”. *Le Français dans le Monde Recherches et Applications, numéro spécial: “Théories Linguistiques et Enseignement du Français aux Non-Francophones”*.  
([https://www.kuleuven-kulak.be/~dboistra/doc/LFDLM\\_RA.pdf](https://www.kuleuven-kulak.be/~dboistra/doc/LFDLM_RA.pdf))
- Fraisse, P. and Breyton, M. 1959. “Comparaisons entre les Langages Oral et Ecrit”. *L’année Psychologique* 1. pp.61-71.
- Fung, L. and R. Carter. 2007. “Discourse Markers and Spoken English: Native and Learner Use in Pedagogic Settings”. *Applied Linguistics* 28(3). pp. 410-429.
- Gadet, F. 1996. “Une Distinction Bien Fragile : Oral/Ecrit”. *TRANEL* 25. pp. 13-27.
- , 2012. “Un Grand Corpus de Français Parlé: le CIEL-F- Choix Epistémologiques et Réalisations Empiriques”. *Revue Française de Linguistique Appliquée* 13. pp. 39-54.

- Gilquin, G. 2008. "Hesitation Markers among EFL learners: Pragmatic Deficiency or Difference?". In: Romero-Trillo J. (ed.) *Pragmatics and Corpus Linguistics: A Mutualistic Entente*. Berlin, Heidelberg, New York: Mouton de Gruyter. pp. 119-149.
- Gilquin, G., De Cock, S. and Granger, S. 2010. *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Granfeldt, J., Nugues, P., Agren, M., Thulin, J. Persson, E. and Schlyter, S. 2006. "CEFLE and Direkt Profil: A New Computer Learner Corpus in French L2 and a System for Grammatical Profiling." In. *Proceedings of the 5th International Conference on Language Resources and Evaluation*. pp. 565-570.
- Goodfellow, R., Lamy, MN. and Jones, G. 2002. "Assessing Learners' Writing Using Lexical Frequency". *ReCALL* 14(1). pp. 133-145.
- Granger, S. 1998. "The Computer Learner Corpus: a Versatile New Source of Data for SLA Research". In. Granger, S. (ed.) *Learner English on Computer*. London and New York: Longman. pp. 3-18.
- , 2002. "A Bird's-Eye View of Computer Learner Corpus Research". In. S. Granger., J. Hung. and S. Petch-Tyson. (eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam and Philadelphia: Benjamins. pp. 3-33.
- , 2003. "The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research". *TEASOL Quarterly* 37(3). pp. 538-546.
- , 2004. "Computer Learner Corpus Research: Current Status and Future Prospects". In U. Connor. and T. A. Upton. *Applied Corpus Linguistics, a Multidimensional Perspective*. Amsterdam: Rodopi. pp. 123-145.
- , 2005. "Pushing Back the Limits of Phraseology. How Far Can We Go?". In. C. Cosme., C. Gouverneur., F. Meunier. and M. Paquot. (eds.). *Proceedings of the Phraseology 2005 Conference, Louvain-La-Neuve, 13-15 October 2005*. pp. 165-168.
- ([http://dial.academielouvain.be/vital/access/services/Download/boreal:75667/DOC\\_01](http://dial.academielouvain.be/vital/access/services/Download/boreal:75667/DOC_01))
- Granger, S. and Rayson, P. 1998. "Automatic Profiling of Learner Texts". In S. Granger. (ed.) *Learner English on Computer*. pp. 119-131.
- Granger, S. and Meunier, F. 2008. "Phraseology in Language Learning and Teaching: Where to from Here?". In. F. Meunier. and S. Granger. (eds.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam and Philadelphia: Benjamins: pp. 247-252.
- Granger S. and Paquot M. 2008. "Disentangling the Phraseological Web." In. S. Granger. and F. Meunier. (eds). *Phraseology: An interdisciplinary Perspective*. Amsterdam: John Benjamins. pp. 27-49.

- Grant, L. and Bauer, L. 2004. "Criteria for Re-Defining Idioms: Are We Barking up the Wrong Tree?". *Applied Linguistics* 25. pp. 38-61.
- Grevisse, M. 1986. *Le Bon Usage –Grammaire Française–*. 12e édition. Bruxelles: de Boeck Duculot.
- Hansen, M-B. M. 2005. "From Prepositional Phrase to Hesitation Marker : the Semantic and Pragmatic Evolution of French 'enfin'". *Journal of Historical Pragmatics* 6(1). pp.37-68.
- Herbst, T. 1996. "What are Collocations: Sandy Beaches or False Teeth". *English studies* 4. pp. 379-393.
- Hinkel, E. 2006. "Current Perspectives on Teaching the Four Skills". *TESOL Quarterly* 40(1). pp. 109-131.
- Hirsh, D. and Nation, I. S. P. 1992. "What Vocabulary Size is Needed to Read Unsimplified Texts for Pleasure?". *Reading in a Foreign Language* 8. pp. 689–696.
- Horn, E. 1926. *A Basic Writing Vocabulary -10,000 Words Most Commonly Used in Writing*. College of Education.
- Howatt, A. P. R. and Widdowson, H. G. 2004. *A History of English Language Teaching*. Oxford: Oxford University Press.
- Howes, D. 1966. "A Word Count of Spoken English". *Journal of Verbal Learning and Verbal Behavior* 5(6). pp. 572-604.
- Imbs, P. 1978. *Trésor de la Langue Française. Tome Sixième*. Paris: Éditions du Centre National de la Recherches Scientifiques.
- Ishikawa, S. 2007. "A Corpus-based Study on the Vocabulary of English Speech Presentations by Japanese Learners of English and English Native Speakers". *Conference Proceedings of the 6th Chulalongkorn University Language Institute's International Conference*. pp. 9-18.  
([http://language.sakura.ne.jp/s/ilaa/ishikawa\\_20071218.pdf](http://language.sakura.ne.jp/s/ilaa/ishikawa_20071218.pdf))
- Jacques, M. P. 2005. Pourquoi une linguistique de corpus? In. G. Williams (eds.) *La Linguistique De Corpus*. Rennes. Presses Universitaires de Rennes. pp. 21-30.
- Järvinen, T. 2003. "Bank of English and Beyond". *TreeBanks, Text, Speech and Language Technology* 20. pp. 43-59.
- Johansson, S., Leech, G. N., and Goodluck, H. 1978. *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Department of English, University of Oslo.
- Johansson, V. 2008. *Lexical Diversity and Lexical Density in Speech and Writing: a Developmental Perspective*. Lund University, Department of Linguistics and Phonetics. pp. 61-79.
- Jones, G. 2001. *Compiling French word Frequency Lists for the Vocabulary Assessment Tool: A Feasibility Study*.  
([http://www.er.uqam.ca/nobel/r21270/cgi-bin/F\\_webfreqs/glynn\\_jones.html](http://www.er.uqam.ca/nobel/r21270/cgi-bin/F_webfreqs/glynn_jones.html))

- Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. London: Longman.
- Kjellmer, G. 1991. "A Mint of Phrases", in K. Aijmer and B. Altenberg (eds.) *English Corpus Linguistics*. London and New York: Longman. pp. 111-127.
- Kjellmer, G. 1994. *A Dictionary of English Collocations*. Vols.3. Oxford: Clarendon Press.
- Krishnamurthy, R. 1987. "The Process of Compilation". In J. Sinclair. (ed.). *Looking Up. An Account of the COBUILD Project in Lexical Computing*. London and Glasgow: Collins ELT. pp. 62-85.
- Lambrecht, K. 1981. *Topic, Antitopic and Verb Agreement in Non-Standard French*. Amsterdam: John Benjamins.
- Laufer, B. 1998. "The Development of Passive and Active Vocabulary in a Second Language: Same or Different?". *Applied linguistics* 19(2). pp. 255-271.
- , 2005a. "Focus on Form in Second Language Vocabulary Learning". In S. Foster-Cohen., M. del Pilar García Mayo. and J. Cenoz. (eds.). *EUROSLA Yearbook 5*. Amsterdam: John Benjamins. pp. 223-251.
- , 2005b. "Lexical Frequency Profiles: From Monte Carlo to the Real World. A Response to Meara". *Applied Linguistics* 26. pp. 581-587.
- Laufer, B. and Nation, I.S.P. 1995. "Lexical Richness in L2 Written Production: Can It be Measured?". *Applied Linguistics* 16(3). pp. 307-322.
- Laufer, B. and Goldstein, Z. 2004. "Testing Vocabulary Knowledge: Size, Strength and Computer Adaptiveness". *Language Learning* 54 (3). pp. 399-436.
- Leech, G. 1991. "The State of the Art in Corpus Linguistics". In K. Aijmer. and B. Altenberg (eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman. pp. 8-29.
- , 1998. "Preface". In S. Granger (ed.). *Learner English on Computer*. London and New York: Longman. pp. xiv-xx.
- , 2011. "Frequency, Corpora and Language Learning". In F. Meunier., S. De Cock., G. Guilquin. and M. Paquot. (eds.) *A Taste for Corpora: in Honour of Sylviane Granger*. Amsterdam, John Benjamins. pp. 7-32.
- Leech, G. and Fligelstone, S. 1992. "Computer and Corpus Analysis". In C. S. Butler (ed.). *Computers and Written Texts*. Oxford: Blackwall. pp. 115-140.
- Leech, G., Rayson, P., and Wilson, A. 2001. *Word Frequencies in Written and Spoken English : Based on the British National Corpus*. Harlow: Longman.
- Léon, J. 2008. "Aux Sources de la «Corpus Linguistics»: Firth et la London School". In M. Cori, S. David, et J. Léon (eds.), *Construction des Faits en Linguistique: la Place des Corpus*. *Langage* 171. pp. 12-33.
- Lewis, M. 1993. *The Lexical Approach*. Hove: Language Teaching Publications.
- Lorenzo-Dus, N. and Meara, P. 2005. "Examiner Support Strategies and Test-Taker Vocabulary". *International Review of Applied Linguistics in Language Teaching* 43. pp. 239-258.



- Malvern, D. and Richards, B. 1997. "A New Measure of Lexical Diversity". In A. Ryan. and A. Wray. (eds.). *Evolving Models of Language*. Clevedon: Multilingual Matters. pp. 58-71.
- Malvern, D., Richards, B., Chipere, N., and Duran, P. 2004. *Lexical Diversity and Language Development: Quantification and Assessment*. Hampshire: Palgrave Macmillan.
- Mark, K. L. 1998. "The Significance of Learner Corpus Data in Relation to the Problems of Language Teaching". *Bulletin of General Education* 312. pp. 77-90.
- Marsden, E., Myles, F., Rule, S. and Mitchell, R. 2002. "Oral French Interlanguage Corpora: Tools for Data Management and Analysis". *Centre for Language in Education Occasional Papers* 58. University of Southampton. pp. 3-15.
- Marushima, N., Detey, S., Kawaguchi, Y. 2010. "Caractéristiques Phonétiques des Voyelles Orales Arrondies du Français chez des Apprenants Japonophones". *Flambeau* 36. Tokyo: Université des Langues Étrangères de Tokyo. pp. 53-72.
- McCarthy, M. J. and Carter, R. 1997. "Written and Spoken Vocabulary". In N. Schmitt. and M. McCarthy (eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press. pp. 20-39.
- McCarthy, M. J. and O'Keeffe, A. 2004. "Research in the Teaching of Speaking". *Annual Review of Applied Linguistics* 24. pp. 26-43.
- McEnery, T., Xiao, R., and Tono, Y. 2006. *Corpus-Based Language Studies: An Advanced Book*. New York. Routledge.
- Meara, P. 2005. "Lexical Frequency Profiles: A Monte Carlo Analysis". *Applied Linguistics* 26(1). pp.32-47.
- Mertens, P. 2002. "Les Corpus de Français Parlé ELICOP: Consultation et Exploitation". in J. Binon., P. Desmet., J. Elen., P. Mertens. and L. Sercu. (eds) *Tableaux Vivants. Opstellen over Taal-en-Onderwijs Aangeboden aan Mark Debrock*. Leuven: Universitaire Pers.  
(<http://membres.multimania.fr/anatex/articles/ELICOP.pdf>)
- Miller, G.A. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information". *Psychological Review* 63. pp. 81-97.
- Milton, J. 2006. "Language Lite? Learning French Vocabulary in School". *French Language Studies* 16. pp. 187-205.
- , 2009. *Measuring Second Language Vocabulary Acquisition*. Clevedon: Multilingual matters.
- Moeschler, J. et De Spengler, N. 1981. "Quand Même. De la Concession à la Réfutation", *Cahiers de Linguistique Française* 2. pp. 93-112.
- Moon, R. 1998. *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Oxford: Oxford University Press.
- Myles, F. 2003. "The Early Development of L2 Narratives: A Longitudinal Study". *Marges Linguistiques* 5. pp. 40-55.

- , 2005a. "Interlanguage Corpora and Second Language Acquisition Research". *Second Language Research* 21(4). pp. 373-391.
- , 2005b. "The Emergence of Morphosyntactic Structure in French L2". In J. M. Dewaele. (ed.). *Focus On French as a Foreign Language: Multidisciplinary Approaches. The Modern Language Journal* 91(1). pp. 88-113.
- Myles, F., Hooper, J., and Mitchell, R. 1998. "Rote or Rule? Exploring the Role of Formulaic Language in Classroom Foreign Language Learning". *Language Learning* 48(3). pp. 323-363.
- Myles, F., Mitchell, R., And Hooper, J. 1999. "Interrogative Chunks in French L2: A Basis for Creative Construction?". *Studies in Second Language Acquisition* 21(1). pp. 49-80.
- Nation, P. 2001. "Using Small Corpora to Investigate Learner Needs- Two Vocabulary Research Tools". In. H. Ghadessy., A. Henry. and R. L. Roseberry. (eds.). *Small Corpus Studies and ELT- Theory and Practice*. Amsterdam: John Benjamins. pp. 31-45.
- Nation, P. and Warning, R. 1997. "Vocabulary Size, Text Coverage, and Word Lists". In. N. Schmitt. and M. McCarthy (eds.) *Vocabulary, Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.
- Nation, P. and Meara, P. 2002. "Vocabulary". In. N. Schmitt. (ed.) *An Introduction to Applied Linguistics*. London: Arnold. pp. 35-54.
- Nesselhauf, N. 2004. "What are Collocations?". In. D. J. Allerton., N. Nesselhauf and P. Skandera (eds.). *Phraseological Units: Basic Concepts and their Application*. Basel: Schwabe Verlag. pp. 1-21.
- , 2005. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Norricks, N. R. 2008. "Using Large Corpora of Conversation to Investigate Narrative: The Case of Interjections in Conversational Storytelling Performance". *International Journal of Corpus Linguistics* 13 (4). pp. 438-464.
- Ovtcharov, V., Cobb, T. and Halter, R. 2006. "La Richesse Lexicale des Productions Orales: Mesure Fiable du Niveau de Compétence Langagière". *The Canadian Modern Language Review*. pp. 107-125.
- Palmer, H.E. and Hornby, A.S. 1937. *Thousand-Word English*. London: George Harrap.
- Pawley, A. and Syder, F.H. 1983. "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency". In. J. C. Richards and R. W. Schmidt. (eds.) *Language and communication*. London: Longman.
- Peters, A. 1983. *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- , 1985. "Language Segmentation: Operating Principles for the Perception and Analysis of Language". In. D. I. Slobin (ed.). *The cross-linguistic study of language acquisition* 2. Hillsdale, NJ: Lawrence Erlbaum. pp. 1029-1067.

- Racine, I., Detey, S., Zay, F. and Kawaguchi, Y. 2012. "Des Atouts d'un Corpus Multitâches pour l'Étude de la Phonologie en L2 : l'Exemple du Projet « Interphonologie Du Français Contemporain » (IPFC)". In A. Kamber et C. Skupiens (eds). *Recherches Récentes en FLE*. Berne : Peter Lang. pp. 1-19.
- Racine, I., Detey, S., Bühler, N., Schwab, S., Zay, F. and Kawaguchi, Y. 2010. "The Production of French Nasal Vowels by Advanced Japanese and Spanish Learners of French: A Corpus-Based Evaluation Study". In K. Dziubalska-Kolaczyk, M. Wrembel, and M. Kul. (eds.). *Proceedings of New Sounds 2010 - Sixth International Symposium on the Acquisition of Second Language Speech* [CD-ROM]. Poznan: Adam Mickiewicz University. pp. 367-372.
- Read, J. 1997. "Vocabulary and Testing". In N. Schmitt, and M. McCarthy (eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press. pp. 303-320.
- , 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- , 2007. Second Language Vocabulary Assessment: Current Practices and New Directions. *International Journal of English Studies* 7 (2). pp. 105-125.
- Richards, J., Platt, J. and H. Weber. 1985. *Longman Dictionary of Applied Linguistics*. Harlow: Longman.
- Ringbom, H. 1998. "Vocabulary Frequencies in Advanced Learner English: A Cross-Linguistic Approach". In S. Granger (ed.). *Learner English on computer*. London: Longman. pp. 41-52.
- Rule, S., Marsden, E., Myles, F. and Mitchell, R. 2003. "Constructing a Database of French Interlanguage Oral Corpora". In D. Archer., P. Rayson., E. Wilson. and T. McEnery. (eds.). *Proceedings of the Corpus Linguistics 2003 Conference, UCREL Technical Papers* 16. pp. 669-677.
- Rule, S. 2004. "French Interlanguage Oral Corpora: Recent Developments". In. F. Myles. and R. Towell. (eds.) *The Acquisition of French as a Second Language. Special issue of the Journal of French Language Studies* 14(3). pp. 343-356.
- Rule, S. and Marsden, E. 2006. "The Acquisition of Functional Categories in Early French L2 Grammars: The Use of Finite and Non-Finite Verbs in Negative Contexts". *Second Language Research* 22(2). pp. 188-218.
- Rundell, M. 2002. "Good Old-fashioned Lexicography: Human Judgement and the Limits of Automation". In. M. H. Corréard. (ed.). *Lexicography and Natural Language Processing. Festschrift in Honour of B.T.S. Atkins*. Euralex. pp. 138-156.
- Schmid, H. J. 2003. "Collocation: Hard to Pin Down, but Bloody Useful". *Zeitschrift für Anglistik und Amerikanistik* 51(3): 235-258.
- Schmitt, N. and Redwood, S. 2011. "Learner Knowledge of Phrasal Verbs -a Corpus-Informed Study". In. F. Meunier., S. De Cock., G. Guilquin. and M. Paquot.

- (eds.) *A Taste for Corpora: in Honour of Sylviane Granger*. Amsterdam: John Benjamins. pp. 173-208.
- Scott, M. 1996. *WordSmith Tools*, Oxford: Oxford University Press. ISBN 0-19-458984-6.
- , 1997. "PC Analysis of Key Words –and Key Key Words". *System* 25 (2). pp. 233-245.
- , 2001. "Comparing Corpora and Identifying Key Words, Collocations, Frequency Distributions through the Wordsmith Tools Suite of Computer Programs". In H. Ghadessy., A. Henry. and R. L. Roseberry. (eds.). *Small Corpus Studies and ELT- Theory and Practice*. Amsterdam: John Benjamins. pp. 47-67.
- Scott, M. and Tribble, C. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: Benjamins.
- Séguy, J. 1973. "Les Atlas Linguistiques de la France par Régions". *Langue Française* 18. pp. 65-90.
- Selinker, L. 1969. "Language Transfer". *General Linguistics* 9. pp. 67-92.
- , 1972. "Interlanguage". *International Journal of Applied Linguistics* 10. pp. 209-31.
- Shastri, S.V. 1988. "The Kolhapur Corpus of Indian English and Work Done on Its Basis So Far". *ICAME Journal* 12. pp. 15-26.
- Sinclair, J. M. 1996. "Preliminary recommendations on corpus typology". *EAGLES*.
- , 2001. "Preface". In. H. Ghadessy., A. Henry. and R. L. Roseberry. (eds.). *Small Corpus Studies and ELT- Theory and Practice*. Amsterdam: John Benjamins. pp. vii-xv.
- Stubbs, M. 1995. "Collocations and Semantic Profiles: on the Cause of the Trouble with Quantitative Studies". *Functions of Language* 2(1). pp. 23-55.  
(<http://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1995-cause-trouble.pdf>)
- , 2002. Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics* 7(2). pp. 215-244.
- Stubbs, M. and Barth, I. 2003. "Using Recurrent Phrases as Text-Type Discriminators: A Quantitative Method and Some Findings". *Functions of Language* 10(1). pp. 61–104.
- Sugiyama, K. 2012. "Lexical Profile of French Learner Speech: The Case of Japanese University Students". In Y. Tono. et al. (eds.) *Developmental and Crosslinguistic Perspectives in Learner Corpus Research. Series: Tokyo University of Foreign Studies Studies in Linguistics (TUFS SL)* 4. Amsterdam: Benjamins. pp. 279-298.
- Thomas, A. 2008. "La Mesure des Progrès Lexicaux en FL2 Avancé". In J. Durand, B. Habert, and B. Laks (eds.) *Congrès Mondial de Linguistique Française-CMLF08*. pp. 587-597.

- Thorndike, E. L. 1921. *The Teacher's Word Book*. New York: Teachers College.
- . 1932. *Teacher's Word Book of 20,000 Words*. Columbia: Teachers College.
- Thorndike, E. L. and Lorge, I. 1944. *A Teacher's Word Book of 30,000 Words*. New York. Teachers College. Columbus University.
- Tidball, F. and Treffers-Daller, J. 2007. "Exploring Measures of Vocabulary Richness in Semi-Spontaneous French Speech- A Quest for The Holy Grail?". In H. Daller, J. Milton and J. Treffers-Daller (eds.) *Modelling and Assessing Vocabulary knowledge*. Cambridge: Cambridge University Press. pp. 133-149.
- . 2008. "Analysing Lexical Richness in French Learner Language: What Frequency Lists and Teacher Judgements Can Tell Us about Basic and Advanced Words". *French Language Studies* 18. pp. 299-313.
- Tono, Y. 2003. "Learner Corpora: Design, Development and Applications". *Proceedings Of The Corpus Linguistics 2003 Conference*. Lancaster. United Kingdom. pp. 800-809.
- Treffers-Daller, J. 2009. "Language Dominance and Lexical Diversity: How Bilinguals and L2 Learners Differ in their Knowledge and Use of French Lexical and Functional Items". In B. Richards., M. H. Daller., D. D. Malvern. P. Meara., J. Milton. and J. Treffers-Daller. (eds.). *Vocabulary Studies in First and Second Language Acquisition -The Interface Between Theory and Application*. pp. 74-90.
- . 2010. "Operationalizing and Measuring Language Dominance". *International Journal of Bilingualism* 15. pp. 1-17.
- Treffers-Daller, J., Daller, H. M., Malvern, D., Richards, B., Meara, P. and Milton, J. 2008. "Introduction: Special Issue on Knowledge and Use of the Lexicon in French as a Second Language". *French Language Studies* 18. pp. 269-276.
- van Hout, R. and Vermeer, A. 2007. "Comparing Measures of Lexical Richness". In H. Daller, J. Milton and J. Treffers-Daller (eds.) *Modelling and Assessing Vocabulary knowledge*. Cambridge: Cambridge University Press. pp. 93-115.
- Verlinde, S. and Selva. T. 2001. "Corpus-Based Versus Intuition Based Lexicography: Defining a Word List for a French Learners' Dictionary". *Proceedings of the Corpus Linguistics 2001 conference*. Lancaster University. pp. 594-598.
- Vermeer, A. 2001. "Breadth and Depth of Vocabulary in Relation to L1/L2 Acquisition and Frequency of Input". *Applied Psycholinguistics* 22. pp. 217-234.
- . 2004. "The Relation Between Lexical Richness and Vocabulary Size". In P. Bogaards & B. Laufer (eds.) *Vocabulary in a Second Language*. Amsterdam: John Benjamins. pp. 173-189.
- Véronique, G. D. 2008. "A l'Intersection de l'Analyse des Productions d'Apprenants en Français Langue Etrangère, de l'Evaluation de leur Compétences et de la Programmation de l'Enseignement : Regards sur un Domaine de la Didactique

- du Français”. In J. Durand., B. Habert. and B. Laks. (eds.). *Congrès Mondial de Linguistique Française- CMLF08*. pp. 621-630.
- Viberg, A. 1996. “Cross-Linguistic Lexicology. The Case of English Go and Swedish ga”. in K. Aijmer, B. Altenberg and M. Johansson (eds.). *Languages in Contrast*. Lund: Lund University Press. pp. 151-182.
- Weinert, R. 1995. “The Role of Formulaic Language in Second Language Acquisition: a Review”. *Applied Linguistics* 16(2). pp. 180-205.
- West, M. 1927. *The New Method Readers (New Series)*. Bombay and Calcutta: London: Longmans, Green.
- , 1953. *A General Service List of English Words*. London: Longman.
- Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge. Cambridge University Press.
- Yu, G. 2010. “Lexical Diversity in Writing and Speaking Task Performances”. *Applied Linguistics* 31. pp. 236 - 259.

#### 日本語の文献

- 石川慎一郎(2008)『英語コーパスと言語教育』,大修館書店.
- (2012)『ベーシックコーパス言語学』,ひつじ書房.
- 和泉絵美,内元清貴,伊佐原均(2004)『日本人 1200 人の英語スピーキングコーパス』, アルク.
- 上村隆一(1997)「データベースで調べる」,『日本語学 16 (11)』, 明治書院. pp. 60-68.
- 鎌田 修,山内 博之(1999)『KY コーパス』. Ver 1.1.  
([http://opi.jp/shiryo/ky\\_corp.html](http://opi.jp/shiryo/ky_corp.html))
- 近藤野里,川口裕司(2009)「IPFC と中間言語としての現代フランス語研究」,『ふらんぼー 34』,東京外国語大学, pp.51-67.
- 鈴木綾乃,海野多枝(2006)「学習者言語コーパスに基づいた教材開発の可能性—日本語コロケーション習得を目的とした教材開発の試み—」,根岸雅史,海野多枝,吉富朝子(編)『言語情報学研究報告 14 TUFSS 言語モジュール』,21 世紀 COE プログラム「言語運用を基盤とする言語情報学拠点」,東京外国語大学大学院地域文化研究科. pp. 45-66.
- 杉山香織(2011a)「フランス語初級学習者の話し言葉における使用語彙の複雑さ—母語話者との比較から—」,『外国語教育研究 14』, pp. 20-34.
- (2011b)「フランス語初級学習者の発話における使用語彙の特徴」,『ロマンス語研究 44』, pp. 51-60.
- 投野由紀夫(編)(2007)『日本人中高生一万人の英語コーパス—中高生が書く英文の実態とその分析』,小学館.
- 前川喜久雄(2006)「概説」,『報告書「日本語話し言葉コーパスの構築法」』. 国立国語研究所. pp. 1-19.

#### ウェブサイト(日本語サイト)

- KY コーパス : [http://opi.jp/shiryo/ky\\_corp.html](http://opi.jp/shiryo/ky_corp.html)
- KOTONOHA 『現代日本語書き言葉均衡コーパス』 少納言(国立国語研究所) : <http://www.kotonoha.gr.jp/shonagon>
- KOTONOHA 『現代日本語書き言葉均衡コーパス』 中納言(国立国語研究所) : <https://chunagon.ninjal.ac.jp/login>
- 日本語話し言葉コーパス(国立国語研究所) : <http://www.ninjal.ac.jp/csi>
- 日本語学習者言語コーパス(東京外国語大学) : <http://cblle.tufs.ac.jp/lc/ja/index.php?menulang=ja>
- 日本語学習者による日本語作文と、その母語訳との対訳データベース(国立国語研究所) : [http://www.ninjal.ac.jp/publication/catalogue/kokken\\_mado/04/05/](http://www.ninjal.ac.jp/publication/catalogue/kokken_mado/04/05/)

#### ウェブサイト(外国語 サイト)

- ELICOP : <http://bach.arts.kuleuven.be/elicop/>
- Etude Linguistique de la Langue Parlée (ELILAP) : <http://bach.arts.kuleuven.be/elicop/projetELILAP.htm>
- French Learner Language Oral Corpora (FLLOC) : <http://www.flloc.soton.ac.uk>
- FRANTEXT : <http://www.frantext.fr/>
- International Corpus of Learner English (ICLE) : <http://www.uclouvain.be/en-cecl-icle.html>
- LANgue et COMmunication (LANCOM) : <http://bach.arts.kuleuven.be/elicop/ProjetLANCOM.htm>
- Louvain International Database of Spoken English Interlanguage (LINDSEI) : <http://www.uclouvain.be/en-cecl-lindsei.html>
- Phonologie du français contemporain : usages, variétés et structure(PFC) <http://www.projet-pfc.net/>
- VAriété LIngustiques du français en BELgique (VALIBEL) <http://www.uclouvain.be/81834.html>

## 謝辞

本研究を進めるにあたり、大変多くの方々に御世話になりました。ここに深く感謝の意を表します。

研究活動全般にわたり格別なる御指導と御高配を賜りました東京外国語大学川口裕司先生に甚大なる謝意を表します。修士課程入学以前から約 10 年間という長きにわたりご指導を賜りました。研究の基礎、研究発表の方法や研究論文の書き方の基礎から懇切丁寧に指導していただきました。心より感謝申し上げます。

副指導教員として、本研究に対して貴重な御指導と御助言を頂いた東京外国語大学根岸雅史先生および同大学海野多枝先生に心より感謝いたします。また、審査委員を引き受け、本論文を精査してくださった東京外国語大学投野由紀夫先生および神戸大学石川慎一郎先生に深く謝意を表します。

続いて、フランス語母語話者の意見が必要な際に、いつも快く質問に答えてくださった Rémy Prulon さんに感謝いたします。

留学生活においては、幸いにも多数の友人たちとの出会いに恵まれ、たくさんの刺激を受けました。特に森内悠佳子さん、萩原香織さん、Sébastien Lugan さん、古賀健太郎さん、新保未来さん、板垣祐子さんには大変お世話になりました。また、ボルドー第三大学の Shuaijun Mallet さん、Antoine Pascaud さんの支え無しには、博士論文を完成することが出来ませんでした。博士論文の提出という明確な目標を持った 2 人とともに同じ研究室の中で博士論文を執筆することができたことは幸いなことでありました。誠にありがとうございました。そして、快適な研究環境を提供してくださったボルドー第三大学 Laurence Labrune 先生をはじめとする CLLE-ERSSàB (Cognition, Langue, Langages, Ergonomie - Equipe de Recherche en Syntaxe et en Sémantique à Bordeaux) の先生方に感謝いたします。

また、筆者が所属する川口裕司研究室の学生の皆様には、学内研究会などで数々のコメントをいただき、論文の問題点などを指摘していただきました。心より御礼を申し上げます。

そして、本論文は様々な経済支援を受けて完成されたものでもあります。まず、東京外国語大学で実施していたグローバル COE プログラム「コーパスに基づく言語学教育研究拠点」の御支援のおかげで、様々な発表機会を得ることができ、博士論文の下地を築くことができました。また、コーパス研究の見識を深めることができたのは、ベルギーのルーバンカトリック大学への留学のおかげです。この留学は、ロータリー財団の御支援により実現したものです。そして、フランス政府および、東京外国語大学「若手研究者国際ナショナル・トレーニング・プログラム(ITP-EUROPA)」の御支援により、フランスのボルドー第三大学へ留学することができ、博士論文を完成させることができました。ここに謝意を表します。

最後になりましたが、博士課程まで進学する機会を与えてくれただけでなく、合計 5 年間半にもわたる三度の長期留学を実現させてくれ、ありとあらゆる場面で私を温かく見守り続けてくれた両親に深く感謝いたします。

ここに重ねて厚く謝意を表し、謝辞といたします。