

Aix 話し言葉コーパスプロジェクト

古賀 健太郎、秋廣 尚恵、川口 裕司

はじめに

本報告は、2005 年度より東京外国語大学が Aix-Marseille 第 1 大学(プロヴァンス大学)¹と共同で行ってきたフランス語話し言葉コーパス(以下 Aix コーパスとする)構築プロジェクトについてまとめたものである。

フランス語話し言葉に関する体系的な研究の歴史はまだ浅く、研究のために不可欠なコーパスの整備がようやく進み始めたばかりである。コーパスについては、書き言葉コーパスに比べて整備が大幅に遅れていると言わざるを得ない。こうした状況は、継続的かつ体系的な研究を進める上で問題である。実際、日本においてもフランス語話し言葉に関するこれまでの研究の数はとても限られている(東郷 1997, 前島 1997, 塩田 2001, 2003, 2005, 小川 2006 etc.)。したがって、話し言葉コーパスの整備は喫緊の課題といえる。Aix コーパス構築プロジェクトはこうしたコーパス整備の必要性に応え、日本をはじめ各国におけるフランス語話し言葉研究の発展に寄与するという目的のもと進められているのである。

以下ではフランス語話し言葉に関する研究の歴史を概観した上で、Aix コーパスプロジェクトについて、その内容とコーパスの使用実績を報告する。

1. フランス語話し言葉研究のあゆみ

音声言語研究のためには記録された音声データの存在が不可欠である。1930 年代のテープレコーダー誕生によって音声を記録し保存・蓄積することが格段に容易になった結果、言語学の諸分野でも録音した音声言語を分析する試みが始まった。しかし、その分野は音韻論(MARTINET 1945, MARTINET & WALTER 1973, DELAIS-ROUSSARIE & DURAND 2003, etc.)や社会言語学(MARCELLESI 1971, GADET 1989, COVENEY 2002 etc.)、あるいは個人語の記述(FRANÇOIS 1974)や方言研究などに限られ、音声言語そのものの構造に関する研究はなかなか現れなかつ

¹ 2012 年 1 月に Aix-Marseille 大学に改称、Aix-Marseille I, II, III の合併による総合大学となった。本報告では便宜上、Aix-Marseille 大学と表記することにする。

た。音声言語の統語構造に注目した本格的な研究は、Aix-Marseille 大学で 1976 年に結成された研究チーム GARS(Groupe Aixois de Recherches en Syntaxe)による研究(BLANCHE-BENVENISTE & JEANJEAN 1987 etc.)が実質的に初めてであったといえるだろう。そうした流れを経たのち、1990 年代以降のコンピュータの普及や録音技術の向上に伴い、以前とは比較にならないほどの大量で質の高い音声データをコーパスとして蓄積することが可能になった²。これによって話し言葉研究のための研究材料がようやく整ってきたといえる。しかし、コーパス化に必要な転写の作業には膨大な時間がかかるため、書き言葉コーパスの整備規模に比べて、その量は未だ充分とはいえないのが現状である。また、話し言葉には声という個人情報が含まれるということもあり、コーパス所有者の使用許可を取らずに自由にアクセスできるコーパス³はかなり限られるというのも、書き言葉コーパスとは異なる事情である。

ところで、転写テキストを元の音声データに忠実でありながら、同時に分析の際に障害にならないようなものにするためには綿密な検討が必要である。コーパスを充実させ多分野で広く使えるようにすることが話し言葉研究の可能性を広げることにつながることから、この問題を野放しにすることはできない。さらにいえば、転写における問題点を検討することそれ自体が、結果的に書き言葉には見られない話し言葉の構造上の特徴を明らかにする作業につながるともいえる。以下ではこれまでの研究で指摘されてきた転写における問題点と、そこから見えてきた話し言葉の特性がどんなものであったかということについて概観する。

GARS による本格的な話し言葉研究の第一歩といえる BLANCHE-BENVENISTE & JEANJEAN (1987)では、これまでにさまざまな研究で蓄積されてきた録音データの転写方法における問題点が詳細に指摘されている。それによると、転写の方法に共通の規則がなく、各研究で独自の転写規則を採用していることが、複数のコーパスを利用しようとした際の大きな障害になっているという。例えば音韻分析のために構築されたコーパスの中には(1)のように音韻現象を反映した表記(t'es, j'dis など)を採り入れているものがあるが、読みやすさの点で問題があるうえ、音韻以外の観点での研究(例えば語彙の頻度調査など)で転写文をそのまま使用することが難しく、汎用性に欠けてしまう。

² フランス語音声コーパス構築の現状についての詳細は CAPPEAU & SEIJIDO (2005)を参照。

³ 中には BLANCHE-BENVENISTE et al. (2002)や ANDERSEN & HANSEN (2000)のように、一般的な書籍と同じ形で(音声データの入ったメディア付)出版されているコーパスも存在する。

- (1) **t'es très bien dans ton costume t'es épatant et alors j'dis « et moi, regarde comment j'suis » et i(l) m'fait « c'est fou s'que tu ressembles à Zaza »**
君の衣装はとても似合ってる、君は素晴らしいね、私なんかは「見てこれ、どんな感じ？」って言っても「おかしいよ、ザザにそっくり」って言われるよ。
(Radio France コーパス⁴, p.11, ARL 010. BLANCHE-BENVENISTE & JEANJEAN 1987, p.133 から引用。太字化は筆者)

それ以外にも転写の際に問題となる現象はいくつかある。その代表的なものとして、ためらいや反復、そして訂正を挙げることができる。これらの現象は自然発話に特に多く見られるものである。書き言葉では通常、文章は完成した状態で受け手に提示されるため、話し手が文を構成し推敲する過程を受け手が見ることはない。その一方で自然発話では発話がその場で組み立てられていくため、そこで繰り広げられる推敲の過程は、発話された以上つぶさに転写テキストに記録しなければならない。しかしながら、これらのいわば冗長的な要素まで転写した(2)のようなテキストから発話の全体的な構成がどうなっているかを把握するのは必ずしも容易ではない。

- (2) Il fallait avoir le euh... ah zut... ...Ah! j'arrive plus à trouver le nom!... Le brevet, le, le diplôme là, de secouriste, le brevet de secouriste
持ってなきゃいけなかった、あの・・・ああもう・・・ああ、もう名前が出てこない！・・・免許証、証書だ、救急隊員の。救急隊員免許証
(Buttaf 91,5,5. BLANCHE-BENVENISTE 2000, p.18 から引用。)

さらに言うと、(2)の例では句読点等が加えられているが、これらの記号の付し方にも注意が必要である。なぜなら、文の解釈上必要となる句読点と、実際の発話で発生するポーズは必ずしも一致するとは限らないからである。例えば、(3)の発話では3つの文がポーズなしで連続しているのに対し、(4)では書き言葉においてであれば読点を打つことのない場所にポーズ(ハイフンで表記)が出現している⁵。

- (3) Venise est faite pour être vue Venise est une ville de coquetterie Venise est une ville qui se met en scène
ヴェニスは見られるためにつくられた ヴェニスは粋な街だ ヴェニスは自らを

⁴ ラジオ番組内の発話を転写したコーパス。教育的目的で1983年に構築された。

⁵ (3)、(4)共に GARS の転写規則にしたがい、句読点を用いていない。

演出する街だ

(Zay 12,12. BLANCHE-BENVENISTE 2000, p.32 から引用。訳は筆者)

- (4) certains Américains parlent – de l’Europe comme nous nous parlons de Venise

我々がヴェニスについて話すように、アメリカ人の中には(・・・)ヨーロッパについて話す人がいる

(Zay 17,5. BLANCHE-BENVENISTE 2000, p.33 から引用。訳は筆者)

句読点に関しては、(5)のような擬似分裂表現(*tournure pseudo-clivée*) や(6)のような左方転位が起こった発話においても同様に問題となる。書き言葉ではこのような種類の文に対し、主文に前置された要素と主文との間に読点を付することがあるが、話し言葉では必ずしもその場所にポーズが置かれるとは限らない。擬似分裂表現と左方転位は話題化などと共に話し言葉の中でよく観察される現象であり、話し言葉における情報伝達の仕方が書き言葉のそれと異なることを示唆している (HALLIDAY 1985)。

- (5) ce que je voudrais c’est recommencer à travailler

したいと思うこと、それは仕事を再開することだ

(DorotII, 38,3, BLANCHE-BENVENISTE 2000, p.99 から引用。訳は筆者)

- (6) les femmes je leur fais pas confiance

女性については、僕は信用しない

(Blasco, 197, BLANCHE-BENVENISTE 2000, p.68 から引用。訳は筆者)

こうしたいくつかの点をふまえ GARS では転写規則として、あくまでも正書法にのっとった表記とし、句読点や引用符は用いず、そこに現れた特筆すべき音韻現象等については注記を付することによって表す方法を提案している (BLANCHE-BENVENISTE & JEANJEAN 1987, pp.179-180. 本稿表1も参照)。また、(2)で挙げた発話の構成に関する問題については、反復されたり訂正された要素を図1のような二次元的な配置にすることを提案している。これによって、発話内で起きた諸現象を容易に把握できるうえ、範列的關係と連辞的關係を整理できるため、統語分析の際に有効である (*ibid*, pp.165-171)。

Il fallait avoir le euh -

(ah zut - - ah j'arrive plus à trouver le nom)

le brevet

le

le diplôme là de secouriste

le brevet de secouriste

図 1 (2)の発話を範列的關係を基に配置した例
(BLANCHE-BENVENISTE 2000, p.18)⁶

転写において問題となる現象としてはもう一点、発話の重なり合いがある。複数の話者が同時に発言することがあるということ自体が話し言葉の特徴のひとつであるが、それだけでなく、発話が重なったときに話者がとる対応（発話を中断する、重なってしまった部分を言いなおす、など）にも話し言葉ならではの特徴が表れているといえる。したがって、転写テキストでは発話の重なり合いが起こった部分に印を施しておくことが望ましいだろう。(7)の例のように、GARS の転写規則では発話の重なり合いは下線で示してある。

(7) Loc 1 ouais mais – dix kilos tu vas cacher dans les chaussures

ああ、だけど・・・10 キロ、君は靴に隠すんだ

Loc 2 je te demande

君に聞くけど

Loc 1 tu peux pas

だめだ

Loc 2 comment ils font passer – parce que vous avez vu combien co-

どうやってあいつらは通って・・・だって見ただろどれだけ・・・

Loc 1 peut-être des fois ils s'entendent eh

おそらく奴らはグルってこともあるってことだろ

Loc 2 combien combien combien ça passe là comment

どれだけ、どれだけ、どれだけ通り抜けたよ？どうやって

Loc 1 ouais – ouah

ああ・・・そうさ

(Belle de Mai 75,9. BLANCHE-BENVENISTE 2000, p.32 から引用。訳は筆者)

⁶ 図中にあるハイフンは、ここに短いポーズがあったことを表している。

このように、転写という作業を通して、話し言葉は書き言葉のそれとは幾分異なる構造上の特徴を有していることが分かってきたのである。ためらいや反復、あるいは(7)で見られた発話の中断等の現象は、書き言葉において定義されてきた文や節といった統語的単位を話し言葉にそのまま適用することが難しいことを示唆するものといえる⁷。それは従来考えられてきたような、「書き言葉＝正しい文法」対「話し言葉＝くだけた文法」(BLANCHE-BENVENISTE & JEANJEAN 1987, p.20)という対立に疑問を投げかけるものであった。文法というものはむしろこの両者があって成り立っているものであり、したがって、これまでの書き言葉中心の研究に話し言葉研究を「組み込むこと(l'intégration)」(ibid, p.86)が、言語システムの全容を解明する上で不可欠な作業となる。そして、そのためには話し言葉コーパスのより一層の充実が求められるのである。

2. 「フランス語話し言葉コーパス」構築共同プロジェクト

フランス語話し言葉研究の発展に資するという目的のもと、東京外国語大学では Aix-Marseille 大学と共同でフランス語話し言葉コーパス構築のプロジェクトを 2005 年より行っている。このプロジェクトは 21 世紀 COE「言語運用を基盤とする言語情報学拠点」(2005 年度実施分)及びグローバル COE「コーパスに基づく言語学教育研究拠点」(2009 年度・2011 年度実施分)の枠組みの中で実施されたものである。

まず 21 世紀 COE のもとで行われた調査(2005 年度)について説明する。この調査は Aix-Marseille 大学文学部の協力のもと 2005 年 7 月にフランスのエクス・アン・プロヴァンスにて行われたもので⁸、フランス語母語話者同士の自然会話を録音し転写するという内容の調査である。録音は大学の講義室や商店の中などで行われ、インフォーマントは大学生や大学教員、書籍店や旅行代理店の店員などである。会話数は合計 21 個、転写した総語数 10 万語以上、時間は合計約 7 時間となった。会話には 2 つのタイプがあり、1 つ目は自由な主題による会話(普段の仕事について、大学での勉強について、家族や友人の近況、ヴァカンスの計画など)、もう 1 つは場面と役

⁷ この点に関し BLANCHE-BENVENISTE (2000, pp.111-123)では、話し言葉の構造は文や節といった概念では説明しきれず、したがってそういった枠組みを超えた“Macro-syntaxe”の領域を想定する必要があると指摘している。

⁸ 2005 年 7 月 1 日～17 日に実施。21 世紀 COE の調査チームは、佐野敦至(福島大学教授)をリーダーとし、中田俊介(本学大学院博士後期課程)、小川敬洋(同博士前期課程)が協力者であった。秋廣尚恵(高等実習研究院博士課程)は現地協力者として参加し、Aix-Marseille 大学からは André Valli 教授と José Deulofeu 教授の協力を得た。

柄を設定した即興寸劇(大家と借主の交渉、銀行での口座開設など)である。転写については Aix-Marseille 大学の話し言葉研究チーム GARS の転写規則に基づいたうえで、同大学の学生によって行われた。コーパスは 21 世紀 COE プログラムの成果として公開されており、その一部については同プログラムのウェブページ上で、実際の会話を試聴したり、その転写を閲覧したりすることができる⁹。

次に、グローバル COE のもとで行われた調査(2009 年度・2011 年度)についてである。この調査は 2005 年実施分同様、Aix-Marseille 大学文学部の協力のもとエクサン・プロヴァンスにおいて、2010 年 2 月～3 月¹⁰及び 2011 年 9 月¹¹の 2 回にわたって行われた。調査の内容は 2005 年度と同様に、フランス語母語話者による自然会話の録音と転写である。録音はいずれの回も大学の講義室や個人宅などで行われ、インフォーマントは大学生や大学教員、児童関連施設の職員などである。会話数は 2009 年度分が 34 個、2011 年度分が 32 個、転写については現在行っている最中であるが、2011 年 12 月現在で 484,232 語の転写が完了している。合計録音時間は 2009 年度分、2011 年度分共に 30 時間を超えている¹²。会話の内容は専ら自由な主題によるもので、具体的にはヴァカンスの思い出、学生生活、仕事の内容についてなどであった。転写については専門の業者に委託し、転写規則はフランス語と同様にグローバル COE のもと進行中の他言語(トルコ語、ロシア語等)話し言葉コーパス構築プロジェクトのとはほぼ共通したものを採用している。

2009 年度及び 2011 年度で採用したこの転写規則と、2005 年度の調査で適用した上掲の GARS 転写規則の間にはいくつかの相違点がある。これについてまとめたのが以下の表 1 である。この表から、同一の現象についてそれぞれ異なった記号を用いている場合(項目 1～6)や、一方の転写規則でマークすることになっている現象や項目がもう一方ではマークされていないケース(項目 7～12, 13～18)があることが確認

⁹ 21 世紀 COE「フランス語(エクス)多言語話しことばコーパス」のウェブページ:

http://www.coelang.tufts.ac.jp/multilingual_corpus/fr/index.html

¹⁰ 2010 年 2 月 20 日～3 月 4 日に実施。グローバル COE の調査チームは、佐野敦至(福島大学教授)をリーダーとし、中田俊介(本学大学院博士後期課程)、古賀健太郎(同博士前期課程)が協力者であった。秋廣尚恵(Aix-Marseille 大学講師)は現地協力者として参加し、Aix-Marseille 大学からは André Valli 教授、José Deulofeu 教授、Frédéric Sabio 准教授の協力を得た。

¹¹ 2011 年 9 月 11 日～9 月 17 日に実施。グローバル COE の調査チームは、川口裕司(本学大学院教授)をリーダーとし、杉山香織(本学大学院博士後期課程、Bordeaux 大学博士課程)、松澤水戸(本学大学院博士後期課程)、古賀健太郎(同博士前期課程、Bordeaux 大学修士課程)が協力者であった。秋廣尚恵(Aix-Marseille 大学講師)は現地協力者として参加し、Aix-Marseille 大学からは José Deulofeu 教授、Frédéric Sabio 准教授の協力を得た。

¹² これは会話開始前・終了後しばらくの空白や雑音なども全て含んだ場合の合計時間である。

できる。この点については、今後両者の転写規則を統一し、それに合わせて転写文も修正することで対応する予定である。両者それぞれに利点と不足している点があるため、具体的にどのように折衷していくかは今後の検討課題である。

項目	転写方法	
	GARS (2005 年度)	2009・2011 年度
1. 短いポーズ	+	#
2. 長いポーズ	++	## (より長い場合 : ###)
3. 母音の引き延ばし	quoi:, eh:	Quooooi, Eeeeh
4. 不完全な語、言いかけ	mer- mercredi	mer(.) mercredi
5. 発話の重なり合い ¹³	下線_____	< >
6. 言語ではない音(笑い、咳など)	{sourire}	[sourire]
7. 省略された要素	正書法に則った表記が原則。必要に応じて注釈付記。	a(r)bres, (je) sais pas
8. ためらいの際に出す euh	特定の記号の指定なし。	e (長さに応じて e, ee, eee とする)
9. 疑問符および感嘆符があると思われる個所	記号なし。 例 : ah bon	? または ! を付する 例 : ah bon ?
10. 唇を閉じながら、承認したり、賛成したり、頷いたりするときの声	該当記号なし	m (長さに応じて mm, mmm とする)
11. 転位した要素	該当記号なし	Il vous va très bien, {ce pantalon}.
12. 強調アクセント	該当記号なし。必要に応じて注釈付記。	下線_____
13. Schwa の添加	donc°, sur°	該当記号なし

¹³重なりあいの記号については、2005 年度(GARS)版では、双方の話者の発話該当部分に対して付すようになっているのに対し、2009・2011 年度版では割り込んできた側の発話に対してのみ付すようになっている、という相違点もある。これについては(8)、(9)の転写例も参照。

14. 長時間の中断	///	該当記号なし
15. 複数の転写可能性(同音)	il(s) travaille(nt)	該当記号なし
16. 複数の転写可能性 (判別が困難な場合)	/j'ai été, j'étais/	該当記号なし
17. その語が発音されたかどうかははっきりしない場合	/d'accord, ø/	該当記号なし
18. リエゾンの欠如のうち、特筆すべきもの	vous # allez	該当記号なし
19. 完全なかたちで聞き取れなかった部分	X (音節数に応じて XX, XXX とする)	X (音節数に応じて XX, XXX とする)
20. インフォーマントの表記法	番号で表記 例:L1, L2 あるいは fr01, fr02 など	名・姓の頭文字＋続 き番号 例: Jean Dubois = JD1, JD2, ... , JDn

表1 2005 年度版(GARS)転写規則と、2009 年度・2011 年度版転写規則の比較

(8) 2005 年度版(GARS)転写規則の使用例

fr01 – bonjour

fr02 – bonjour {sourire} ++

fr01 – je pars en voyage d'affaire euh à Mexico je dois être là-bas lundi

fr02 – bon je vais regarder + mais je vous XX

fr01 – je ne peux pas je ne peux pas partir avant samedi

fr02 – mais je vous préviens que ça va être assez difficile

fr01 – hum /d'accord, ø/ ++

fr02 – il n'y a que trois vols Air France par semaine

fr01 – que- {raclement de gorge} quels jours

fr02 – le lundi le mar- pardon euh + le lundi le mercredi et le vendredi

fr01 – la compagnie ça m'est complètement égal

fr02 – Air Mexique ou Trans-XXXXX {bruit}

fr01 – euh non il n'y a pas d'autres vols

fr02 – apparemment + non

(9) 2009・2011 年度版転写規則の使用例

AM1 – Bonjour

DH1 – Bonjour [sourire] ###

AM2 – Je pars en voyage d'affaire ee à Mexico. Je dois être là-bas lun(di)

DH2 – Bon, je vais regarder #

AM3 – Je ne peux pas <mais je vous xx>, je ne peux pas partir avant samedi

DH3 – Mais je vous préviens que ça va être assez difficile

AM4 – mm xx ##

DH4 – Il n'y a que trois vols Air France par semaine

AM5 – Que(.) [raclement de gorge] quels jours ?

DH5 – Le lundi, le mar(.) pardon, euh # le lundi, le mercredi et le vendredi

AM6 – {La compagnie}, ça m'est complètement égal

DH6 – Air Mexique ou Trans-xxxxx [bruit]

AM7 – eeee Non. (Il n'y) a pas d'autres vols ?

DH7 – Apparemment, # non

なお、これら合計 3 回にわたる調査に加え、Aix-Marseille 大学の Frédéric Sabio 氏より、同氏が 2005 年から 2009 年にわたって収集したフランス語母語話者の会話音声資料の提供を受け、これについても 2010 年 6 月に文字化を行った。録音は個人宅等で行われたもので、インフォーマントは大学生やその家族、会話数は 13 個で時間は合計約 4 時間である。会話の内容は旅行や家族についてなどであり、中にはインタビュー形式のものも含まれている。

3. Aix コーパスを利用した研究

2005 年のプロジェクト開始から現在まで継続して構築されてきたこの Aix コーパスは、話し言葉の音韻的分析はもちろんのこと、統語論や学習者言語に関する研究など、すでにいくつかの分野で利用されている。そして今後もよりさまざまな分野の研究で使用されることが期待される。ここではこれまでに Aix コーパスを利用して行われた研究やプロジェクトについて簡単に説明する。

3. 1. リエゾン研究

音韻論・統語論の分野においては近藤(2010)が、フランス語の日常会話におけ

るリエゾンがどのような条件のもとで実現されているかについて、2005 年に収集された Aix コーパスを使用して分析を行っている。その結果、リエゾンが実現され得る音韻的コンテキストの中には、必ずリエゾンが起こる連辞と、そうでない連辞(選択的リエゾン)があるが、話し言葉では必ずリエゾンが起こる連辞以外ではほとんどリエゾンが実現されない傾向にあることが分かった。さらに、ÅGREN(1973)のラジオコーパスや ENCREVÉ(1988)の政治家演説コーパスとの比較を行い、話し言葉におけるリエゾン実現率がラジオでの発話や政治家の演説におけるそれよりも低いことが観察されたことから、リエゾンの実現率は統語的要素以外の要因によっても左右され得ることも指摘している。

リエゾンについてはさらに、社会言語学的な要因に注目した研究においても Aix コーパスが利用されている。近藤(2012 出版予定)では 2005 年の Aix コーパスにおいて自然会話と即興劇会話の2種類の異なる会話種が収録されていることを利用し、インフォーマントの年齢や性別、発話の状況といった社会言語学的要因がリエゾン出現率にどのような影響を及ぼしているかについて、統計的手法を用いて分析している。その結果、自然会話において性別が及ぼすリエゾン実現率への影響は観察されなかった一方、年齢については選択的リエゾンの実現率との関連を示唆する結果が見られた。また発話の状況についても、自然発話において選択的リエゾン実現率が低い若者でも、丁寧さが要求される状況(職業従事者役を演じる時など)では選択的リエゾンを行う傾向にあることから、少なくとも若者の選択的リエゾン実現率に影響を及ぼす要因とみなすことができると指摘している。

3. 2. 統語論研究

統語論においては秋廣(2006)が、直接目的語の省略が話し言葉においてどのように実現されているかについて、照応的な場合と非照応的な場合の大きく2つに分類した上で、2005 年に収集された Aix コーパスを基に考察している。その結果、話し言葉において高頻度に現れる現象のひとつとして従来考えられてきた(10)の *j'ai lu* (文脈照応的コンテキスト)や(11)の *je remplis* (現場指示的コンテキスト)のような場合における直接目的語省略の出現率については、書き言葉と比較してもさほど大きな差が観察されなかった。一般的傾向として、書き言葉にせよ、話し言葉にせよ照応的目的語省略は少ないことが指摘されている。またこのような省略の例が話し言葉で出現する理由としては、丁寧か、ぞんざいかといった文体の違いによるものであること(ぞんざいな文体で往々にして省略が観察される)、また近接する文脈内に同じ「動詞+目的語」が繰り返し使われていること(11)が考えられる。

(10) L1 : Laaba je lui ai écrit sa préface.

ラーバには前書きを書いてやったんだ。

L2 : ouais ouais je sais j'ai **lu**.

うん、うん、知っている。読んだよ。

(chrabi, 60,3. 秋廣 2006, p.185.)

(11) L2 : vous avez juste à vous mettre dans un coin remplir tout ça

どこかに座ってこれを書き込んで下さい。

L1 : d'accord je **remplis**

分かりました。書きましょう。

(Aix コーパス 2005. 秋廣 2006, p.188.)

秋廣(2006)によると、Aix コーパスの分析によって明らかになった話し言葉に特によく見られる照応的目的語省略の現象としてはむしろ、(12)のような「挿入句」的な用法や、(13)のような遊離構文が挙げられるという。Aix コーパスにおいては、「挿入句」的用法¹⁴は croire, dire, écouter, espérer, imaginer, savoir, trouver, voir, vouloir の9動詞で観察された。また、遊離構文は adorer, aimer, connaître, détester, préférer などのいわゆる「評価動詞」とともに観察された。(秋廣 2006, pp.191-192)

(12) tu as aussi les photos **je crois**

君は写真も持っていると思うんだが

(Aix コーパス 2005. 秋廣 2006, p.192.)

(13) l'ambiance + **j'aime** enfin

雰囲気は好きだね。

(Aix コーパス 2005. 秋廣 2006, p.192.)

3. 3. 学習者言語研究

Aix コーパスは母語話者による会話コーパスであるが、学習者言語の分析においても使用することができる。杉山(2011)では、日本語を母語とするフランス語学習

¹⁴ 挿入句的動詞用法は、発語的動詞としての機能を持つことが指摘されている。BLANCHE-BENVENISTE(1989) BLANCHE-BENVENISTE & WILLEMS (2007) は、この用法を *verbes de recteur faible* と名づけ、本来の他動詞として直接目的語を従える用法(*verbe plein*)とは区別して記述している。挿入句的動詞用法は最近再び注目を集めており、2012年5月には、パリ西ナンテール大学にて「Parenthétique 2012」という会議が企画されている。

者の会話について、フランス語母語話者の会話と比較し、使用語彙の複雑さという観点から見た両者の違いを統計的に分析している。この中で使用しているコーパスは、フランス語学習者話し言葉コーパスと、2005年の Aix コーパス、そして2006年にパリ第13大学の協力のもと構築されたフランス語話し言葉コーパス(以下 Paris コーパスとする)¹⁵の合計3つである。この研究では、延べ語数、異なり語数、そして語彙の多様性の指標である Guiraud 値の3点に基づいた学習者－母語話者間の比較をし、その結果、3点いずれの場合においても語彙の複雑さの点で両者の間に有意差が随所で確認された。同様に杉山(2012出版予定)も上で挙げた3つのコーパスを用いて、学習者の使用語彙に見られる傾向について統計的に分析している。それによると、学習者は母語話者に比べ、否定辞「ne」などを過剰使用する一方、疑問文においてイントネーション型をあまり用いなかったり、「c(ce)」や「ça」を過少使用する傾向にあったという。このことから、使用語彙の観点で学習者と母語話者の間に一定の開きが見られることが確認できたと報告している。また、学習者言語の多様性についても注目した杉山(2013出版予定)では、日本人フランス語学習者による自由会話コーパスと Aix コーパスの2つを用いた統計的分析を通して、学習者と母語話者との間に語彙の豊かさの点で有意差が存在することだけでなく、学習者間においても語彙使用の傾向にばらつきが見られること、さらにそのばらつきはいくつかのグループにまとめられることが指摘され、学習期間や留学経験などの指標によって語彙の豊かさにも違いが現れ得ることが示唆されている。

3. 4. 教材研究

Aix コーパスや Paris コーパスはフランス語教材の研究・開発のために応用することも可能である。馬場(2011)は、日本及びフランスにて出版されている CECR(ヨーロッパ共通参照枠)A1レベル相当のフランス語教科書のうち、ダイアログ形式で示された会話文における使用語彙の特徴を分析している。その中で、Paris コーパスを教科書会話文コーパスとの比較対象に用い、教科書の会話例と自然会話の間に使用語彙の点で違いがあるかどうか統計的分析を行っている。その結果、教科書の会話では二人称における vous や敬称、挨拶表現ならびに基本動詞を多用する点が特徴的であり、またその傾向は出版元(日本／フランス)に関わらず共通してみられることから、自然会話と比較した場合、教科書の会話文に現れる語彙には一定の共通した特徴

¹⁵ 2005年 Aix コーパスと同じく、東京外国語大学21世紀COEプログラムによっておこなわれたプロジェクトである。総語数10万語以上、時間は合計約8時間半。会話のテーマは主に音楽や文化についてである。調査者は本学の François Roussel 客員准教授であり、7月27日～8月29日にかけて実施した。

があることを指摘している。

3. 5. 品詞検索

最後に Aix コーパスの応用研究の例として、品詞検索エンジンについて紹介する。これは Aix コーパスと Paris コーパスのそれぞれの転写テキストの各語に Tree Tagger¹⁶を用いて品詞タグを付与し、検索できるようにしたものである。公開語数は 218,366 語、タグの種類は「略号」、「固有名詞」、「句読点」、「動詞過去分詞」など合計 41 種類である。タグ付けは東京外国語大学の学生が行い、タグの誤りについては修正を随時行っている。この品詞検索エンジンは 21 世紀 COE 拠点のウェブサイト上に公開され、誰でも自由に利用できるようになっている¹⁷。品詞、辞書形、出現形の 3 つから検索できるほか、Aix コーパスと Paris コーパスのどちらか一方だけを選択しての検索も可能となっている(図2)。

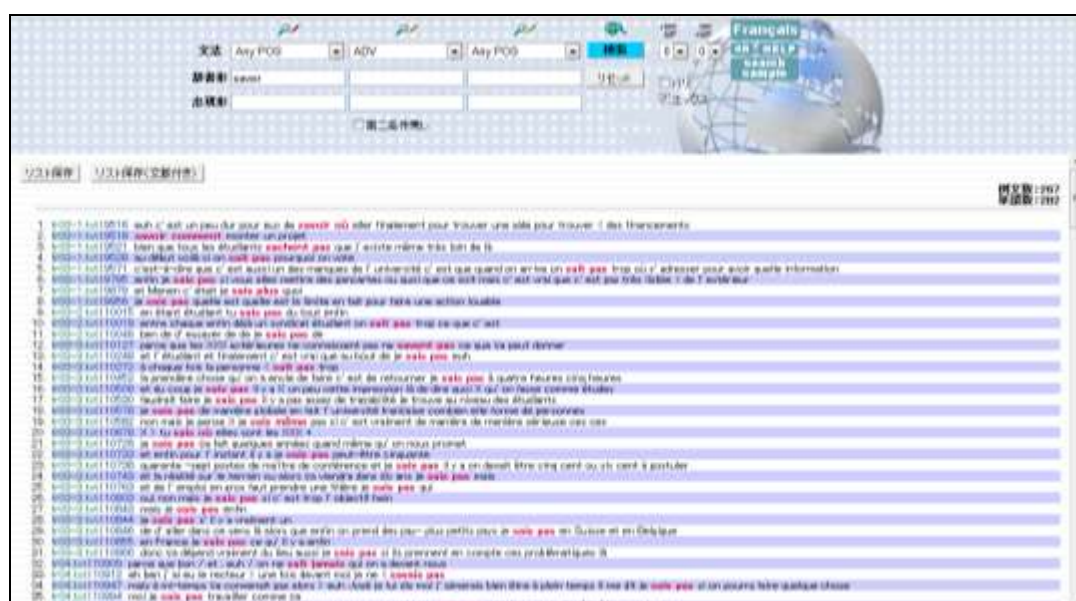


図2 品詞検索エンジンの画面例

(辞書形 « savoir » と文法「ADV(副詞)」の連続について、Aix コーパスのみから検索した場合。« je sais pas » 等、否定文における ne の省略が確認できる)

¹⁶ ドイツ Stuttgart 大学コンピュータ言語学チームの H. Schmid 氏らによって開発されたプログラムで、フランス語版のタグセットは Achim Stein (2003)が作成し、ウェブ上に公開されている。

<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/> を参照。

¹⁷ CbLLE 品詞検索エンジン: <http://cblle.tufts.ac.jp/tag/fr/index.php?menulang=ja>

おわりに

以上、Aix コーパスについて、プロジェクトの現状を報告した。今後は転写規則統一のための検討を行うと共に、転写が完了していない音声データの転写作業を進め、コーパスの質と量の両面の充実を目指すことになる。Aix コーパスが、フランス語話し言葉研究の今後のさらなる発展への一助となることを期待する。

参考文献

- ÅGREN, J.(1973). « Étude sur quelques liaisons facultatives dans le français de conversation radiophonique : fréquences et facteurs ». *Acta Universitatis Upsaliensis*, Studia Romanica Upsaliensia 10.
- AKIHIRO-TRESSEL, H. (2011). «Analyse contrastive entre le français et le japonais, la non-réalisation du complément». *Working Papers in Corpus-based Linguistics and Language Education* 7, Graduate School of Global Studies, Tokyo University of Foreign Studies, pp.71-89.
- ANDERSEN H.L. & HANSEN, A.B. (2000) *Le Français Parlé : Actes du colloque international Université de Copenhague du 29 au 30 octobre 1998*, Aarhus : Museum Tusculanum Press.
- BENZITOUN, C. (2006). «Quand et lorsque sont-ils synonymes ?». *Working Papers in Linguistic Informatics* 12, Graduate School of Area and Culture Studies, Tokyo University of Foreign Studies, pp.167-182.
- BILGER, M. (1982). « Contribution à l'analyse en grilles ». *Recherches sur le français parlé*, 4, pp.195-215.
- . (2000). *Corpus : méthodologie et applications linguistiques*, Paris : H. Champion.
- . (2008). *Données orales : Les enjeux de la transcription*, Perpignan : Presses Universitaires de Perpignan.
- BILGER, M., BLASCO, M., CAPPEAU, P., PALLAUD, B., SABIO, F., & SAVELLI, M.-J. (1997). « Transcription de l'oral et interprétation ; illustration de quelques difficultés ». *Recherches sur le français parlé*, 14, pp.57-86.
- BLANCHE-BENVENISTE, C. (1989). «Construction verbale 'en incise' et réaction faible des verbes». *Recherche sur le français parlé*. 9. pp. 53-74
- . (2000). *Approches de la langue parlée en français*, Paris : Ophrys.
- . (2007). « Linguistic Analysis of Spoken Language : The Case of French Language ». KAWAGUCHI, Y. et al. (eds.). *Spoken Language Corpus and Linguistic*

- Informatics*, John Benjamins, Amsterdam/Philadelphia, pp.35-66.
- BLANCHE-BENVENISTE, C. & JEANJEAN, C. (1987). *Le français parlé : transcription et édition*, Paris : Didier Érudition.
- BLANCHE-BENVENISTE, C., ROUGET, C. & SABIO, F. (2002). *Choix de textes de français parlé : 36 extraits*, Paris : Honoré Champion.
- BLANCHE-BENVENISTE, C. et WILLEMS, D. (2007). « Un nouveau regard sur les verbes faibles ». *Bulletin de la société linguistique de Paris*, 102/1, pp.217-154.
- CAPPEAU, P., SABIO, F., BILGER, M. CHANET, C. (2004). *Autour du corpus de référence du français parlé*, Aix-en-Provence : Publications de l'Université de Provence.
- CAPPEAU, P., SEIJIDO, M. (2005). Les corpus oraux en français (inventaire 2005 v.1.0), <http://www.dglf.culture.gouv.fr/recherche/corpus_parole/Presentation_Inventaire.pdf> consulté le 23 janvier 2012.
- COVENEY, A. (1996). *Variability in spoken French: a sociolinguistic study of interrogation and negation*, Exeter : Elm Bank Publications.
- DELAIS-ROUSSARIE, É. et DURAND, J. (2003). *Corpus et variation en phonologie du français : méthodes et analyses*, Toulouse : Presses Universitaire du Mirail.
- DELATTRE, P. (1966). *Studies in French and Comparative phonetics*, La Haye : Mouton.
- DEULOFEU, J. (1986). « Syntaxe de *que* en français parlé et le problème de la subordination ». *Recherches sur le français parlé*, 8, pp.79-104.
- (1992). « Variation syntaxique : recherche d'invariants et étude des attitudes des locuteurs devant la norme ». *Langages*, 108, pp.66-78.
- DEULOFEU, J. & BLANCHE-BENVENISTE, C.. (2007). «C-ORAL-ROM –French Corpus–», KAWAGUCHI, Y. et al. (eds.). *Spoken Language Corpus and Linguistic Informatics*, John Benjamins, Amsterdam/Philadelphia, pp.181-198.
- ENCREVÉ, P. (1988). « La liaison avec et sans enchaînement ». *Phonologie tridimensionnelle et usages du français*, Paris : Editions du Seuil.
- FRANÇOIS, D. (1974). *Français parlé: analyse des unités phoniques et significatives d'un corpus recueilli dans la région parisienne*, Paris : S.E.L.A.F.
- GADET, F. (1989). *Le français ordinaire*, Paris : Colin.
- HABERT, B. (2005). *Instruments et ressources électroniques pour le français*, Paris : Ophrys.
- HALLIDAY, M.A.K. (1985). *Spoken and written language*, Oxford : Oxford University Press.
- KAWAGUCHI, Y., T. TAKAGAKI, N. TOMIMORI & Y. TSURUGA. (2007). *Corpus-Based Perspectives in Linguistics*, Amsterdam/Philadelphia : John benjamins.
- KAWAGUCHI, Y., S. ZAIMA & T. TAKAGAKI. (2006). *Spoken Language Corpus and Linguistic Informatics*, Amsterdam/Philadelphia : John Benjamins.
- LABRIE, V. (1982). *Précis de transcription de documents d'archives orales*, Québec :

- Institut Québécois de Recherche sur la Culture, coll. « Instruments de travail », 4.
- MARCELLESI, J.-B. (1971). *Le congrès de Tours (Décembre 1920) : études sociolinguistiques*, Paris : Le Pavillon.
- MARTINET, A. (1945). *La prononciation du français contemporain*, Paris : Droz.
- MARTINET, A. & WALTER, H. (1973). *Dictionnaire de la prononciation française dans son usage réel*, Paris : France Expansion.
- SABIO, F. (2002). « L'opposition de modalité en français parlé: étude macro-syntaxique ». *Recherches sur le français parlé*, 17, pp.55-77.
- (2006). « L'antéposition des compléments dans le français contemporain : l'exemple des objets directs ». *Linguisticae Investigationes*, 29/1, pp.173-182.
- VALLI, A. (1999). « Remarques sur le français parlé de locuteurs de la région de Marseille : usage régional du français et "régionalisme" ». *Recherches sur le français parlé*, 15, pp.59-86.
- WILLEMS, D. (1997). « Histoire, linguistique et sources orales ». *Recherches sur le français parlé*, 14, pp.10-20.

- 秋廣尚恵. 2000. 「直接目的語の諸問題について—先行研究のまとめと今後の展望」. 『ふらんぼー』No.35-26, 東京外国語大学フランス語研究室, pp.107-125.
- . 2003. 「フランス語話し言葉における動詞 *aimer*」. 『ふらんぼー』No.32-33, 東京外国語大学フランス語研究室, pp.66-84.
- . 2005. 「直接目的語の現れない他動詞構文」. 『フランス語を語る』, 三修社, pp.88-102.
- . 2006. 「フランス語会話コーパスにおける直接目的語省略の問題について」. 『言語情報学研究報告』No.12, 東京外国語大学, pp.183-199.
- 小川敬洋. 2006. 「口語コーパスにおける婉曲を表すフランス語の条件法」. 『言語情報学研究報告』No.11, 東京外国語大学大学院地域文化研究科, pp.383-398.
- 近藤野里. 2010. 「フランス語話し言葉におけるリエゾン —Aix-en-Provence コーパスを用いた統語・音韻分析—」. 東京外国語大学地域文化研究科提出修士論文.
- . 2012 出版予定. 「フランス語のリエゾンにおける社会言語学的要因の一考察」. 『ロマンス語研究』No.44, 日本ロマンス語学会.
- 塩田明子. 2001. 「時制と連結辞—話し言葉の語りにおいて」. 『フランス語学研究』No.35, 日本フランス語学会, pp.43-48.
- . 2003. 「話し言葉における *présent de narration*」. 『ふらんぼー』No.29, 東京外国語大学フランス語研究室, pp.21-39.
- . 2005. 「語りの構造と時制」. 『フランス語を語る』, 三修社, pp.180-192.
- 杉山香織. 2011. 「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一」.『外国語教育研究』No.14, 外国語教育学会, pp.20-34.
- . 2012 出版予定.「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』No.44, 日本ロマンス語学会.
- . 2013 出版予定.「話しことばにおける語彙の豊かさ一言語コーパスに基づく中間言語研究」.『フランス語学の諸問題 IV』, 三修社.
- 東郷雄二. 1997.「会話フランス語のストラテジー」.『フランス語学研究』 No.31, 日本フランス語学会, pp.15-26.
- 馬場淳子. 2011.「日本とフランスで出版されたフランス語教科書の特徴 ダイアログで示された会話の語彙分析」.『外国語教育研究』No.14, 外国語教育学会, pp.81-96.
- 前島和也. 1997.「口語における文末の là」.『フランス語学研究』 No.31, 日本フランス語学会, pp.34-39.