フランス語学習者話し言葉コーパスのアノテーションと分析

東京外国語大学　秋廣尚恵

本発表の流れ

1. フランス語学習者話し言葉コーパスのデータ
2. アノテーション
3. 誤用分析について
4. 今後の展望

1. フランス語学習者話し言葉コーパスのデータ

2.1.　12月時点で収録済みのコーパスの概要

- 学習者レベル：B1－C1
- インフォーマント数：39名（女31名、男8名）
- インタビュー：46件
- 自由会話：67件
- 全体の時間数：26時間
- フォーマット：音声と転写表記のアライメント方式（Transcriber)

2. アノテーション

2.1. 共有アノテーションの方式

- 母音や子音（あるいはその複合的なもの）が変化している場合：

  [v]_　あるいは[c]　　[v/c]__

- エラーと思われる形式とそれに対する正しい形式

  [e]_…_[xxx]

  例）[e]_à_[en]

  Allemagne, je [e]_vas_[vais]

- エラーと思われる形式とそれをインフォーマント自らが自分で訂正している場合

  [ac]_[ ]:

  例）je suis allé [ac]_à Allemagne euh je suis allé en Allemagne_[ ]

- ポーズ　#
- 発話の重なり部分　<　　>
- 単語の言いかけ　X-

GARSやIPFCのアノテーションを参考。

report

[no speaker]

AH1 - ok moi je veux parler [e]_sur_[de] euh # respecter s- l'heure

YI1 - mh

AH1 - parce qu'à Québec # à Montréal il y avait beaucoup de gens q- [e]_[v:]_de_[du] Japon de France

YI1 - <mh mh>

AH1 - <[e]_du_[de] Montréal> # du Québec # et euh # les gens du Jap- les Japonais et les Québécois # respectaient l'heure parc- <donc euh>

YI1 - <ah c'est> vrai ?

AH1 - ouais ouais

YI1 - <oh>

AH1 - donc euh l- les Québecois c'était comme

YI1 - <mh>

AH1 - <les Japonais> par exemple euh

YI1 - <[c:]_mh>

AH1 - <[v]_quand on dit euh> # (X) on va [e]__[se] voir euh à lib- # [v:]_à [e]__[la] biblio

YI1 - <mh mh>

AH1 - <bibliothèque> à douze heures

YI1 - <ouais>

jpto2ah1yi1_1l

report

(no speaker)

AH1 - ok moi je veux... l'heure | YI ...h | AH1 - parce qu'à Québec #... France | AH1 - <[e]_du_[de] Montréal> # du Québec... <donc euh> | YI ...? | A s | YI ...h> | AH1 ...me | AH1 ...euh | Y > | AH1 - <[v]_quand on dit... [e]__[la] biblio | AH1 -... ...heures | AH1 - <ils ... ...moins cinq ou | AH1 - <dc ... euh

0 5 10 15 20 25 30

３．誤用分析について

■ 形態統語的・文法的なアノテーションに関して

誤用とその訂正（転写者による訂正と自己訂正）

→　誤用の分析を行う必要がある。

誤用のタイプには実に様々なものが存在する。

３.１． 学習者の言語の多様性

Sylviane Granger, Gaëtanelle Gilquin and Fanny Menier (2015) *Learner Corpus Research*, Cambridge, CUP.

Granger et al. (2015: 39)

「学習者の産出するものは、スペルミス、語形成、文法的誤用の影響を受けるので、自動でターゲット形式を抜粋することが困難である。（中略）研究者は学習者データを扱う際、学習者がネイティブと同じ構造やパターンを必ず使用していると断言することができないという点に留意する必要がある。」

Contrastive Interlanguage Analysis （対照的中間言語分析）

学習者の言語をネイティブの参照言語と対照させつつ「両者の間の量的、質的違いという観点から分析、解釈を行う」(Granger et al 2015 : 40)

その際に、対照するべき双方の言語が多様性を持つものである点に注意する必要がある。

「 Reference language varieties（参照言語の多様性）」
「 Interlanguage varieties （中間言語の多様性）」

(Granger et al 2015 : 40)

3.2.　Computer Aided Error Systemの開発

- 形態
- 文法
- 語彙
- レジスター
- 文体
- 語の繰り返し
- 語の欠如
- 語彙文法
- 分節

などのカテゴリーに分けて誤用タグを付すシステムの開発が様々な研究者によって、目下行われている。

→人的、時間的、財政的にコストの大きな研究分野である。

→文字化された転写に基づくデータを対象とする。転写方式により分析は左右される。

→分析者の主観の影響を受けやすい。

3.3. エラーと誤り

■ Errors （エラー）

口が滑ってしまった、うっかり言い間違えたもの
（習得はしているが、注意を怠ってしまったために出る誤用）

■ Mistakes（誤り）

習得ができておらず、知識が欠落しているために出る誤り。

3.4. 非用のケースの扱い

■ 誤用のタグ付けは誤用にのみ焦点を当ててしまい、学習者のデータの全体像を見失わせる場合がある。

→非用のケースがある。（Granger et al. 2015:41)

Mistakes / Misuses

「ある形式の過剰な使用は明らかに規範に反する誤用である。非用は誤用はないが、学習者がその形式を学ぼうとしているかどうかは疑わしい。」（Granger et al. 2015:41)

3.5. 誤用研究の方法論

- 量的調査を行い、学習者の overuse（過剰使用）、underuse （過少使用）されている形式を特定する。

- 質的調査を行い、学習者の誤用を特定する。
  - その誤用は、ある特定のグループの学習者に特有のものなのか、あるいは広く一般に見受けられるものかを調査。
  - その誤用は、ある特定の状況においてのみ観察されるものか、あるいはどんなときにも観察されるものであるか調査。
  - 他言語の転用の影響が見られるかどうかを調査。

（*Granger et al. 2015*：39）

3.6. コーパスに見られる誤用例と分析

- Jpto1 について誤用タグの付された要素は、全部で41件。誤用の内訳を分析する。

①自己訂正のケース：5件

- 語彙の訂正：3件
- 文法の訂正（冠詞の言い換え）：2件

③誤用：36件

- 誤りに気が付いて訂正しようとしたができなかった：2件（語彙１、文法１：発話の中断、ためらい）
- 形態13件（名詞、動詞の屈折、形容詞の一致など）
- 語彙4件（いずれも同じものの繰り返し）
- 文法17件
- 談話1件

4. 今後の課題（計画段階）

- 談話レベルの誤用をどう定義するか。
  - 流暢性と非流暢性
  - 談話標識の誤用と非用
  - 文法性よりも適切性の観点からの分析
- 話者と状況に応じて特有な誤用か、あるいは一般的にみられる誤用かの見極めをするためには、量的、質的分析の双方を行う必要がある。
  - データの統計化ー量的分析
  - 学習者個々人による差異→質的分析（社会言語学的調査、インタビューなどの追跡調査が必要）

- 転写の終了したコーパスすべてのマルチレベルの分析アノテーションを行う。
  - 分析用アノテーション方式を定める。
  - *TXM→Excel*にデータを移し、分析用アノテーションを手作業で付与。

- 統計的な処理による量的分析により、全体的な傾向性が現れるかを調査。（例外的な誤用は質的分析の対象とする）

- 質的分析
  - 学習者ごとに誤用のパターンを観察（ケーススタディ）
  - 追跡調査（インタビュー）