

ゼミ論：  
NICT JLE コーパスストーリーテリングに基づいた  
学習者の特徴抽出とスピーキング指導への応用

言語文化学部3年

7413104

山崎加奈

# 目次

- 研究概要
- 先行研究
- NICT JLEコーパス・ストーリーテリングの概要
- 仮説
- 分析1~5
- 結論：学習者の特徴・指導への応用
- 今回の研究の限界・問題点

# 研究概要

- 対象コーパス：NICT JLEコーパス
- 対象タスク：ストーリーテリング
- 目的：4技能型大学入試が導入されるにあたり、スピーキングテストに頻出のストーリーテリングの学習者の特徴を知り、指導へ活かす
- 目標とする教材・指導：ストーリーテリングを指導するステップ・コツ
- 使用ツール：AntConc
- 分析方法：SSTレベル4~8の中級学習者とNSを比較  
N-gramと目視から仮説を立て、検証

# 先行研究

- 投野由紀夫. (2008). NICT JLE vs. JEFLL--n-gram を用いた語彙・品詞使用の発達 (シンポジウム 英語学習者コーパスの新展開--会話 (NICT JLE) vs 作文 (JEFLL) コーパスの比較と分析). *英語コーパス研究*, (15), 119-133.

# NICT JLE Corpus

- 学習者スピーキングコーパス
  - 受験者：主に大学生～社会人
  - 3つのタスク：イラスト描写・ロールプレイ・ストーリーテリング
  - SSTレベル1~9（1~3：初級学習者・4~8：中級・9：上級）
- 
- NS：20人分（39タスク）
  - NNS：80人分（80タスク）

# Stage4 ストーリーテリング

- 6または4枚の絵を見ながら「話」をつくる
- One day last weekから発話を始めるよう指示

## テキスト構成

- `<stage4></stage4>`
- `<task></task>` : 学習者がストーリーテリングをしている部分
- `<followup></followup>` : タスク後の雑談

# 仮説を立てる

- 1. NICT JLEコーパスの元データから、B（学習者）の発話のみを抽出
- 2. Stage4のみを抽出
- 3. n-gram分析（trigram, 4-gram）

n-gramトークン数      NNS : 23,684

NS : 23,318

- 4. テキストを実際に見る

# N-gramからわかったこと

- NS : butやandなどの接続詞を含むn-gramが上位に
  - NNS : 接続詞を含むn-gramがない (43位 and after thatが最上位)
  - →学習者はスピーキングで接続詞を適切に使えてないのではないか？
- 
- NS : itを含むn-gramが多い
  - →学習者は無生物主語 (it) を使えていないのではないか？
- 
- NNS : go, wentの多用
  - →ある場面を見ると「～へ行った」という表現から始める傾向？



# テキストを見て気づいたこと

- NS : 登場人物に名前をつける
  - NNS : a lady, a woman
  - →詳しい人物設定をした方が話しやすいのでは？
- 
- NNS : 現在形を使う。直接話法
  - NS : 過去形を使う。間接話法
  - →ストーリーテリングで過去形で話すことが身についていない？

# 仮説：学習者の特徴

- 接続詞を適切に使用できていないのではないか（分析3：接続詞使用頻度）
- 接続詞を適切に使用できていないとすれば、NNSの発話は1文の長さが短く、文の数が多いのではないか（分析1：文の数と長さ）
- 過去形で話せていないのではないか（分析2：動詞の時制）
- 直接話法を頻繁に用いているのではないか（分析4：直接話法vs.間接話法）
- 人名の固有名詞の使用頻度が低いのではないか（分析5：固有名詞の使用）

# テキスト整形

- 1. 学習者の発話のみ抽出（<B></B>）
  - 2. <stage4></stage4>と<task></task>のみ抽出
  - 3. 1行1文（置換：\。→.\r\n）
  - 4. 行頭行末にタグ付与（<s></s>）
  - 5. 品詞タグ付与（Free CLAWS WWW tagger: CLAWS5）
- 
- NS：20人分（39タスク）
  - NNS：80人分（80タスク）

# 分析<sub>1</sub>：文の数と長さ

- 文の数
- 分析方法：AntConcで <s> の数を数える
- NS：711/39task → 18.2文/1task
- NNS：1222/80task → 15.3文/1task
- 結論：NNSはNSより文の数が少ない

# 分析<sub>1</sub>：文の数と長さ

- 文の長さ
- NS：総語数7995 11.2語/文
- NNS：総語数14234 11.6語/文
- →NNS：フィラー, 言い淀み, 繰り返しが多いためでは？
  
- <nvs></nvs><F></F><R></R><SC></SC>カット
- NS：総語数7155 文667 10.72語/文
- NNS：総語数9928 文1044 9.5語/文
  
- 結論：内容量で見ると、NNSの方が1文の語数が少ない

# 分析2：動詞の時制

分析方法：

- 1. AntConcでV.B（原形） V.Z（現在形）をそれぞれ検索
- 2. 目視で確認
- 3. 以下の不要な項目を除外
  - 非言語音：<nvs></nvs> e.g. cough, sigh
  - フィラー：<F></F> e.g. err, urm
  - 連語のフィラー：you know, I mean
  - I think, I guess, I mean
  - thank you

# 分析2：動詞の時制

結果：原形/現在形

- NS : 155/146 計.301 7.7/ 1task
- NNS : 442/221 計.663 8.2/ 1task

- 結論：過去形で話すことを促すタスクであるにもかかわらず、原形・現在形の動詞使用が多い

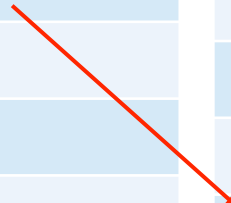
# 分析3：接続詞使用頻度

- 分析方法：AntConcでCJ.（接続詞）を検索
- CJC（等位接続詞） e.g. and, or
- CJS（従位接続詞） e.g. although, when
- CJT e.g. that



# 分析3：接続詞使用頻度

	NS		NNS
521	and (67%)	710	and (66%)
55	that	94	but
46	but	85	so
32	so	43	because
22	because	38	that
21	or	26	after
17	as	21	when
:	:		
4	after		
Total. 772		Total. 1075	



## 共通点

- andの割合
- but, so, becauseの順番

## 相違点

- NNS：that使用頻度が低い
- NNS：after使用頻度が高い

# 分析3：接続詞使用頻度

	NS		NNS
511	and (67%)	669	and (66%)
53	that	93	but
46	but	85	so
32	so	38	because
22	because	34	that
20	or	23	after
17	as	19	when
:	:		
4	after		
Total. 758		Total. 1014	

<R></R>削除

- ・ andのリピート率が高い
- ・ 頻度順は変わらない

c.f.

NS : 711文      文<接続詞

NNS : 1222文    文>接続詞

NNS :

文の数に対する接続詞の数が少ない

# 分析3：接続詞使用頻度

NNS : n-gram (n=2)

	X + that
35	after that
8	and that
6	on that
5	er that

Cluster Types/ Cluster Tokens :  
64/ 138

	that + X
20	that's
11	that day
6	that movie
6	(after) that, they
5	that man
5	that was
5	(after) that, she

61/ 138

高頻度コロケーション

- 「after that」
- 「that + 普通名詞」計. 92  
→thatを代名詞として使用

# 分析3：接続詞使用頻度

NS : n-gram (n=2)

	X + that
9	decided that
6	and that
4	so that

Cluster Types/ Cluster Tokens :  
61/ 91

	that + X
14	that they
8	that's
7	that she
6	that he
5	that it

44/ 91

高頻度のコロケーション

- ・ 「decided that」
  - ・ 「that + 代名詞」
- thatを接続詞として使用

# 分析3：接続詞使用頻度

- 結論：
- NNSの特徴
- 文の数に対する接続詞の数が少ない
- 接続詞thatの使用頻度が低い
- 「after that, SV...」の多用（NNS：35 NS：3）
- 代名詞thatの用法が多い（⇔NS：接続詞thatの用法が多い）
  
- 考察：接続詞、特にthatを適切に使用できていないのではないか

# 分析3：接続詞使用頻度

NS：品詞 + that

	NS
19	名詞 + that e.g. the reason that, 関係代名詞
18	動詞 + that e.g. decide, think, see, suggest
7	形容詞 + that e.g. so + 形容詞 + that SV 感情を表す形容詞 + that SV

結論：関係代名詞・so + that構文・動詞 + that節など、  
中学生で習う項目が学習者は会話で産出できていない  
考察：既習項目を使った会話練習が必要ではないか

# 分析4：直接話法 vs. 間接話法

分析方法：

- 1. AntConcで直接話法を検索 (\\".\*?\\")
  - 2. 目視で確認
  - 3. "固有名詞"と "One day last week?" (タスク冒頭の聞き返し) を除外
- 
- NS : 37 0.95/ 1task
  - NNS : 57 0.71/ 1task
- 
- 結論 : NSの方が直接話法を使用している

# 分析4：直接話法 vs. 間接話法

## 直接話法の前のコロケーション

- NS : S+say/said 18 c.f. be+like 10
- NNS : S+say/said (間にフィラー有) : 33 c.f. ask : 9 tell : 1
- 結論 : 直接話法の台詞の部分が発話しながら考えている  
直接話法を正しく使用できていない
- 分析3結論 : NNSはthat使用率が低い
- 考察 : thatを使わない直接話法の方が負荷が低いのではないか



# 分析5：固有名詞の使用

分析方法：

- 1. 固有名詞（NPo）をAntConcで検索
- 2. 目視で確認・人名のみ抽出
  
- NS：120      3.08/ 1task      e.g. Bob, Suzie, Vincent
- NNS：53      0.66/ 1task      e.g. Ken, Hanako, Taro
  
- 結論：NSはストーリーテリングの際に人名を使用する傾向にある
- 考察：人名を設定した方が話しやすいのではないか

# 結論：学習者の特徴

- 文の数が少なく、1文中の語数も少ない
- 過去の文脈で動詞の過去形を使えていない
- 文の数に対する接続詞の数が少ない
- 接続詞thatの使用頻度が低い
- 代名詞thatの使用頻度が高い
- 「after that, SV...」の多用
- 直接話法の使用頻度がやや低い（askやtellを使用）
- 人名の固有名詞の使用頻度が低い

# 結論：スピーキング指導への応用

ストーリーテリングで高得点をとるために...

- 過去時制で話す練習
- 接続詞、特にthatを会話で使う練習
- 中学レベルの文法事項を使った会話練習
- 「S say, “～”」の型を覚えさせ、直接話法を正しく使用させる
- 登場人物に名前を付けさせる（日本名で可）

# 今回の研究の限界・問題点

- タスク統制を一切していない  
→タスクによる影響がある可能性がある
- 繰り返しやフィラーの扱い方  
→スピーキングという特性をどう考えるか
- 仮説と分析の対象が違う      NNS14,232    NS7,993  
    仮説 : <task></task>+<followup></followup>  
    分析 : <task></task>