

2016/12/03
 国立国語研究所第1回学習者コーパスワークショップ

世界の英語学習者コーパス研究の潮流：HowからWhyへ —学習者コーパス研究の3つの課題をめぐって— (発表資料：公開用)

石川 慎一郎
 神戸大学



コーパス言語学の目指すこと

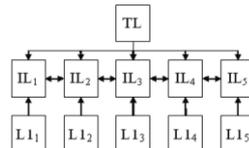
- 直観ではなく、データを根拠にする
- 頻度データを通して言語の実相を探る
- 頻度によって見えない言語を可視化する
- 母語話者・辞書・文法書などの常識の見直し
- 語や言語の「振る舞い」を正確に記述する

2

中間言語対照分析

- Contrastive interlanguage analysis (Granger, 1998)
- L1話者×L2話者 (日本語学習者 vs 日本語母語話者)
- L2話者×L2話者 (中国人日本語学習者 vs アメリカ人日本語学習者)

誤用 (misuse) だけでなく、過剰・過小使用 (over/underuser) も分析



Leriko-Szymańska (2008)

従来の学習者コーパスは比較可能か？

Essay	L1	ProfLev	Topic	Length	Dict
A	中国語	A2	文化祭	100語	辞書使用
B	韓国語	B2	水力発電	300語	辞書なし

得られた「差」は何を意味しているのか？

The ICNALE

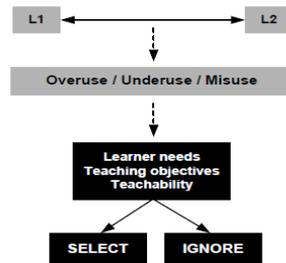
- International Corpus Network of Asian Learners of English (Ishikawa, 2013/2014)
- 世界最大のアジア圏英語学習者コーパス
- 180万語
- 3,550 人の学習者
- 10 カ国・地域
- 4,400 本の発話
- 5,600 本の作文
- 10,000 本の産出サンプル



学習者コーパスの3つの留意点

- (1) 対照分析で見つかった差異はすべからく教授の内容に含むべきか
- (2) 異なる種類のデータを用いて対照を行うことに問題はないのか
- (3) 学習者との比較の規準となる母語話者のL1運用の安定性を前提視してよいのか

(1) 教授項目化 (Granger, 2009)



(2) 対照可能性

- ICANLE・・・トピックは2つに。時間・分量も固定。
- I-JAS・・・統制的O P I形式
- イラストを描写するストーリーテリング (5分)
- 対話 (30分)
- ロールプレイ (10分)
- ストーリーライティング

(3) NS規準の安定性

- 1) 属性的多様性
- The conventional prescriptive view has been that the goal of foreign language learning is to approximate closer and closer to the performance of native speakers. Yet which native speakers? American, Australian, British or Caribbean? Highly educated or less so? Old or young? Such questions as these cause difficulties, although in practice teachers probably have covert answers to them. The problem becomes more noticeable when we compare learner corpora with a native-speaker 'reference corpus'.... Native-speaking students do not necessarily provide models that everyone would want to imitate. And, when we come to examine a reference corpus of native-speaker speech, the less admirable features of the native speaker's performance can show up especially clearly... (Leech, 1998, p. xix)
- 2) 産出的多様性
- 従来, 十分に研究されていない
- 点と点から面と面へ

実例研究

I-JASのNSデータの内的多様性

- RQ1 母語話者によるL1の書き言葉産出データは, 文法的正確性においてどの程度安定しているか?
- RQ2 母語話者によるL1の書き言葉産出データは, 語数および句読点・高頻度語の使用状況においてどの程度安定しているか?
- RQ3 母語話者によるL1の書き言葉産出データは, 高頻度語の使用率に着目した場合, どの程度一体的なものでありうるか?

RQ1 文法的正確性

- (1) 犬が飛び出して (→飛び出して) きました。 (JJJ03)
- (2) そんな仲睦ましい (→仲睦まじい) 二人を… (JJJ026)
- (3) 行く場所を地図で確認してる [→している?] 隙に… (JJJ35)
- (4) 出かける前に二人が地図を見ている間に…
- (5) 行く場所も決定し歩きだす二人。目的地に着いて「さあ食べよう」とバスケットを開けるとそこからは犬が・・・驚く二人。 (JJJ09)
- (6) ケンはバスケットを地面に下ろすと、まっぴっくり。… (JJJ10)
- (7) …するとあろう事が、飼い犬が飛び出したではありませんか。… (JJJ26)

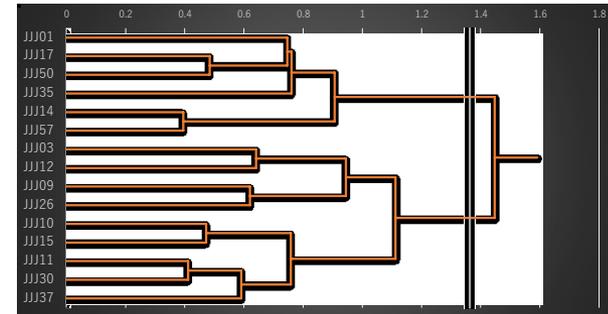
RQ2-1 語数・句読点

	語数1	語数2	読点	句点	句読点	読/句	読点%	句点%	句読点%
Av	108.1	119.3	4.7	6.5	11.2	1.1	4.1	5.4	9.4
SD	30.4	33.1	1.8	3.4	3.9	1.4	1.6	1.8	1.9
CV	0.3	0.3	0.4	0.5	0.3	1.2	0.4	0.3	0.2
Max	180.0	199.0	7.0	14.0	20.0	6.0	6.5	8.2	11.7
Min	81.0	90.0	1.0	1.0	7.0	0.1	0.9	0.6	4.4
Max/Min	2.2	2.2	7.0	14.0	2.9	42.0	7.5	13.1	2.7

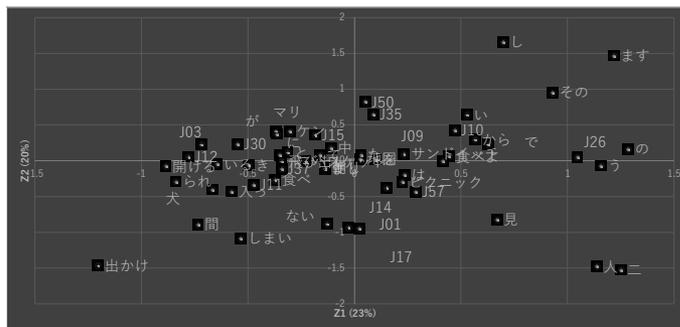
RQ2-2 高頻度形態素

	JJJ01	JJJ03	JKK09	JJJ10	JJJ11	JJJ12	JJJ14	JJJ15
1	に	に	た	た	た ¹	た	た ¹	た
2	た	た	に	と	と ¹	と	を ¹	て
3	て	を	を	は	て ¹	に	て ¹	まし
4	は	が ¹	が ¹	て	に ¹	が ¹	まし	と
5	まし	と	と	を	まし ¹	まし	は	に
	JJJ17	JJJ26	JJJ30	JJJ35	JJJ37	JJJ50	JJJ57	
1	て ¹	た ¹	に	を	て	て	た ¹	
2	を ¹	の ¹	を	に	た	た	を ¹	
3	た	を	と	た	と	を	まし	
4	まし	は	た	て	は	に	は	
5	と	に	まし	し	まし	まし	て	

RQ3 内的一体性



RQ3 内的一体性



学習者コーパス研究者が考えるべきこと

- 学習者の個別差に徹底的に寄り添う
- そうすると「学習者」という概括的存在は消える
- “an ideal speaker/hearer” (Chomsky)
- “an ideal learner” (Granger)
- そこに逃げ込むことは許されるのか？
- これに代わる学習者の定義をどう作るか？
- 学習者コーパス研究者には実践と理論構築の両面が求められる