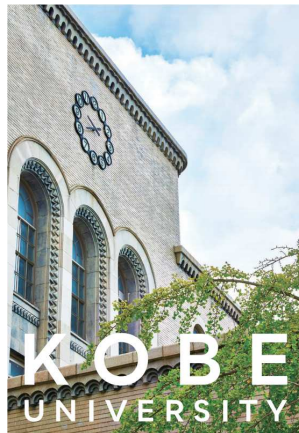


2020/12/05
日本語OPI研究会第104回定例研究会 (online)

学習者コーパスを用いた 発話研究の展望

L2英語学習者とL2日本語学習者を事例として

石川 慎一郎
iskwshin@gmail.com



学習者コーパスは若い研究分野



	60s	70-80s	90s	2000s	10s	20s
LI	64 Brown		94 BNC	08- COCA	11 BCCWJ	
IL				02/09/20 ICLE	10 LINDSEI 13- ICNALE	20 I-JAS

先駆的業績
98 KY

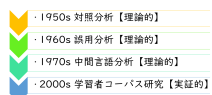
[LC are relatively new but] “highly influential in the world of ELT publishing” (McEnery & Hardie, 2011, p.82)

対照分析仮説 (1950s)



- 50年代＝言語を規則体系ととらえる構造主義言語学 (Broomfield, 1933)
- Lado (1957) L2が習得できないのは、言語規則がL1と異なるから
- L1→正の転移 (positive transfer) / 負の転移 (negative transfer)
- 負の転移→干渉 (interference) → 誤り (error)
- 対照分析仮説 (contrastive analysis hypothesis: CAH)
- L1とL2の異なりを明らかにし、そこに焦点を当てて学ぶことが効果的
- L1がX語なら必ずL2で特定の誤り (error) を犯すと言えるのか?
- Wardhaugh (1970)
- 習得難度を予測→「強化版CA」(strong version)
- 習得難度を説明→「弱化版CA」(weak version) → 誤用分析に発展

誤用分析 (1960s)



- 1960年代～
- 旧来の行動心理学モデルと、新しい生成文法モデルの対立 (※行動心理学をベースとしたAudio-lingual methodの見直し)
- Corder (1967) 2つの言語システムの界面として誤用 (error) に注目
- 誤用パタンの個人差が大きければ行動心理学モデル、個人差が(少)なければ生成文法モデルを支持することになる
- Corderの結論
- 誤用には体系性と再現性がある。
- 生得的システム (built-in syllabus) の存在を証明。

中間言語分析 (1970s)



- 1970年代～
- 誤用に限らず学習者の言語運用を総体的に研究する機運が高まる
- 学習者のL2使用を「中間言語 (interlanguage)」とみなし、固有の研究対象に格上げ
- 化石化 (fossilization)
- 1) 言語転移 (linguistic transfer) : L1の負の転移
- 2) 過剰般化 (overgeneralization) : 習得ルールを過剰適用
- 3) 訓練転移 (transfer of training) : 不自然なスキット、教材、ドリル
- 4) 学習方略 (strategy) : チャンク学習方略→「*飲んだのほうがいい」
- 5) コミュニケーション方略 (communication strategy) : 苦手回避

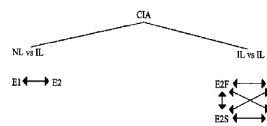
学習者コーパス研究 (2000s～)



- 各種の学習者コーパスの開発が進む
- 理論中心の学習者研究に、データに基づく実証的視点が加わる
- 学習者コーパス研究の主な対象
- (1) 過剰・過小使用される言語特徴
- (2) 母語転移 (L1 transfer)
- (3) コミュニケーション上の回避方略
- (4) 母語話者の・非母語話者の言語使用特性
- (5) 特定母語話者が苦手とする言語領域

対照中間言語分析 (contrastive interlanguage analysis: CIA)

• Granger 1996 (CIA)

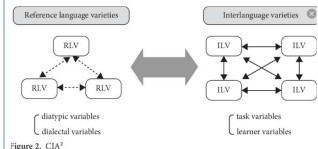


批判

NSは正しい英語を書くと言えるのか？

NSモデルの安易な押し付けにつながるのでは？

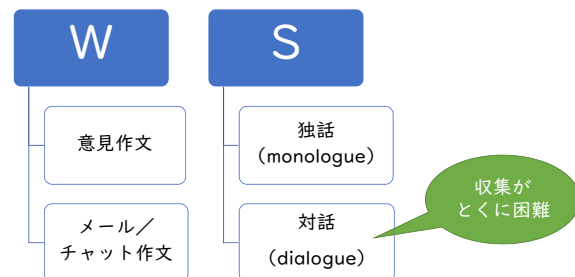
• Granger 2015 (CIA2)



参考：Davies, A. (2011). Does language testing need the native speaker? *Language Assessment Quarterly*, 8, 291-308.

学習者コーパスの類型論

媒体（書く・話す）
方向性（一方向・双方向）



主要学習者対話コーパスのインタビュー設計

LINDSEI	NICT-JLE	Trinity Lancaster	ICNALE SD	I-JAS
1) トピック発話（3題から選択。準備後3-5分会話） 2) 質疑（トピック他について） 3) 絵描写（4枚）	1) 導入会話 2) 絵描写（1枚） 3) RP（質問・依頼・交渉より選択） 4) 絵描写（4/6枚） 5) まとめ会話	1) テーマプレゼン（事前準備） 2) 質疑（プレゼン内容について） 3) やりとり（設定を選んで会話） 4) 学術対話（テーマに基づく）	1) 導入質疑 2) 絵描写 I 3) 関連質疑 4) RP I（説得） 5) 関連質疑 6-9) 別課題 10) ふりかえり 11) LIふりかえり	1) ST（4/5コマ） 2) 対話（15問） 3) RP（依頼／拒否） 4) 絵描写（1枚） 5) SW（ST作文）

I-JASの主要特徴

- ・豊富なデータ量（1,050人，対面発話808万語）
- ・幅広い学習者（母語12種／教室・自然環境）
- ・幅広いタスク（7種：S/W）
- ・詳細な学習者背景調査（習熟度＋学習環境）
- ・参照用母語話者データの収集
- ・収集データの全面公開（文字＋音声）

10

各国中級学習者の副詞使用（石川，2020a） RQ1 使用量

母語	人数	全体トーン数	1名あたりトーン数	タイプ数
中国語	75	4,758	63.4	142
韓国語	26	1,863	71.7	94
タイ語	24	1,678	69.9	84
ベトナム語	18	1,171	65.1	77
インドネシア語	15	1,217	81.1	80
英語	22	2,872	130.5	106
ドイツ語	18	1,973	109.6	87
フランス語	3	246	82.0	36
スペイン語	4	458	114.5	37
ロシア語	17	1,608	94.6	83
ハンガリー語	17	1,996	117.4	76
トルコ語	9	1,015	112.8	69
日本語	50	12,629	252.6	382

- ・1名あたりトークン
- ・NNS (60-130) < NS (252)
- ・副詞を豊かに使えることが「NSらしさ」の一部を構成
- ・NNSの中では、中韓よりも、英・独・ハンガリー等が相対的に副詞を多く使う

RQ2 高頻度使用項目

母語	高頻度使用副詞（頻度順）
中国語	そう／ちょっと／たぶん／とても／もう
英語	そう／ちょっと／たぶん／とても／まあ
フランス語	たぶん／そう／あまり／とても／ちょっと
ドイツ語	そう／ちょっと／とても／もう／たぶん
ハンガリー語	そう／ちょっと／まあ／たぶん／あんまり
インドネシア語	そう／まあ／ちょっと／たぶん／よく
韓国語	そう／ちょっと／あんまり／たぶん／まあ
ロシア語	そう／ちょっと／とても／もちろん／もう
スペイン語	そう／たぶん／／とても／ちょっと／あまり
タイ語	そう／ちょっと／あまり／たぶん／たくさん
トルコ語	そう／ちょっと／あんまり／もう／もちろん
ベトナム語	そう／ちょっと／たぶん／よく／たくさん
日本語	そう／もう／ちょっと／こう／まあ

- ・【NS/NNS共通】
- ・「そう」「一寸」は全NNSと共通
- ・「もう」「まあ」は半数のNNSと共通
- ・【NSのみ】
- ・「こう」：『グリー』は、高校生で…グリーグってゆう、さう、合唱？…ちょっとさうアクションってゆうか… (JJJ02)
- ・【NNSのみ】
- ・「たぶん」(9)「とても／あ(ん)まり」(6)
- ・NS: 語調を整えるフィラー的副詞使用
- ・NNS: 主張をやわらげるヘッジ的副詞使用

RQ3 特徴使用項目

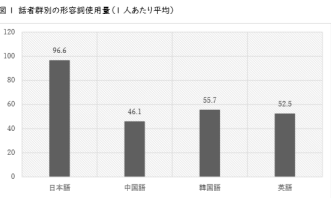
表3 学習者の過剰・過小使用語 (上位 10 語)

過剰使用語 (学習者>母語話者)				過小使用語 (学習者<母語話者)			
語	統計量	効果量		語	統計量	効果量	
とても	689.1	17.0	こう	1,036.7	0.08		
あまり	307.2	6.5	やっぱり	217.4	0.03		
たくさん	294.5	6.4	そう	208.4	0.69		
たぶん	288.1	2.8	もう	191.8	0.3		
もっと	249.6	7.5	やっぱり	165.7	0.41		
ちよっと	174.8	1.7	まあ	163.9	0.49		
よく	166.8	2.8	結構	147.1	0.32		
たとえば	163.9	3.6	けっこう	130.8	0.14		
一番	156.2	2.9	なかなか	74.4	0.16		
まだ	104.3	2.7	たまたま	65.2	0.08		

注：過小使用語の「結構」と「けっこう」は使用したデータ内の表記に誤れがあったため二重で登録されたものである。

- 簡単にNSっぽくなる方法
- 〜こう、やっぱさ、そうなんだよね。で、もう、やっぱり、まあ、結構なかなかなんだけど。たまたまなんだよね…
- 中保(2018)上級者マーカー=「結構」などの漢語の副詞的使用
- 山内(2020) KYコーパスでの超級特徴語:けれども、こう、そういって、まあ、思うんですけど、すげ、んてうすね、んてすよ、んてすよね(In 野田・迫田(編))

L1中・韓・英学習者による形容詞使用 (石川, 2020b) RQ1 使用量



- 1名あたりトークン
- NNS (46-56) < NS (97)
- 前述の副詞に加え、形容詞を豊かに使えることも「NSらしさ」の一部を構成
- 副詞の場合、欧州>アジア圏学習者の構造があったが、形容詞では英語≒中韓となる

RQ2 高頻度使用項目

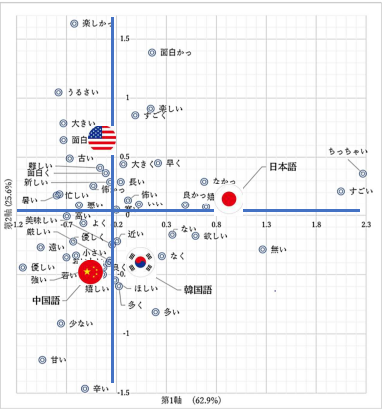
表1 高頻度形容詞 (上位 20 語, 用者に複数語が並ぶ場合は上階級数まで任意に語を選択)

#	日本語話者		中国語話者		韓国語話者		英語話者	
	形容詞	%	形容詞	%	形容詞	%	形容詞	%
1	ない	22.7	ない	15.7	ない	22.3	ない	27.6
2	いい	11.3	いい	10.7	いい	10.2	いい	24.1
3	すごい	7.0	小さい	4.6	なく	6.1	すごく	9.2
4	なく	3.5	多い	3.3	小さい	3.4	面白い	7.8
5	すごく	2.9	なく	2.7	多い	3.3	小さい	6.7
6	多い	2.7	怖い	2.5	なかっ	2.9	難しい	6.6
7	なかっ	2.6	面白い	2.3	難しい	2.0	怖い	4.9
8	怖い	2.2	難しい	2.1	すごく	2.0	なく	4.8
9	難しい	1.5	悪い	1.9	辛い	1.7	大きい	4.5
10	小さい	1.4	辛い	1.9	悪い	1.6	楽しかつ	4.3

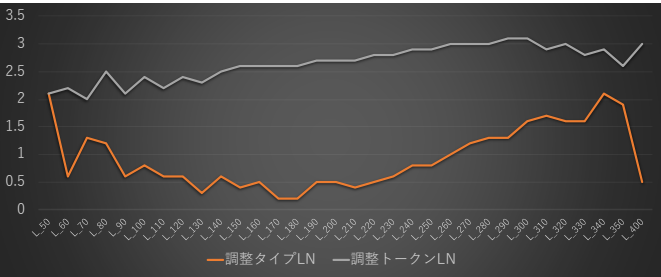
- 【NSのみ】「すごい」「ちっちゃい」
- NNSは全体として丁寧なモードを保持して発話
- NSは内容(視点)に応じて丁寧度のモードを連続的に変化
- <自分の過去への言及>
- |まあ|それ|ちっちゃい|頃|から|して|てー|(JJJ16)
- <一般的な事象への言及>
- |子供|が|小さい|うち|は| (JJJ16)

RQ3 相互関係

- NNSの中では、韓がNSに近い
- NSの特徴形容詞
- 【夕接続】なかっ(た), 良かつ(た), 嬉しかっ(た)
- 【口語的】すごい, ちっちゃい
- NNSの特徴形容詞
- 中国語話者
- 【個人感情】嬉しい, 欲しい
- 【イ型】若い, 強い, おいしい
- 英語話者
- 【イ型】大きい, 面白い, 古い
- 韓国語話者 ※機能語的用法を多用
- 【存在・量】多い(く), ない(く)



学習者習熟度の簡易測定 (石川, 2021) RQ1 習熟度による使用量の変化



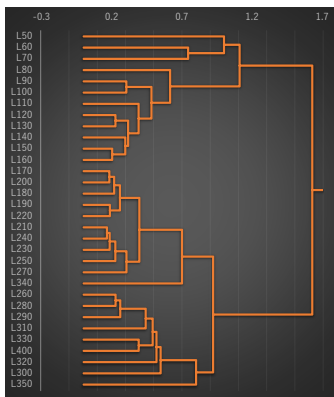
RQ2 習熟度と頻度が相関する動詞群

相関の程度	動詞
ρ>=.7 (7語)	飛び出す／知る／気付く／居る／来る／飼う／仕舞う
ρ>=.4 (27語)	着く／出掛ける／持つ／開ける／為る／空く／終わる／向かう／驚く／飛び出る／決める／飛び込む／入り込む／楽しむ／困る／成る／話し合う／食う／齧る／慌てる／見る／思う／隠れる／覗く／繋ぐ／忍び込む／残る

RQ3 発達の節目

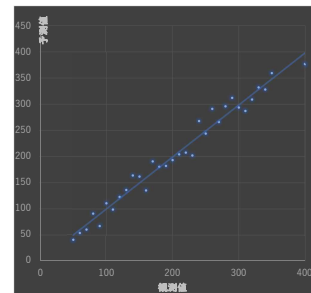
- ・ 頻度上位100語による階層クラスター分析
- ・ 隣接スコアは同一クラスターに分析される
- ・ 学習者は2つの大クラスター, 3つの中クラスターに区分される

I		II	
I a	I b	II a	II b
50~70	70~170	170~270	260~400



RQ4 上位20語による習熟度予測モデル

- ・ J-CAT総合スコア予測式=
- ・ 「飛び出す」*191
- ・ + 「出掛ける」*185
- ・ + 「出る」*143
- ・ + 「知る」*115
- ・ + 「飼う」*111
- ・ + 「入る」*40
- ・
- ・ 「飛び出す」と「出掛ける」の2種の複合動詞が同時に使用できていれば推定スコアは376/400となって母語話者相当



ICNALE 世界最大のアジア圏英語学習者コーパス

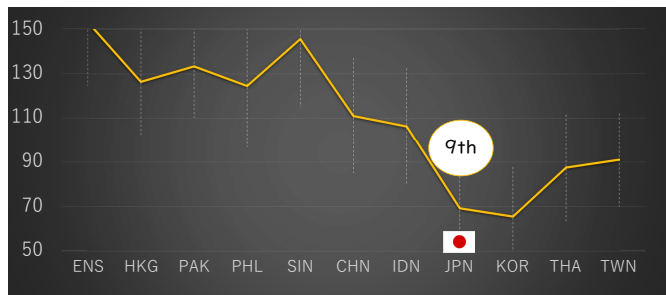
4,600+ participants / 10,000+ samples / 3.5M+ words



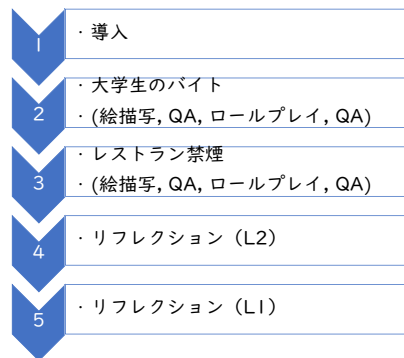
ICNALEの主要理念



I 分間発話量の国際比較 (BI_2)



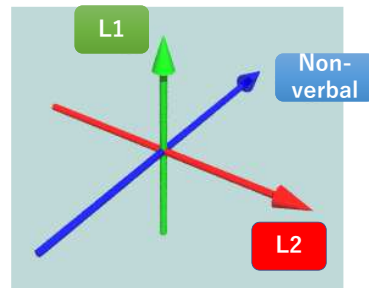
インタビューフロー



BI上級学習者のタスク別流暢性比較 (ENSを100として比率値に換算)



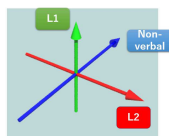
流暢性三元分析モデル(石川, 2019) 日本人学生は「英語が」流暢でないのか?



26

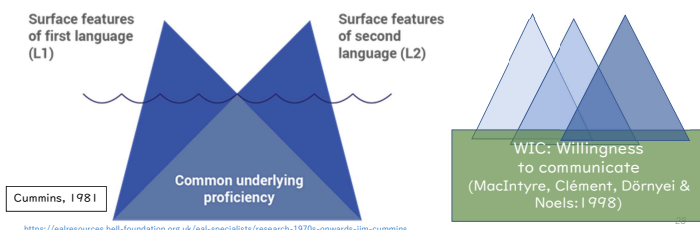
三元分析の結果(Ishikawa, 2020他)

- L1流暢性影響
- L2語数 TOEICスコア($r=.23$) < L1流暢性($r=.44$)
- 身体流暢性(ボディランゲージ使用度)影響
- L2語数 Hand($r=0.16$), Head($r=-0.15$)
- L2語彙多様性(TTR) Hand($r=-0.2$), Head($r=-0.24$)
- L2語彙難度(MWL) Hand($r=-0.43$), Head($r=-0.28$)
- ボディランゲージをL2の補償手段として使用している可能性



Cummins(1981)再訪:L2能力とL1能力

- L2能力とL1能力には基底部がある
- 基底部は「WIC」に大きく影響される
- Cumminsモデルにボディランゲージを組み込んだ拡張モデルの必要性



残された課題(1) NSモデル問題

- 1) 倫理的問題(NSの安易な理想化の危険性)
- Timmis(2013) ... why encumber learners with the 'luxury' items of NS spoken language? ... does teaching these features involve imposing a false identity on the learners?
- Prodromou(1998) What is real for NS...may be unreal for EFL learner in Greece...
- 2) 言語的問題(NSの多様性)
- Leech(1998) NSの地域・教養・年齢差は?/NS学生の作文は手本にならず、話し言葉はさらにひどい
- 「NSモデル問題」への(便宜的な)対応
- CIA分析モデル...CIA v2では、NS側データの多元化を要請
- ICNALE... NSデータを母語と職業の2観点で多元化
- I-JAS...今後、50人のNSデータの拡張が望まれる

残された課題(2) 他分野連携 —OPI専門家との協力の必要性—

- コーパス開発のためのインタビュー設計には、OPIの知見が不可欠
- 1) インタビューの流れの決定
- ウォームアップ→レベルチェック(できることの確認)→突き上げ(probe)(できないことの確認)→ワインドダウン
- 2) インタビューの内容の決定
- タスク=場面+話題+機能(山内, 2005, p.16)
- 3) 分析対象項目の決定(ibid, pp.44-50)
- 文で話せるか・やりとりができるか(初級/中級)
- 結束性があるか(中級/上級)
- 論拠・結論・多面性があるか(上級/超級)