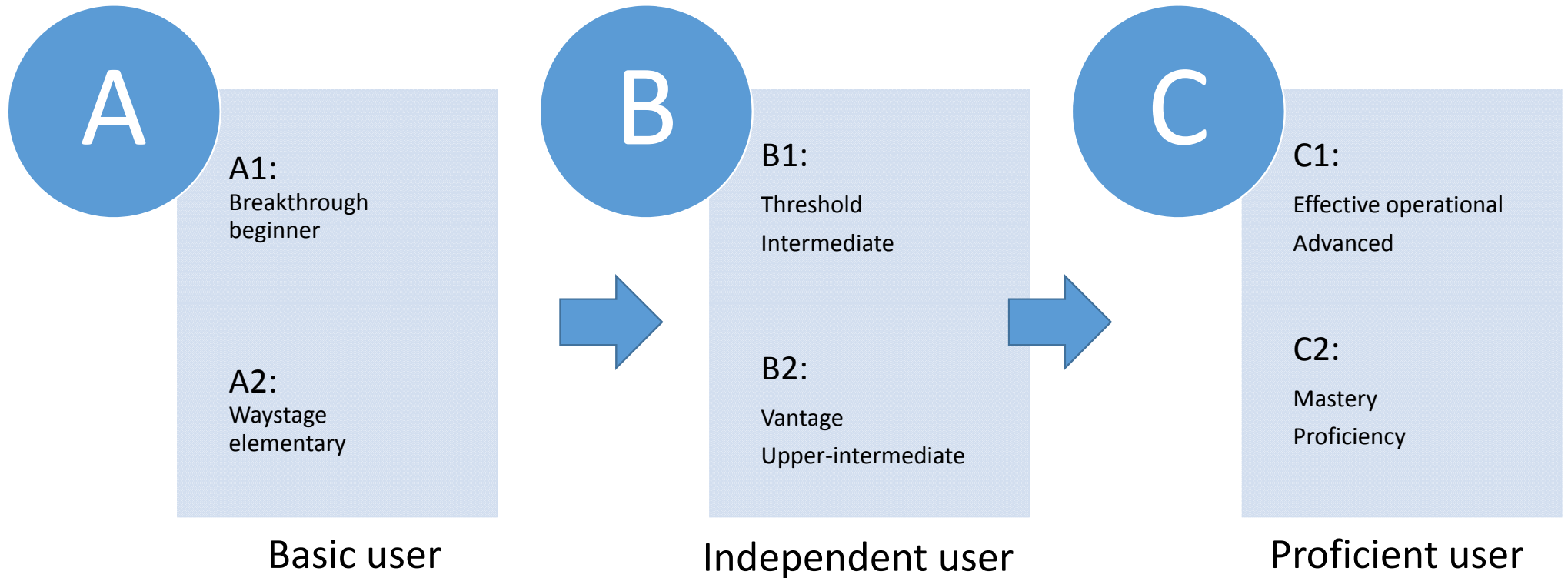


CEFR-Jに基づく英語口頭能力 テスト開発の必要性和可能性

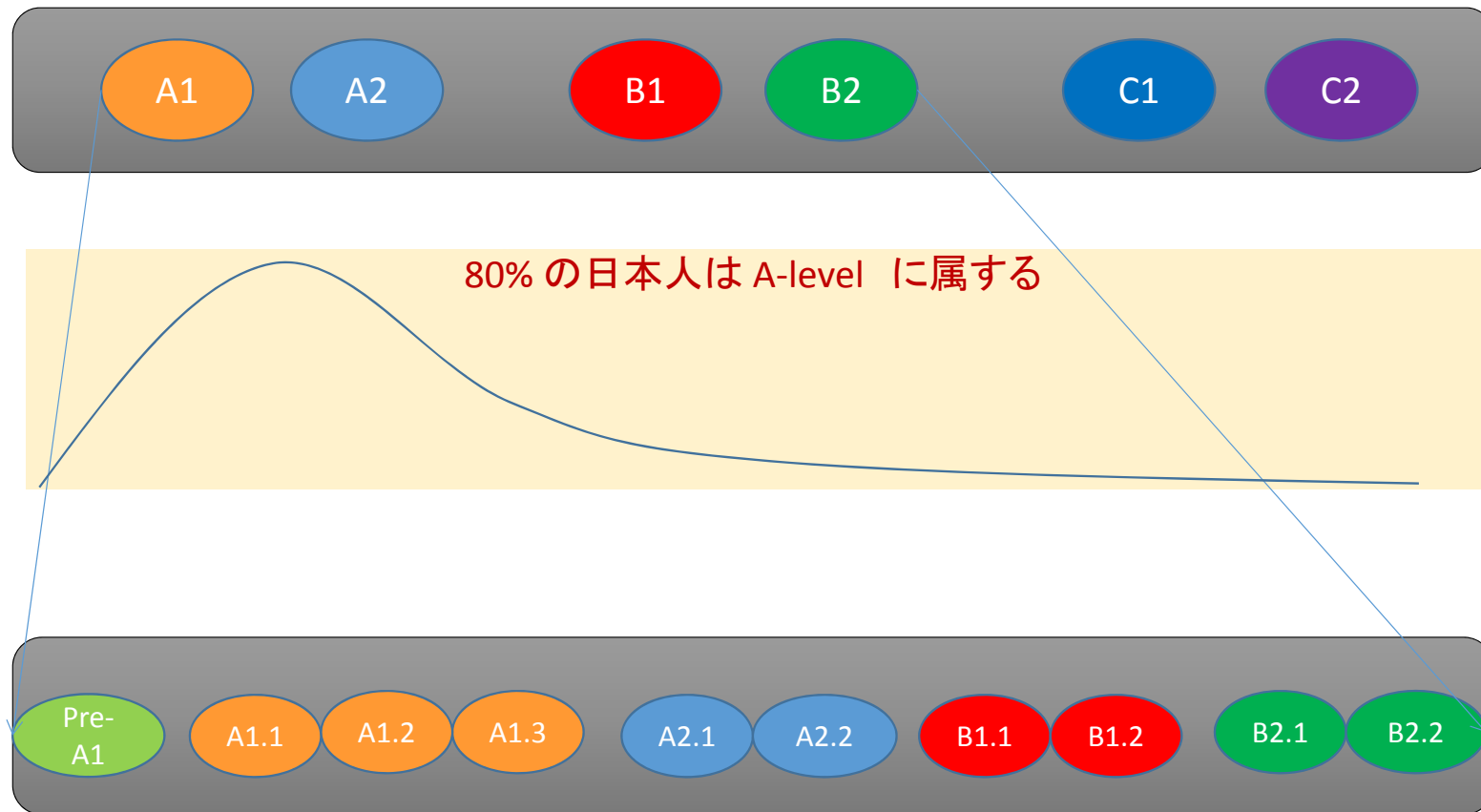
東京外国語大学

根岸雅史

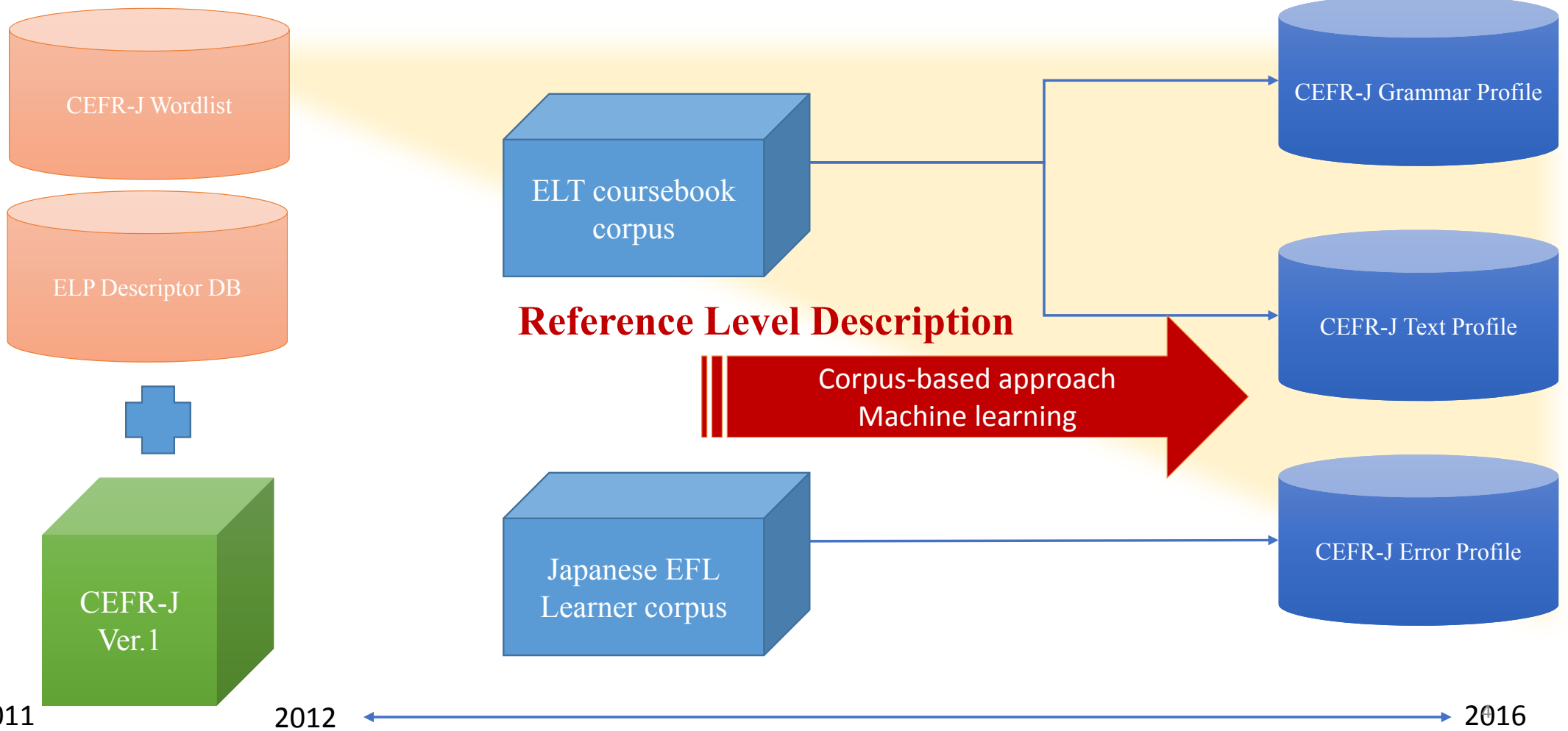
CEFR-levels



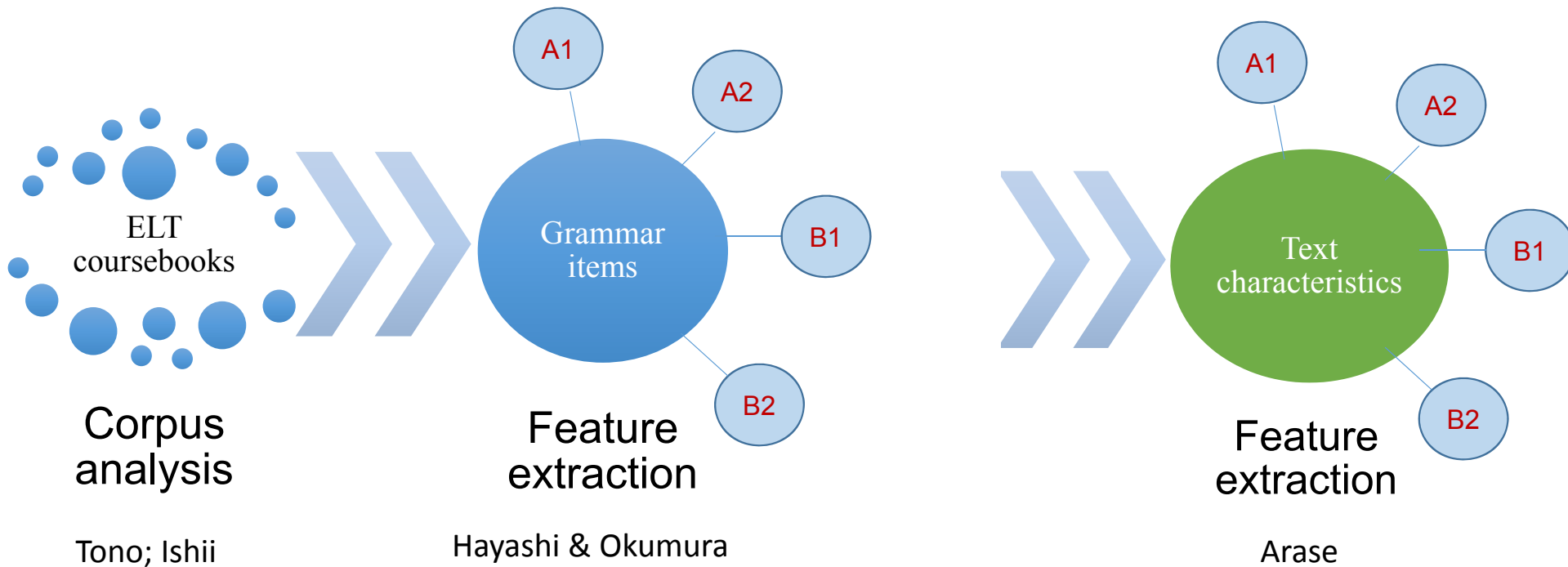
CEFR-J のレベル設定：12段階



The CEFR-J Project: Phase 2 (2012-2015)



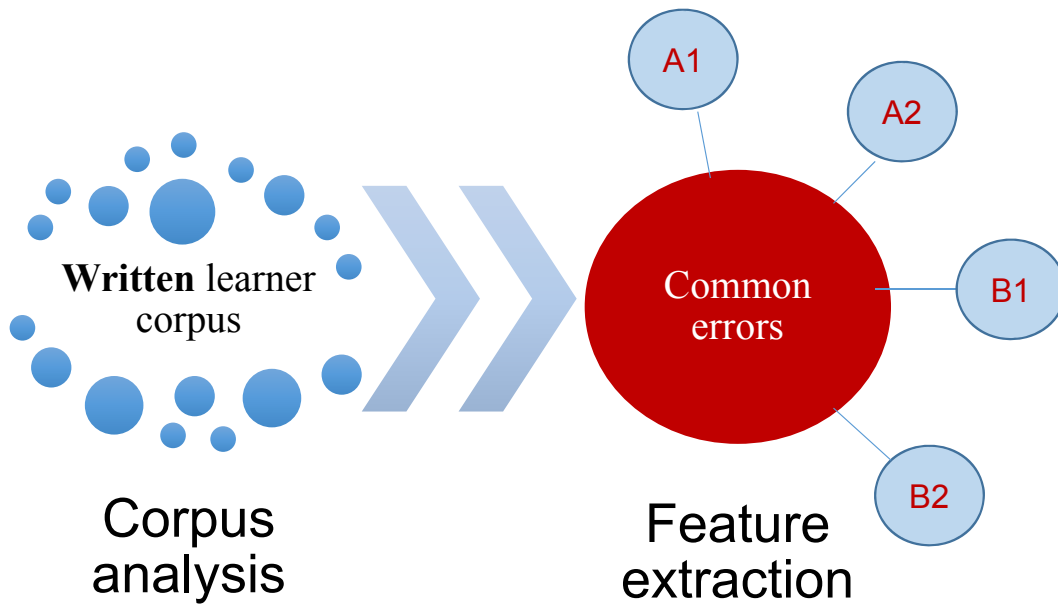
The process of the CEFR-J RLDs



Grammar Profile

Text Profile

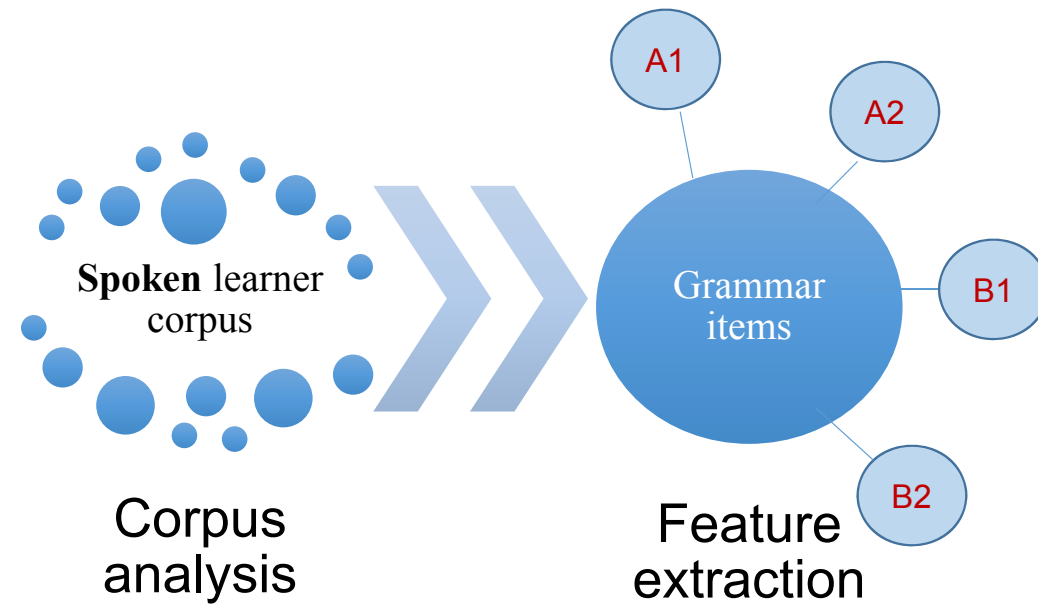
The process of the CEFR-J RLDs



Tono

Hayashi & Okumura

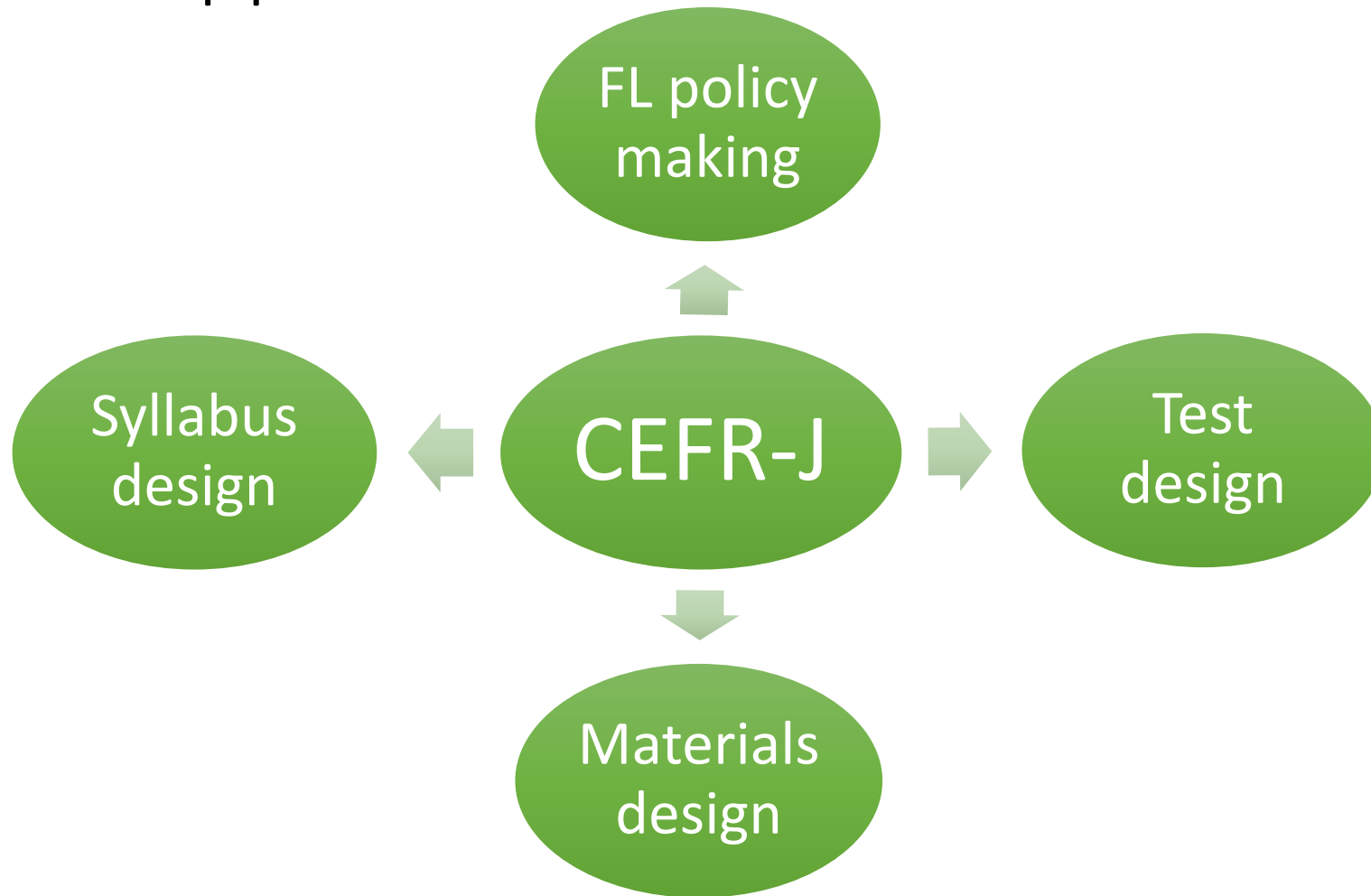
Error Profile



Brendan, Izumi, Kaneko
& Hirokawa

Learner Speech Profile

Possible applications



CEFR-Jプロジェクト

- CEFR-Jプロジェクトの成果
 - 「産出モード(スピーキング・ライティング)」における
 - 日本人英語学習者の発達段階に応じた「文法・語彙などの基準特性」
 - 「受容モード(リスニング・リーディング)」における
 - 「テキストの基準特性」
- 新規プロジェクト
 - 基準特性を基にした、CEFR-Jベースの指導タスクと評価テストの開発
 - 実証データをとることで、指導タスクと評価テストの妥当性を検証
 - 指導タスクと評価テストのアイテム・バンクを構築・公開

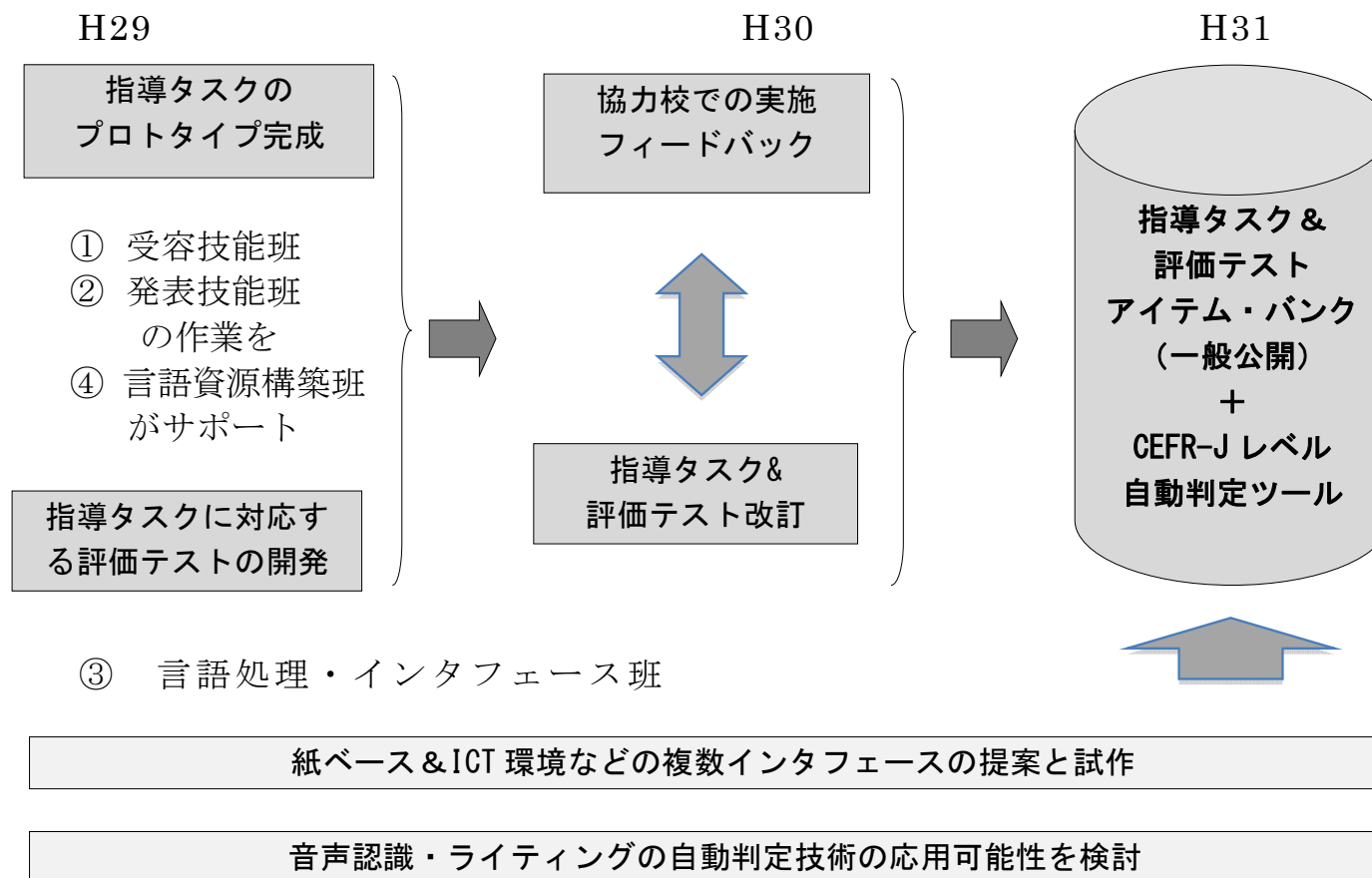
CEFR-J新規プロジェクト

- 指導タスク＋評価テストの開発
- 学際的なチーム
 - 言語教育学
 - 言語テスト
 - コーパス言語学
 - 自然言語処理
 - 音声認識

CEFR-J新規プロジェクト

- 教材作成: コーパス言語学や言語処理の専門家により導き出された言語資源を言語教育学の専門家が教材として具現化
- 言語テスト開発: 先の言語資源をもとに、
 - 言語教育学と言語テストの専門家が共同してテスト項目を作成
 - そのテスト・データをテスト統計の専門家が分析
 - こうして開発されたテストは、音声認識の専門家の力を借りることで、スピーキング・テストの自動採点までを実現
- 本研究により、4技能型の大学入試システムの開発に貢献

CEFR-J新規プロジェクト研究計画



英語の入試はようになる？

- センター試験の後継＋個別入試
- 英語4技能試験 情報サイト
<http://4skills.eiken.or.jp/education/innovation.html>
- 外部試験の利用
- 4技能の測定
- 4技能のバランス

【参考】各試験団体のデータによるCEFRとの 対照表

| CEFR | Cambridge English | 英検 | GTEC CBT | IELTS | TEAP | TOEFL iBT | TOEFL Junior Comprehensive | TOEIC / TOEIC S&W |
|------|-------------------|------------------|-----------|---------|---------|-----------|----------------------------|-----------------------------------|
| C2 | CPE (200+) | | | 8.5-9.0 | | | | |
| C1 | CAE (180-199) | 1級 (2810-3400) | 1400 | 7.0-8.0 | 400 | 95-120 | | 1305-1390 L&R 945~ S&W 360~ |
| B2 | FCE (160-179) | 準1級 (2596-3200) | 1250-1399 | 5.5-6.5 | 334-399 | 72-94 | 341-352 | 1095-1300 L&R 785~ S&W 310~ |
| B1 | PET (140-159) | 2級 (1780-2250) | 1000-1249 | 4.0-5.0 | 226-333 | 42-71 | 322-340 | 790-1090 L&R 550~ S&W 240~ |
| A2 | KET (120-139) | 準2級 (1635-2100) | 700-999 | 3.0 | 186-225 | | 300-321 | 385-785 L&R 225~ S&W 160~ |
| A1 | | 3級-5級 (790-1875) | -699 | 2.0 | | | | 200-380 L&R 120~ S&W 80~ |

平成26年度 英語教育改善のための英語力調査事業報告書

- 全国の**無作為抽出**による高等学校第3学年約**7万人**(約480校)を対象に、英語に関する**4技能**(読むこと、聞くこと、書くこと、話すこと)がバランスよく育成されているかという観点から本調査を実施し、生徒の英語力や英語の学習状況の把握・分析を通じて、学校における生徒への指導の充実や学習状況の改善に活用する。
- なお、本調査は、高校生の英語教育の多様性を踏まえ、世界標準に基づいて日本の高校生全体の英語力を測定するため、**CEFR**(Common European Framework of Reference for Languages:ヨーロッパ言語共通参照枠)を参照して測定することとした。
- A1～B2

平成26, 27年度英語力調査(高等学校)

【生徒全体のスコア分布(公立)】

<読むこと> 43問 (約45分)

<聞くこと> 36問 (約23分)

| Reading | | 平成26年度 | | 平成27年度 | | Listening | | 平成26年度 | | 平成27年度 | |
|---------|--------|--------|--------|--------|-------|-----------|-----|--------|-------|--------|-------|
| CEFR | 得点 | 人数 | 割合 | 人数 | 割合 | CEFR | 得点 | 人数 | 割合 | 人数 | 割合 |
| B2 | 320 | 14 | 0.0% | 30 | 0.1% | B2 | 320 | 55 | 0.1% | 123 | 0.2% |
| | 310 | 3 | | 14 | | 56 | | | | | |
| | 300 | 7 | | 35 | | 62 | | | | | |
| B1 | 290 | 11 | 1.2% | 41 | 2.0% | B1 | 290 | 29 | 1.2% | 77 | 2.1% |
| | 280 | 34 | | 51 | | 90 | | | | | |
| | 270 | 36 | | 73 | | 176 | | | | | |
| | 260 | 47 | | 122 | | 172 | | | | | |
| | 250 | 82 | | 175 | | 238 | | | | | |
| | 240 | 108 | | 250 | | 342 | | | | | |
| | 230 | 188 | | 347 | | 414 | | | | | |
| | 220 | 272 | | 503 | | 607 | | | | | |
| A2 | 210 | 404 | 23.5% | 730 | 29.9% | A2 | 210 | 444 | 20.3% | 751 | 24.2% |
| | 200 | 556 | | 1007 | | 1046 | | | | | |
| | 190 | 854 | | 1365 | | 1377 | | | | | |
| | 180 | 1204 | | 1957 | | 1770 | | | | | |
| | 170 | 1707 | | 2580 | | 2241 | | | | | |
| | 160 | 2367 | | 3648 | | 2835 | | | | | |
| | 150 | 3324 | | 5063 | | 3683 | | | | | |
| | 140 | 5031 | | 7144 | | 4700 | | | | | |
| A1 | 130 | 7989 | 75.3% | 9963 | 68.0% | A1 | 130 | 5268 | 78.4% | 6111 | 73.6% |
| | 120 | 11631 | | 12793 | | 7526 | | | | | |
| | 110 | 12398 | | 12821 | | 8718 | | | | | |
| | 100 | 7740 | | 9486 | | 8936 | | | | | |
| | 90 | 4663 | | 4891 | | 7788 | | | | | |
| | 80 | 1813 | | 2038 | | 6794 | | | | | |
| | 70 | 598 | | 696 | | 4289 | | | | | |
| | 60 | 206 | | 240 | | 2594 | | | | | |
| | 50 | 75 | | 105 | | 1299 | | | | | |
| | 40 | 50 | | 35 | | 642 | | | | | |
| | 30 | 18 | | 35 | | 331 | | | | | |
| | 20 | 1 | | 1 | | 144 | | | | | |
| | 10 | 0 | | 0 | | 147 | | | | | |
| 0 | 282 | 332 | 529 | | | | | | | | |
| 平均 | 126.7 | | 131.9 | | 平均 | 117.1 | | 120.7 | | | |
| 調査対象 | 65,711 | | 78,569 | | 調査対象 | 65,711 | | 78,569 | | | |

平成26, 27年度英語力調査(高等学校)

<書くこと> 2問 (約27分)

<話すこと> 3問 (対面約10分)

| Writing | | 平成26年度 | | 平成27年度 | |
|---------|--------|--------|--------|--------|-------|
| CEFR | 得点 | 人数 | 割合 | 人数 | 割合 |
| B2 | 140 | 0 | 0.0% | 0 | 0.0% |
| | 135 | 0 | | 1 | |
| | 130 | 0 | | 0 | |
| B1 | 125 | 2 | 0.3% | 2 | 0.7% |
| | 120 | 6 | | 18 | |
| | 115 | 10 | | 46 | |
| | 110 | 59 | | 179 | |
| A2 | 105 | 101 | 10.7% | 288 | 17.2% |
| | 100 | 306 | | 679 | |
| | 95 | 420 | | 726 | |
| | 90 | 829 | | 1,370 | |
| | 85 | 737 | | 1,577 | |
| | 80 | 1,465 | | 2,130 | |
| A1 | 75 | 1,525 | 89.1% | 3,515 | 82.1% |
| | 70 | 1,752 | | 3,563 | |
| | 65 | 1,668 | | 4,518 | |
| | 60 | 2,109 | | 3,709 | |
| | 55 | 1,876 | | 4,130 | |
| | 50 | 2,400 | | 3,651 | |
| | 45 | 2,039 | | 2,435 | |
| | 40 | 2,346 | | 3,208 | |
| | 35 | 1,940 | | 2,234 | |
| | 30 | 2,441 | | 2,668 | |
| | 25 | 2,045 | | 2,861 | |
| | 20 | 2,226 | | 3,551 | |
| | 15 | 2,151 | | 4,621 | |
| 10 | 2,529 | 12,844 | | | |
| 5 | 2,889 | 0 | | | |
| 0 | 29,973 | 14,303 | | | |
| 平均 | 24.9 | | 37.5 | | |
| 調査対象 | 65,904 | | 78,827 | | |
| 0点 | 20,059 | 30.4% | 14,303 | 18.1% | |

| Speaking | | 平成26年度 | | 平成27年度 | |
|----------|----|--------|-------|--------|-------|
| CEFR | 得点 | 人数 | 割合 | 人数 | 割合 |
| B1 | 14 | 166 | 1.0% | 211 | 1.2% |
| A2 | 13 | 193 | 9.5% | 239 | 9.8% |
| | 12 | 330 | | 390 | |
| | 11 | 418 | | 422 | |
| | 10 | 559 | | 611 | |
| A1 | 9 | 621 | 89.5% | 748 | 89.0% |
| | 8 | 718 | | 905 | |
| | 7 | 898 | | 1,026 | |
| | 6 | 1,143 | | 1,168 | |
| | 5 | 1,602 | | 1,569 | |
| | 4 | 1,085 | | 1,028 | |
| | 3 | 1,629 | | 1,601 | |
| | 2 | 1,444 | | 0 | |
| | 1 | 2,816 | | 3,918 | |
| | 0 | 2,210 | | 3,149 | |
| 平均 | | 4.2 | | 4.3 | |
| 調査対象 | | 15,032 | | 16,905 | |
| 0点 | | 2,210 | 14.0% | 3,149 | 18.5% |

平成27年度英語力調査(中学校)

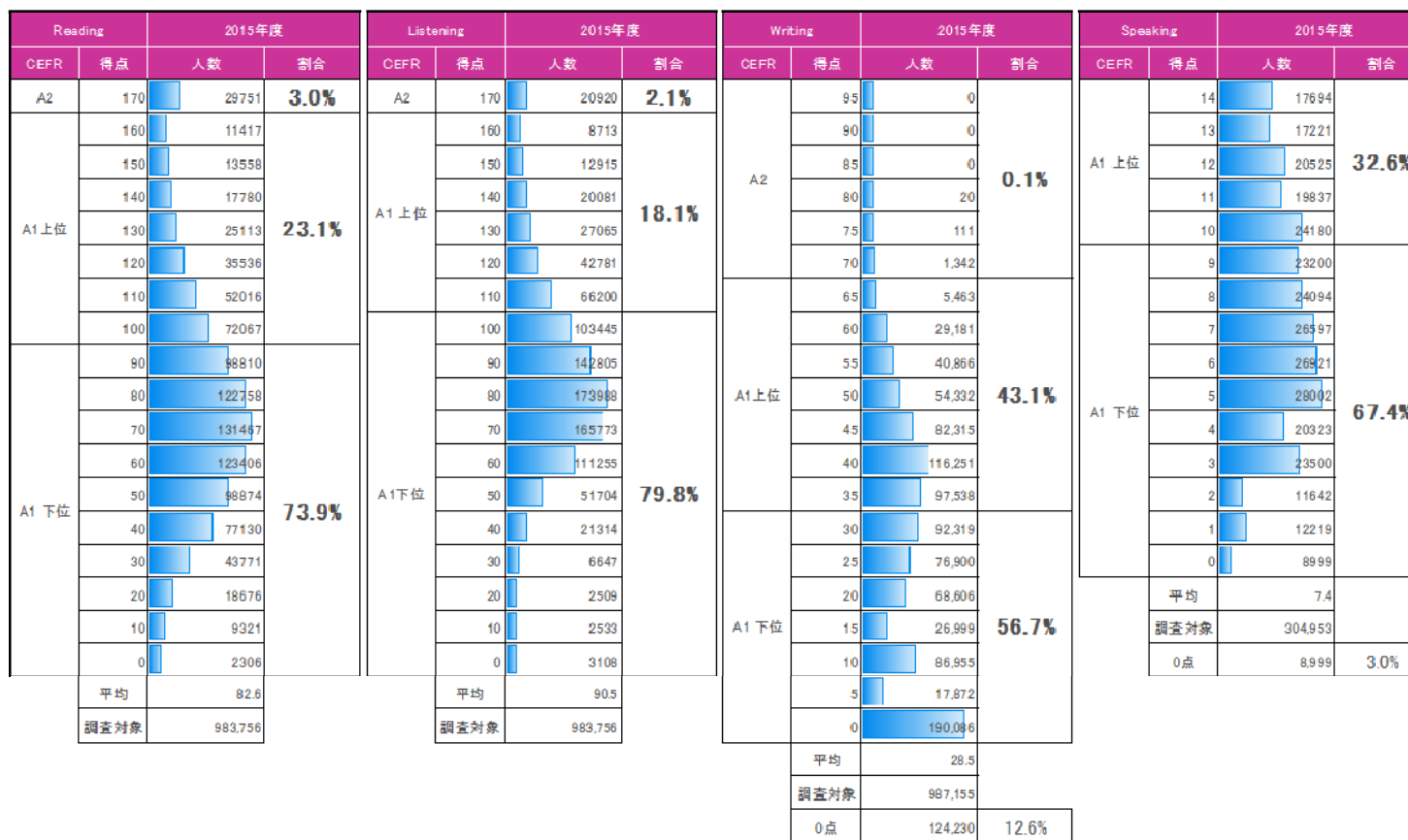
【生徒全体のスコア分布(公立)】

<読むこと> 28問 (約32分)

<聞くこと> 32問 (約18分)

<書くこと> 2問 (約25分)

<話すこと> 3問 (対面約10分)



【参考】各試験団体のデータによるCEFRとの対照表

| CEFR | Cambridge English | 英検 | GTEC CBT | IELTS | TEAP | TOEFL iBT | TOEFL Junior Comprehensive | TOEIC / TOEIC S&W |
|------|-------------------|------------------|-----------|---------|---------|-----------|----------------------------|-----------------------------------|
| C2 | CPE (200+) | | | 8.5-9.0 | | | | |
| C1 | CAE (180-199) | 1級 (2810-3400) | 1400 | 7.0-8.0 | 400 | 95-120 | | 1305-1390 L&R 945~ S&W 360~ |
| B2 | FCE (160-179) | 準1級 (2596-3200) | 1250-1399 | 5.5-6.5 | 334-399 | 72-94 | 341-352 | 1095-1300 L&R 785~ S&W 310~ |
| B1 | PET (140-159) | 2級 (1780-2250) | 1000-1249 | 4.0-5.0 | 226-333 | 42-71 | 322-340 | 790-1090 L&R 550~ S&W 240~ |
| A2 | KET (120-139) | 準2級 (1635-2100) | 700-999 | 3.0 | 186-225 | | 300-321 | 385-785 L&R 225~ S&W 160~ |
| A1 | | 3級-5級 (790-1875) | -699 | 2.0 | | | | 200-380 L&R 120~ S&W 80~ |

見えてきた:

大学入学希望者学力評価テスト(仮称)の「英語」の問題イメージ

- 独自開発?
- 「聞くこと」「読むこと」「書くこと」の三技能及び技能統合型合計で110分程度、別に「話すこと」10分程度のテスト構成・時間を想定
- CEFRのA1上位層からB1層を中心として測定が可能であり、CEFRにおけるレベルとの関連づけをすること
- 等化(IRTを含む。)

大規模テストにおけるスピーキング・テスト

- チャレンジ:
 - 50万人の受験者
 - 同日受験
 - 50万人の受験者の採点
 - 短い採点期間

- “... we cannot afford not to introduce a test with a powerful beneficial backwash effect. (Hughes, 2002)”

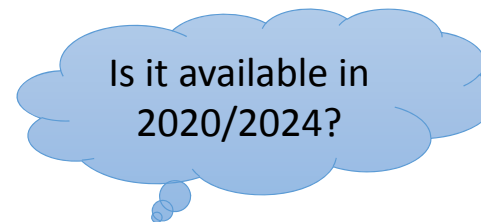
Potential solutions

-scoring-

- **By machine**

“Advances in automated speech recognition and computational linguistics have now made it possible for machines to score written and even spoken performance, especially when responses are more guided or controlled (Xi, 2012) The prospect is for more flexible automated scoring systems to emerge: systems that are capable of rating language elicited through different prompts without having to be trained for each individual task.”

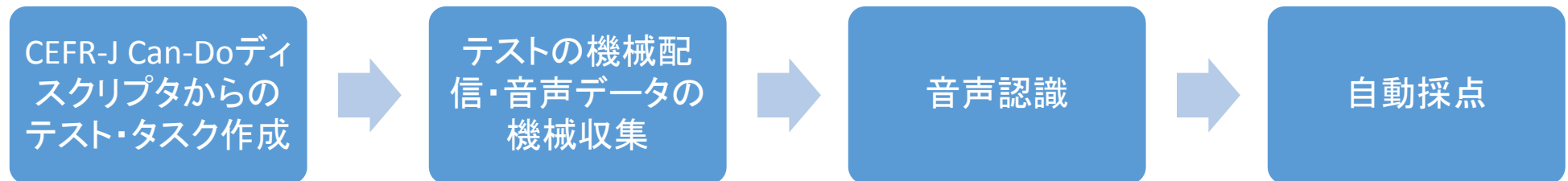
-Green (2014)



CEFR-J新規プロジェクト

- スピーキングやライティングの評価テスト
 - テストの機械配信
 - 音声認識をもとにした自動採点
 - 従来人間がやってきた採点も学習者の発達段階ごとの基準特性を利用して、個別の指導タスクに依存しない自動採点を実現する。

CEFR-J新規プロジェクト



課題と可能性

| 課題 | 可能性 |
|------------------------------|---------------------------------------|
| タスクの限定性 | 一斉実施が可能なので、多様なタスクの実施が可能 |
| 「やりとり」の限定性 パソコンに向かって話す違和感 | 試験官ロボット? |
| 音声認識の精度 | 認識可能な英語音声かどうかが最低基準になる? |
| 同時解答の問題 | ノイズ・キャンセリング? |
| 採点の信頼性 | 人間とほぼ同程度(0.8程度を実現) |
| 採点の妥当性 | 採点のプロセスの説明可能性が高い ただし、学習者・教師が解釈可能か? |
| コスト | 初期投資を官民共同? |

人による実施に対するメリット

| 観点 | メリット |
|----------------|----------------------|
| コスト | ランニング・コストが安い |
| 試験官・採点官のトレーニング | 不要 |
| 採点のモニタリング | 不要 |
| 採点期間 | 結果の即時性 |
| パフォーマンスの安定性 | 多様なタスクにより、パフォーマンスは安定 |

テストのインパクト

- “The research into washback suggests that teachers tend to pay close attention to the format of a test, rather than considering the underlying construct –the knowledge, skills or abilities targeted by the assessment. Just focussing on practice with the test format might not really help test takers to improve in the abilities being assessed (Green, 2007). A new spoken interview assessment based on personal information intended to assess impromptu conversation skills might, in practice, encourage memorisation of formulaic exchanges or short memorised speeches on obscure topics, even though these may be of rather limited value outside (or even inside) the examination room (Shohamy et al ., 1996).”

-Green. (2014)

- 教師や受験者に根底にあるテストの構成概念とその指導法を正しく伝える必要性