A Symposium on the CEFR-J RLD Project:
Developing Grammar, Text, and Error Profiles
Using Textbook & Learner Corpora

5-6 December, 2016
AGORA GLOBAL, TUFS
Funded by JSPS KAKENHI (Grant No. 24242017)
Acknowledgements

• JSPS KAKENHI (Grant No. 24242017)
• MEXT (Ministry of Education, Culture, Sports, Science and Technology)
• Tokyo University of Foreign Studies (TUFS)

• Sponsors:
  ◆ Pearson Education
  ◆ Cambridge University Press
  ◆ Cambridge English Language Assessment
  ◆ Oxford University Press
Programs: 4 keynote speeches

• Keynote speakers

Brian North
Paula Buttery
Programs

Day 1

Research reports
- CEFR-J Grammar Profile
- CEFR-J Text Profile
- CEFR-J Error Profile

Day 2

Teacher's workshop
- What is CEFR-J?
- How to use CEFR-J for assessment
- How to use profile information for teaching

Individual papers:
- Studies on criterial features
What is the CEFR-J?

• An adapted version of the Common European Framework of Reference for Languages (CEFR) for English language teaching in Japan

• Released in March, 2012

• More than 5,000 downloads by various institutes and individuals
• Has been used as a base framework for creating can-do descriptors for primary/secondary/tertiary schools
• Textbooks or dictionaries have started to include the information from the CEFR-J (e.g. descriptors, levels, wordlist)
What is the CEFR-J?

80% of university graduates belong to A-levels.

• KAKEN 2008-2011 (Leader: Yukio Tono)
  • Validation research:
    • Expert comments and revision
    • Large-scale self-assessment surveys using CEFR-J Beta:
      • 5,468 participants in total
        • 1,685 junior high school students
        • 2,538 senior high school students
        • 1,245 university students
      • Scaling descriptors by the method similar to the original CEFR
  • The version 1 was publicly released in March, 2012.

Preparation Phase
- Re-examining Can-do Descriptors
  - Teacher Survey

Revision Phase
- Expert Survey
  - Rewriting descriptors
  - CEFR-J alpha

Validation Phase
- Descriptor Sorting Exercise
  - Comparing Self-assessment & Actual Skills
  - Empirical Data & Final Revision
  - Students’Self-assessment
  - CEFR-J beta

Resources:
- Sample tasks
- CEFR-J Wordlist
- ELP Descriptor DB

CEFR-J Ver.1

Time:
- 2008
- 2009
- 2010
- 2011

CEFR-J Wordlist

ELP Descriptor DB

CEFR-J Ver. 1

2011

2012

ELT coursebook corpus

Japanese EFL Learner corpus

Reference Level Description

Corpus-based approach
Machine learning

CEFR-J Grammar Profile

CEFR-J Text Profile

CEFR-J Error Profile

2016
Reference Level Description: definition


- Main purpose: “to describe or transpose the Framework descriptors that characterise the competences of users/learners (at a given level) in terms of linguistic material specific to that language and considered necessary for the implementation of those competences.” (ibid: p.4)
Methodologies for RLD (Guide of RLD, 2005)

<table>
<thead>
<tr>
<th>Inventories of statistical frequency, including oral usage</th>
<th>The perusal of large bodies of texts</th>
<th>Knowledge considered as established in relation to the acquisition of the language concerned, in the “natural” and institutional environments, possibly in the form of acquisition sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners’ productions, particularly language data produced in evaluation situations (body of language examinations)</td>
<td>The approaches and concepts used to analyse discourse (including conversation analysis)</td>
<td>Established knowledge concerning spontaneous oral usages and their social variations</td>
</tr>
<tr>
<td>The intuitive and reflexive command of the language possessed by knowledgeable and highly competent informers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How to form a RLD team

RLD Team

1. automatic processing of language or information technology
2. assessment and certification of language skills
3. didactics of languages and cultures
4. linguistics, sociolinguists, speech analysis, lexicology, etc.
基盤研究（A）研究分担者
KAKEN CEFR-J RLD Project members

・ 代表：投野由紀夫（東京外国語大学）
・ 相川真佐夫（京都外国語大学）
  外国語教育政策, CEFR-J
・ 石井 康毅（成城大学）
  認知言語学, コーパス言語学
・ 和泉 絵美（同志社大学）
  自然言語処理, 学習者コーパス
・ 内田 諭（九州大学）
  認知言語学, 語用論
・ 奥村 学（東京工業大学）
  自然言語処理

・ 金子 恵美子（会津大学）
  スピーキング, 学習者コーパス
・ 寺内 一（高千穂大学）
  ESP, CEFR-J
・ 中谷 安男（法政大学）
  ESP, CEFR-J
・ 根岸 雅史（東京外国語大学）
  テスト理論, 評価, CEFR-J
・ 能登原祥之（同志社大学）
  語彙習得, 学習者コーパス

研究協力者：
廣川左千男（九州大学）
荒瀬 由紀（大阪大学）
The CEFR-J RLD project: features & aims

• To provide an inventory of the linguistic material necessary to implement the competences defined in the CEFR(-J) descriptors.

• To develop the descriptions using the CEFR(-J) descriptors and explain exactly how to pass from the descriptors to the inventories of forms.

• The descriptions should have a description of the approaches used to establish the inventories of forms.
  • To this end, we aimed to make our approach as transparent and replicable as possible.

• Indications as to whether the proposed forms should be known for reception only or also for production.
  • We prepared the inventories based on the ELT coursebook data as input and then have them accompanied with the information from learner corpora as output.
The process of the CEFR-J RLDs

**Grammar Profile**
- A1
- A2
- B1
- B2
- Tono; Ishii
- Hayashi & Okumura

**Text Profile**
- A1
- A2
- B1
- B2
- Arase
The process of the CEFR-J RLDs

**Error Profile**

- Written learner corpus
  - Corpus analysis: Tono
  - Feature extraction: Hayashi & Okumura

**Learner Speech Profile**

- Spoken learner corpus
  - Corpus analysis
  - Feature extraction: Brendan, Izumi, Kaneko & Hirokawa

- Grammar items
  - Feature extraction: A1, A2, B1, B2
Grammar Profile: output

- For each grammar item, the following information is provided:
  - Frequencies of A1 – C1 coursebooks (per 1 million)
  - Dispersion (range)
  - Tentative CEFR(-J) level based on FREQ * RANGE (Utility measure)
  - Core Inventory level
  - EGP level
  - GSE-LO level
  - SVM weights

<table>
<thead>
<tr>
<th>Grammar Item</th>
<th>Sentence Type</th>
<th>Word Forms</th>
<th>Freqs</th>
<th>Disp</th>
<th>CEFR</th>
<th>Core Inventory</th>
<th>EGP</th>
<th>GSE-LO</th>
<th>SVM weight</th>
<th>SVM weight</th>
<th>SVM weight</th>
<th>SVM weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODAL/AUX: be going to</td>
<td>AFF. DEC.</td>
<td>be going to AFF.</td>
<td>862</td>
<td>1,298</td>
<td>420</td>
<td>448</td>
<td>601</td>
<td>1</td>
<td>25</td>
<td>35</td>
<td>29</td>
<td>20</td>
</tr>
<tr>
<td>MODAL/AUX: be going to</td>
<td>NEG. DEC.</td>
<td>be going to NEG.</td>
<td>198</td>
<td>97</td>
<td>30</td>
<td>33</td>
<td>40</td>
<td>1</td>
<td>6</td>
<td>8</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>MODAL/AUX: be going to</td>
<td>AFF. INT.</td>
<td>be going to INT. AFF.</td>
<td>75</td>
<td>68</td>
<td>14</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>MODAL/AUX: be going to</td>
<td>NEG. DEC.</td>
<td>be going to INT. NEG.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MODAL/AUX: can</td>
<td>AFF. DEC.</td>
<td>can AFF.</td>
<td>3,336</td>
<td>2,577</td>
<td>3,123</td>
<td>2,840</td>
<td>2,402</td>
<td>17</td>
<td>21</td>
<td>26</td>
<td>23</td>
<td>18</td>
</tr>
<tr>
<td>MODAL/AUX: can</td>
<td>NEG. DEC.</td>
<td>can NEG.</td>
<td>864</td>
<td>39</td>
<td>111</td>
<td>62</td>
<td>59</td>
<td>3</td>
<td>6</td>
<td>56</td>
<td>26</td>
<td>6</td>
</tr>
<tr>
<td>MODAL/AUX: can</td>
<td>AFF. INT.</td>
<td>can INT. AFF.</td>
<td>906</td>
<td>667</td>
<td>419</td>
<td>206</td>
<td>208</td>
<td>14</td>
<td>17</td>
<td>20</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>MODAL/AUX: can</td>
<td>NEG. DEC.</td>
<td>can INT. NEG.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MODAL/AUX: could</td>
<td>AFF. DEC.</td>
<td>could AFF.</td>
<td>55</td>
<td>305</td>
<td>920</td>
<td>1,067</td>
<td>1,048</td>
<td>4</td>
<td>12</td>
<td>26</td>
<td>29</td>
<td>23</td>
</tr>
<tr>
<td>MODAL/AUX: could</td>
<td>NEG. DEC.</td>
<td>could NEG.</td>
<td>24</td>
<td>169</td>
<td>320</td>
<td>276</td>
<td>304</td>
<td>12</td>
<td>12</td>
<td>25</td>
<td>31</td>
<td>8</td>
</tr>
<tr>
<td>MODAL/AUX: could</td>
<td>AFF. INT.</td>
<td>could INT. AFF.</td>
<td>18</td>
<td>83</td>
<td>94</td>
<td>57</td>
<td>37</td>
<td>3</td>
<td>7</td>
<td>14</td>
<td>11</td>
<td>4</td>
</tr>
</tbody>
</table>
Possible applications

CEFR-J

FL policy making

Syllabus design

Test design

Materials design
Day 2: From research to pedagogy

- KN3: Automated marking for learner language (P. Buttery)
- KN4: What does the CEFR imply for teaching (B. North)
- The Global Scale of English (Mike Mayor, Pearson Education)
- Basic lectures:
  - What is the CEFR-J (Tono) (J)
  - Assessing with the CEFR-J (Negishi) (J)
  - How to use the GP (Tono) (J)
  - How to use the TP (Negishi) (J)
  - How to use the EP (Notohara) (J)
- How to use the CEFR-J Profile for speaking (Aikawa) (J) / writing (Nakatani) (J)
- Applications of Grammar profile for textbook analysis (Terauchi) (J)
- Aspects of criterial features:
  - Sentence patterns (Notohara)
  - Modal usage (Nakatani)
  - Information structure (Izumi)
  - Relative clauses (Takahashi)
  - Pragmatics (Miura)
  - Verb collocations (Uchida)
Enjoy the conference!

We’d appreciate your feedback or suggestions!