Corpus Linguistics (4): Learner corpora & error tagging

Yukio Tono (TUFS)
LCR and related fields

- What is a corpus?
- How can we use a corpus?

- How can we make sense of the data?
- What answer are we looking for?

- How can we apply corpus findings?
- How can we improve our teaching?
Three corpus-linguistic methods

1. Frequency lists & collocate lists
   - Most decontextualized methods

2. Colligations (Collostructions)
   - Lexical elements + grammatical element or structure

3. Concordances (of search expressions)
   - The occurrence of a match of the search expression
   - Most context-rich
Important concepts in CL (2)

- Frequency vs distribution (dispersion)
- Collocation (lexical n-grams; prefabs; multi-word units)
  - node vs collocate
  - N-word cluster
- Colligation
  - Lexico-grammatical co-occurrence
- Concordance
  - KWIC; Search [node] –word; right/left contexts; sorting;
Data relevant to SLA/ELT

- How does the input language pattern?
- How does the native language of the learners pattern?
- How does the target language pattern?
- What are the differences between how the native language and the target language pattern? (Contrastive/Cross-linguistic Analysis)
Data relevant to SLA/ELT

- How does the input language pattern?
  - Textbook corpus / Classroom interaction corpus

- How does the native language of the learners pattern?
  - L1 corpus

- How does the target language pattern?
  - TL corpus (e.g. English native corpus)

- What are the differences between how the native language and the target language pattern? (Contrastive/Cross-linguistic Analysis)
Important concepts in CL (3)

- General vs. Specialized corpora
- Spoken vs. Written corpora
- Balanced vs. Monitor corpora
Output of the learner

- How does the interlanguage pattern?
  - Learner corpora

- Which kinds of errors do the language learners commit?
  - Computer-aided error analysis (Granger)
Types of Learner Corpora

- Proficiency levels:
  - Fixed vs. varied (cross-sectional/longitudinal)

- L1 background:
  - Fixed vs. varied (learners with various L1s)

- Mode of production:
  - Written (essay)
  - Spoken (speech; retelling; conversation)

- Levels of annotation (POS; parsed; error-tagged)
Error Analysis (EA)

- Based on nativist views of language learning
  - Interlanguage (Selinker 1972)
  - Idiosyncratic dialect (Corder 1971)

- Basic steps:
  - Collection of a sample of learner language
  - Identification of errors
  - Description of errors
  - Explanation of errors
  - Error evaluation
Error descriptions 1

Linguistic taxonomy:
- Basic sentence structure
- Verb phrase (tense/ aspect/ subjunctive/ auxiliary/ non-finite verb)
- Verb complementation
- Noun phrase
- Prepositional phrase
- Adjunct
- Coordinate & subordinate constructions
- Sentence connection
Error description 2

- Surface structure (modification) taxonomy:
  - Omission
  - Addition
    - Regularization: e.g. *eated for ate
    - Double-marking: e.g. He didn’t *came
    - Simple addition: e.g. regularization/double-marking 以外
  - Misinformation
    - Regularization: e.g. *Do they be happy? \(\rightarrow\) Are they happy?
    - Archi-forms: e.g. It’s not *me. Me don’t care.（両方 me）
    - Alternating forms: e.g. Don’t watch. & No watch.
  - Misordering: e.g. She fights all the time her brother.
CL methods and LC

- Overuse vs. underuse

- Use vs. misuse (errors)
  - Linguistic classification of errors
    - Lexical vs. grammatical (POS + tense/agreement/etc)
  - Surface strategy taxonomy
    - Omissions/additions/misinformations/misorderings (Dulay, Burt & Krashen 1982)
SLA and CLR

- Description → Explanation
- SLA theories:
  - UG-Based SLA (Hawkins, White)  left more focus on lexicon
  - Processibility Hypothesis (Pienemann)  left Levelt & LFG
  - Competition Model (MacWhinney)  left very much frequency-based
- Related disciplines:
  - Cognitive linguistics; Usage-based approach
  - Systemic-functional grammar
  - Natural language processing
  - Data mining; Neural network
LCR & ELT applications

- **Indirect use:**
  - Lexicography
  - Wordlist
  - Syllabus/course design
  - Materials design (textbooks; vocabulary books; classroom tasks)
  - Test development (CEFR; Criterion; SST)

- **Direct use:**
  - Corpus use in the classroom
  - Data driven learning
  - CALL implementations
  - Teacher training
Main areas to be covered

- State-of-the-art articles in LCR
- Error annotation
- CEFR-based LCR
- Automatic detection of errors using LC
- Applications of LCR in iCALL
- Spoken vs. Written LC
Error tagging

- Annotation on language learners’ errors
- Error-tagged corpora:
  - NICT JLE Corpus: partially error-tagged
  - JEFLL Corpus: partially error-tagged
  - Cambridge Learner Corpus
  - HKUST Corpus of Learner English
- Generic error tagsets:
  - NICT JLE/ICLE
- Tagging is usually done manually
Learner corpus projects in Japan

- **NICT JLE Corpus (Izumi et al. 2005)**
  - 2 million words
  - Spoken
    (based on the OPI-like interview scripts)
  - 1,283 subjects
  - Distributed by NICT

- **JEFLL Corpus (Tono et al. 2007)**
  - 669,281 words
  - Written in-class essays (w/o dictionary)
  - 10,038 subjects (junior & senior high)
  - Freely accessible on the web:
    - http://scn02.corpora.jp/~jefll04dev/
L2 vocabulary profile: Crucial differences

Most learner corpora available now: e.g. ICLE/ LLC/ CLC

- advanced
- intermediate
- novice / lower-intermediate

NICT JLE
JEFLL
Systematizing LC descriptions

A series of studies on criterial features of L2 developmental stages based on JEFLL and NICT JLE:

- Morpheme orders: Tono (1998); Izumi (2005)
WORDLIST ANALYSIS
Use of top 100 words (JEFLLL)
Tokyo University of Foreign Studies

Distributions of top 100 words (JEFLL)
Overuse/underuse of words in top 10

<table>
<thead>
<tr>
<th>J1</th>
<th>J2</th>
<th>J3</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>BNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>Freq</td>
<td>Word</td>
<td>Freq</td>
<td>Word</td>
<td>Freq</td>
<td>Word</td>
</tr>
<tr>
<td>I</td>
<td>7071</td>
<td>I</td>
<td>7235</td>
<td>I</td>
<td>7513</td>
<td>I</td>
</tr>
<tr>
<td>IS</td>
<td>2412</td>
<td>AND</td>
<td>2350</td>
<td>AND</td>
<td>2624</td>
<td>THE</td>
</tr>
<tr>
<td>AND</td>
<td>2328</td>
<td>VERY</td>
<td>2101</td>
<td>THE</td>
<td>2350</td>
<td>AND</td>
</tr>
<tr>
<td>VERY</td>
<td>2293</td>
<td>TO</td>
<td>2020</td>
<td>TO</td>
<td>2344</td>
<td>TO</td>
</tr>
<tr>
<td>BUT</td>
<td>1755</td>
<td>THE</td>
<td>1979</td>
<td>WAS</td>
<td>1983</td>
<td>IS</td>
</tr>
<tr>
<td>MY</td>
<td>1743</td>
<td>IS</td>
<td>1898</td>
<td>WAS</td>
<td>1893</td>
<td>MY</td>
</tr>
<tr>
<td>THE</td>
<td>1606</td>
<td>MY</td>
<td>1730</td>
<td>MY</td>
<td>1597</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>1581</td>
<td>MY</td>
<td>1739</td>
<td>A</td>
<td>1641</td>
<td>IT</td>
</tr>
<tr>
<td>LIKE</td>
<td>1268</td>
<td>BUT</td>
<td>1666</td>
<td>BUT</td>
<td>1627</td>
<td>IS</td>
</tr>
<tr>
<td>TO</td>
<td>1245</td>
<td>A</td>
<td>1573</td>
<td>VERY</td>
<td>1585</td>
<td>VERY</td>
</tr>
</tbody>
</table>

freq = per 100,000 words
COLLOCATION/COLLIGATION ANALYSIS
The use of “make + Noun”
L2 vocabulary profile: LC perspectives

Structural complexities:
- have + P.P.
- have + NP
- have + NP + V
- have + NP + Ving

Semantic complexities:
- have + NP:
  - abstract N delexical use
  - concrete N

Profiling based on NS corpora only

Learner corpora

LC descriptions needed to describe the gap between NS and NNS performances
Use of the verb *have*

- Underuse: `have + p.p.`
- Slightly overuse: `have + noun`

**Normalized freq. (per 1 million)**
<table>
<thead>
<tr>
<th>BNC Top 10</th>
<th>NS</th>
<th>JEFLL-JH</th>
<th>JEFLL-SH</th>
</tr>
</thead>
<tbody>
<tr>
<td>have + time</td>
<td>*****</td>
<td>******************</td>
<td>******************</td>
</tr>
<tr>
<td>have + right(s)</td>
<td>*****</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>have + problem</td>
<td>****</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>have + effect</td>
<td>****</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>have + look</td>
<td>****</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>have + child</td>
<td>***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>have + idea</td>
<td>***</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>have + chance</td>
<td>***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>have + place</td>
<td>***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>have + power</td>
<td>***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>JEFL Top 10</td>
<td>NS</td>
<td>JEFL-JH</td>
<td>JEFL-SH</td>
</tr>
<tr>
<td>------------------</td>
<td>----</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>have + breakfast</td>
<td>*</td>
<td>********</td>
<td>********</td>
</tr>
<tr>
<td></td>
<td></td>
<td>********</td>
<td></td>
</tr>
<tr>
<td>have + bread</td>
<td>-</td>
<td>********</td>
<td>********</td>
</tr>
<tr>
<td></td>
<td></td>
<td>********</td>
<td></td>
</tr>
<tr>
<td>have + rice</td>
<td>-</td>
<td>********</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>********</td>
<td></td>
</tr>
<tr>
<td>have + time</td>
<td>*****</td>
<td>********</td>
<td>*****</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>have + money</td>
<td>**</td>
<td>********</td>
<td>********</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>have + dream</td>
<td>-</td>
<td>********</td>
<td>********</td>
</tr>
<tr>
<td></td>
<td></td>
<td>********</td>
<td></td>
</tr>
<tr>
<td>have + food</td>
<td>-</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>have + lunch</td>
<td>*</td>
<td>****</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>have + break</td>
<td>*</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>have + thing</td>
<td>*</td>
<td>**</td>
<td>*</td>
</tr>
</tbody>
</table>
have + noun (3)

- Fewer abstract nouns
- More concrete objects

Very little use of delexical verbs (e.g. have a look)

- rice
- time
- money
- idea
- bread
L2 vocabulary profile: Dealing with the gap more efficiently

have

I have sth

I have to

SH1-2

have sb do/p.p.

JH1...

JH2...

JH3

CONCRETE

ABSTRACT/ DELEXICAL

USE OF PERFECT TENSE
WORD/POS N-GRAM ANALYSIS
**Note:** Sentence-initial “but” in blue

<table>
<thead>
<tr>
<th>Freq</th>
<th>Seq</th>
<th>J1</th>
<th>Freq</th>
<th>Seq</th>
<th>J2</th>
<th>Freq</th>
<th>Seq</th>
<th>J3</th>
<th>Freq</th>
<th>Seq</th>
<th>S1</th>
<th>Freq</th>
<th>Seq</th>
<th>S2</th>
<th>Freq</th>
<th>Seq</th>
<th>S3</th>
</tr>
</thead>
<tbody>
<tr>
<td>578</td>
<td>#JP#</td>
<td>426</td>
<td>. So</td>
<td>844</td>
<td>I don't</td>
<td>224</td>
<td>I don't</td>
<td>600</td>
<td>I don't</td>
<td>235</td>
<td>I don't</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>441</td>
<td>I like</td>
<td>425</td>
<td>I don't</td>
<td>529</td>
<td>. So</td>
<td>170</td>
<td>#JP#</td>
<td>1</td>
<td>450</td>
<td>I want to</td>
<td>188</td>
<td>I want to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>417</td>
<td>But</td>
<td>408</td>
<td>#JP#</td>
<td>527</td>
<td>But</td>
<td>166</td>
<td>. But</td>
<td>346</td>
<td>. But</td>
<td>166</td>
<td>. So</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>408</td>
<td>I bring</td>
<td>425</td>
<td>#JP#</td>
<td>184</td>
<td>So</td>
<td>345</td>
<td>So</td>
<td>141</td>
<td>So</td>
<td>141</td>
<td>. So</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>347</td>
<td>. So</td>
<td>347</td>
<td>. So</td>
<td>371</td>
<td>I was</td>
<td>153</td>
<td>I was</td>
<td>295</td>
<td>It is</td>
<td>131</td>
<td>It was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>335</td>
<td>#JP#</td>
<td>335</td>
<td>. I like</td>
<td>335</td>
<td>I will bring</td>
<td>147</td>
<td>I want to</td>
<td>291</td>
<td>I can't</td>
<td>126</td>
<td>. It is</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>268</td>
<td>I will bring</td>
<td>275</td>
<td>I was</td>
<td>336</td>
<td>I want to</td>
<td>139</td>
<td>school festival</td>
<td>281</td>
<td>I think</td>
<td>122</td>
<td>in the morning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>#JP#</td>
<td>255</td>
<td>I will</td>
<td>327</td>
<td>I like</td>
<td>130</td>
<td>It was</td>
<td>274</td>
<td>I was</td>
<td>118</td>
<td>and so on</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>in the morning</td>
<td>233</td>
<td>I want to</td>
<td>283</td>
<td>I will</td>
<td>111</td>
<td>our school festival</td>
<td>273</td>
<td>I can't</td>
<td>116</td>
<td>I can't</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>in the morning</td>
<td>217</td>
<td>I want to</td>
<td>283</td>
<td>I will</td>
<td>111</td>
<td>our school festival</td>
<td>273</td>
<td>I can't</td>
<td>116</td>
<td>I can't</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>189</td>
<td>very much</td>
<td>193</td>
<td>the morning</td>
<td>264</td>
<td>And</td>
<td>109</td>
<td>in the morning</td>
<td>239</td>
<td>a lot of</td>
<td>105</td>
<td>school festival</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>178</td>
<td>the morning</td>
<td>170</td>
<td>very much</td>
<td>263</td>
<td>I can't</td>
<td>103</td>
<td>a lot of</td>
<td>237</td>
<td>and so on</td>
<td>105</td>
<td>so on</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>177</td>
<td>Our class</td>
<td>160</td>
<td>I was</td>
<td>254</td>
<td>I do</td>
<td>99</td>
<td>So</td>
<td>235</td>
<td>#JP#</td>
<td>103</td>
<td>Our school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>175</td>
<td>Our school festival</td>
<td>159</td>
<td>#JP#</td>
<td>228</td>
<td>I think</td>
<td>88</td>
<td>I was</td>
<td>226</td>
<td>And</td>
<td>101</td>
<td>our school festival</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>It's</td>
<td>152</td>
<td>very #JP#</td>
<td>218</td>
<td>I think</td>
<td>90</td>
<td>I will bring</td>
<td>224</td>
<td>But</td>
<td>98</td>
<td>a lot of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>173</td>
<td>very #JP#</td>
<td>151</td>
<td>So</td>
<td>212</td>
<td>the morning</td>
<td>88</td>
<td>And</td>
<td>221</td>
<td>If</td>
<td>95</td>
<td>I think</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>171</td>
<td>I will</td>
<td>149</td>
<td>It is</td>
<td>209</td>
<td>Because</td>
<td>88</td>
<td>I think</td>
<td>216</td>
<td>so on</td>
<td>94</td>
<td>I was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>161</td>
<td>Our school</td>
<td>148</td>
<td>It was</td>
<td>196</td>
<td>don't have</td>
<td>87</td>
<td>Because</td>
<td>215</td>
<td>I do</td>
<td>92</td>
<td>#JP# is</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>158</td>
<td>I do</td>
<td>143</td>
<td>#JP#</td>
<td>195</td>
<td>I was very</td>
<td>87</td>
<td>Our school festival</td>
<td>210</td>
<td>I have</td>
<td>92</td>
<td>I will bring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>154</td>
<td>and #JP#</td>
<td>142</td>
<td>I want</td>
<td>189</td>
<td>very much</td>
<td>88</td>
<td>I like</td>
<td>210</td>
<td>our school festival</td>
<td>92</td>
<td>our school festival</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>I usually</td>
<td>138</td>
<td>#JP#</td>
<td>185</td>
<td>It is</td>
<td>84</td>
<td>Our school</td>
<td>208</td>
<td>school festival</td>
<td>91</td>
<td>the morning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>I usually</td>
<td>138</td>
<td>#JP#</td>
<td>185</td>
<td>It is</td>
<td>84</td>
<td>Our school</td>
<td>208</td>
<td>school festival</td>
<td>91</td>
<td>the morning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>131</td>
<td>#JP#</td>
<td>127</td>
<td>I usually</td>
<td>148</td>
<td>He was</td>
<td>86</td>
<td>So</td>
<td>177</td>
<td>I want</td>
<td>84</td>
<td>If</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>129</td>
<td>morning</td>
<td>126</td>
<td>Urashima Taro</td>
<td>147</td>
<td>If</td>
<td>87</td>
<td>a big earthquake</td>
<td>170</td>
<td>a big earthquake</td>
<td>83</td>
<td>I want</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>I'm</td>
<td>126</td>
<td>I usually have</td>
<td>147</td>
<td>So</td>
<td>66</td>
<td>don't have</td>
<td>168</td>
<td>Because</td>
<td>80</td>
<td>And</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>I have</td>
<td>123</td>
<td>Our school</td>
<td>144</td>
<td>#JP#</td>
<td>66</td>
<td>too</td>
<td>167</td>
<td>but</td>
<td>78</td>
<td>I have</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>too</td>
<td>122</td>
<td>And</td>
<td>125</td>
<td>I didn't</td>
<td>65</td>
<td>school festival</td>
<td>184</td>
<td>too</td>
<td>71</td>
<td>I will</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>I #JP#</td>
<td>122</td>
<td>But</td>
<td>138</td>
<td>One day</td>
<td>61</td>
<td>so</td>
<td>150</td>
<td>so</td>
<td>68</td>
<td>Urashima Taro</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>Urashima Taro</td>
<td>118</td>
<td>Our school festival</td>
<td>134</td>
<td>a lot of</td>
<td>81</td>
<td>the school festival</td>
<td>152</td>
<td>I think that</td>
<td>65</td>
<td>I could not</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>I want</td>
<td>117</td>
<td>It's</td>
<td>133</td>
<td>And he</td>
<td>60</td>
<td>I have</td>
<td>148</td>
<td>Our school festival</td>
<td>63</td>
<td>I usually</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>#JP#</td>
<td>115</td>
<td>Urashima Taro was</td>
<td>126</td>
<td>.</td>
<td>60</td>
<td>I want</td>
<td>143</td>
<td>and</td>
<td>63</td>
<td>very much</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>#JP#</td>
<td>115</td>
<td>Urashima Taro was</td>
<td>126</td>
<td>.</td>
<td>60</td>
<td>I want</td>
<td>143</td>
<td>and</td>
<td>63</td>
<td>very much</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>is very #JP#</td>
<td>110</td>
<td>a lot of</td>
<td>125</td>
<td>I'm</td>
<td>59</td>
<td>#JP#</td>
<td>59</td>
<td>#JP#</td>
<td>59</td>
<td>#JP#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>I usually have</td>
<td>106</td>
<td>#JP#</td>
<td>125</td>
<td>So he</td>
<td>58</td>
<td>I could not</td>
<td>142</td>
<td>I will</td>
<td>59</td>
<td>He was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>#JP# and #JP#</td>
<td>106</td>
<td>I can't</td>
<td>123</td>
<td>#JP#</td>
<td>58</td>
<td>and #JP#</td>
<td>137</td>
<td>for breakfast</td>
<td>59</td>
<td>It's</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>I'm</td>
<td>100</td>
<td>I</td>
<td>114</td>
<td>dream</td>
<td>62</td>
<td>but</td>
<td>136</td>
<td>For example</td>
<td>58</td>
<td>Because</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>And</td>
<td>101</td>
<td>Then</td>
<td>111</td>
<td>morning</td>
<td>55</td>
<td>It is</td>
<td>127</td>
<td>It's</td>
<td>57</td>
<td>When</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>rice and #JP#</td>
<td>98</td>
<td>I think</td>
<td>111</td>
<td>morning</td>
<td>55</td>
<td>It is</td>
<td>127</td>
<td>It's</td>
<td>57</td>
<td>When</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>school festival is</td>
<td>97</td>
<td>it</td>
<td>109</td>
<td>and so on</td>
<td>54</td>
<td>#JP#</td>
<td>54</td>
<td>#JP#</td>
<td>54</td>
<td>#JP#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>#JP#</td>
<td>97</td>
<td>very happy</td>
<td>107</td>
<td>don't like</td>
<td>54</td>
<td>I went to</td>
<td>126</td>
<td>I'm</td>
<td>56</td>
<td>I was very</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>#JP# is #JP#</td>
<td>95</td>
<td>I went to</td>
<td>106</td>
<td>I will</td>
<td>54</td>
<td>very much</td>
<td>124</td>
<td>So</td>
<td>55</td>
<td>for me</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>do like</td>
<td>96</td>
<td>#JP#</td>
<td>106</td>
<td>I was</td>
<td>54</td>
<td>very much</td>
<td>124</td>
<td>So</td>
<td>55</td>
<td>for me</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>do like</td>
<td>96</td>
<td>#JP#</td>
<td>106</td>
<td>I was</td>
<td>54</td>
<td>very much</td>
<td>124</td>
<td>So</td>
<td>55</td>
<td>for me</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>#JP# in the</td>
<td>93</td>
<td>I have</td>
<td>104</td>
<td>so</td>
<td>51</td>
<td>I will</td>
<td>118</td>
<td>For example</td>
<td>51</td>
<td>#JP#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>I like #JP#</td>
<td>92</td>
<td>#JP# and #JP#</td>
<td>102</td>
<td>breakfast</td>
<td>63</td>
<td>Urashima Taro</td>
<td>115</td>
<td>the morning</td>
<td>50</td>
<td>And he</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>bread and milk</td>
<td>92</td>
<td>#JP#</td>
<td>101</td>
<td>Then</td>
<td>52</td>
<td>very happy</td>
<td>112</td>
<td>there is a</td>
<td>50</td>
<td>One day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J1</td>
<td>J2</td>
<td>J3</td>
<td>S1</td>
<td>S2</td>
<td>S3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1161</td>
<td>PPIS1 VV0</td>
<td>036.</td>
<td>PPIS1 VV0</td>
<td>118.</td>
<td>PPIS1 VV0</td>
<td>0115</td>
<td>PPIS1 VV0</td>
<td>1315</td>
<td>PPIS1 VV0</td>
<td>492</td>
<td>NN1 , PPIS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>692</td>
<td>PPIS1 VV0 NN1</td>
<td>719</td>
<td>NN1 , PPIS1</td>
<td>1033</td>
<td>PPIS1 VV0</td>
<td>413</td>
<td>PPIS1 VM VV</td>
<td>1259</td>
<td>PPIS1 VV0</td>
<td>469</td>
<td>PPIS1 VM VV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>326</td>
<td>AT PPIS1</td>
<td>673.</td>
<td>PPIS1 VM VV</td>
<td>519.</td>
<td>AT PPIS1</td>
<td>378.</td>
<td>AT PPIS1 VM VV</td>
<td>1148</td>
<td>AT PPIS1 VM VV</td>
<td>421.</td>
<td>AT PPIS1 VM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>447</td>
<td>APPGE NN1</td>
<td>677.</td>
<td>RO JJ</td>
<td>750</td>
<td>RO JJ</td>
<td>362</td>
<td>APPGE NN1</td>
<td>963</td>
<td>AT NN1</td>
<td>417</td>
<td>AT NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>529</td>
<td>J# PPIS1</td>
<td>542.</td>
<td>APPGE NN1</td>
<td>699.</td>
<td>JJ NN1</td>
<td>341</td>
<td>AT NN1</td>
<td>934</td>
<td>JJ NN1</td>
<td>414</td>
<td>JJ NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>426</td>
<td>RO JJ</td>
<td>449.</td>
<td>JJ NN1</td>
<td>745</td>
<td>JPPIS1</td>
<td>333</td>
<td>NN1 NN1</td>
<td>809</td>
<td>APPGE NN1</td>
<td>375</td>
<td>NN1 NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>424</td>
<td>CCB PPIS1</td>
<td>481.</td>
<td>RO JJ</td>
<td>816</td>
<td>AT NN1</td>
<td>322</td>
<td>RO JJ</td>
<td>783</td>
<td>AT NN1</td>
<td>328</td>
<td>AT NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>APPGE NN1</td>
<td>429.</td>
<td>PPIS1 VD0 XX</td>
<td>587</td>
<td>PPIS1</td>
<td>319</td>
<td>APPGE NN1</td>
<td>322</td>
<td>APPGE NN1</td>
<td>322</td>
<td>APPGE NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>401</td>
<td>PPIS1 VM VV</td>
<td>410.</td>
<td>PPIS1 VV0</td>
<td>583</td>
<td>VD0 XX VV</td>
<td>302</td>
<td>VBDZ RG J</td>
<td>722</td>
<td>CS PPIS1</td>
<td>329</td>
<td>APPGE NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>322</td>
<td>PPIS1 VV0 J#</td>
<td>399</td>
<td>PPIS1 VV0</td>
<td>561</td>
<td>APPGE NN1</td>
<td>281</td>
<td>PPIS1 VV0</td>
<td>689</td>
<td>AT NN1</td>
<td>302</td>
<td>APPGE NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>326</td>
<td>APPGE NN1</td>
<td>384</td>
<td>PPIS1 VV0</td>
<td>565</td>
<td>CS PPIS1</td>
<td>257</td>
<td>APPGE NN1</td>
<td>861</td>
<td>PPIS1 VM</td>
<td>271</td>
<td>AT NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>316</td>
<td>PPPI1 VBZ</td>
<td>382.</td>
<td>J# PPIS1</td>
<td>503</td>
<td>CS PPIS1</td>
<td>248</td>
<td>AT NN1</td>
<td>860</td>
<td>VM XX VV</td>
<td>262</td>
<td>RO JJ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>299</td>
<td>NN1 NN1</td>
<td>365.</td>
<td>NN1 NN1</td>
<td>536</td>
<td>AT NN1</td>
<td>246</td>
<td>II APPGE NN1</td>
<td>654</td>
<td>APPGE NN1</td>
<td>251</td>
<td>PPIS1 VM XX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>288</td>
<td>NN1 CC NN1</td>
<td>362</td>
<td>CCB PPIS1</td>
<td>608</td>
<td>APPGE NN1</td>
<td>226</td>
<td>PPIS1 VD0 XX</td>
<td>446</td>
<td>APPGE NN1</td>
<td>261</td>
<td>VBDZ RG J</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>287</td>
<td>RR PPIS1</td>
<td>349</td>
<td>VD0 XX VV</td>
<td>499</td>
<td>PPIS1 VV0</td>
<td>274</td>
<td>CS PPIS1</td>
<td>635</td>
<td>NN1 NN1</td>
<td>245</td>
<td>NN1 CC NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>284</td>
<td>NN1 J#</td>
<td>321</td>
<td>APPGE NN1</td>
<td>477</td>
<td>PPIS1 VM</td>
<td>206</td>
<td>AT1 JJ NN1</td>
<td>610</td>
<td>PPIS1 VV0</td>
<td>245</td>
<td>VV0 TO VV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>273</td>
<td>VD0 XX VV</td>
<td>307</td>
<td>VBZ RG J</td>
<td>467</td>
<td>VM XX VV</td>
<td>198</td>
<td>VD0 XX VV</td>
<td>609</td>
<td>PPIS1 VD0 XX</td>
<td>238</td>
<td>PPIS1 VM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>255</td>
<td>NN1 CC</td>
<td>292</td>
<td>AT NN1</td>
<td>458</td>
<td>PPIS1 VM XX</td>
<td>193</td>
<td>RR PPIS1</td>
<td>606</td>
<td>VD0 TO VV</td>
<td>231</td>
<td>CS PPIS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>252</td>
<td>NN1 CCB</td>
<td>275</td>
<td>PPIS1 VV0</td>
<td>412</td>
<td>AT NN1</td>
<td>182</td>
<td>II APPGE NN1</td>
<td>579</td>
<td>PPIS1 VV0</td>
<td>230</td>
<td>PPIS1 VM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>236</td>
<td>J# PPIS1</td>
<td>275</td>
<td>NN1 CCB</td>
<td>422</td>
<td>II APPGE NN1</td>
<td>182</td>
<td>PPIS1 VV0</td>
<td>577</td>
<td>PPIS1 VV0</td>
<td>230</td>
<td>PPIS1 VM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>234</td>
<td>PPIS1 VM</td>
<td>275</td>
<td>PPIS1 VV0 TO</td>
<td>416</td>
<td>J# PPIS1</td>
<td>182</td>
<td>VM XX VV</td>
<td>568</td>
<td>PPIS1 VV0</td>
<td>221</td>
<td>AT1 JJ NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>232</td>
<td>PPIS1 VM</td>
<td>275</td>
<td>PPIS1 VV0 TO</td>
<td>416</td>
<td>J# PPIS1</td>
<td>182</td>
<td>VM XX VV</td>
<td>562</td>
<td>J# PPIS1</td>
<td>229</td>
<td>RR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>229</td>
<td>JJ NN1</td>
<td>267</td>
<td>PPPI1 VBZ</td>
<td>385</td>
<td>NN1 , CS</td>
<td>172</td>
<td>PPIS1 VM XX</td>
<td>541</td>
<td>PPIS1 VM XX</td>
<td>213</td>
<td>PPIS1 VM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>NN1 , CCB</td>
<td>246</td>
<td>NN1 CC NN1</td>
<td>380</td>
<td>NN1 , RR</td>
<td>171</td>
<td>CCB PPIS1</td>
<td>513</td>
<td>PPIS1 VM</td>
<td>205</td>
<td>NN1 , PPIS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>224</td>
<td>J#</td>
<td>244</td>
<td>J#</td>
<td>360</td>
<td>J#</td>
<td>181</td>
<td>PPIS1 VV0</td>
<td>602</td>
<td>NN1 CC NN1</td>
<td>201</td>
<td>II PPIS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>207</td>
<td>J# CCB</td>
<td>238</td>
<td>PPIS1 VV0</td>
<td>376</td>
<td>VBZ TO VV</td>
<td>189</td>
<td>PPIS1 CC NN1</td>
<td>482</td>
<td>VBZ RG J</td>
<td>199</td>
<td>AT1 NN1 IO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>203</td>
<td>II AT NN1</td>
<td>237</td>
<td>II AT NN1</td>
<td>371</td>
<td>PPIS1 VBDZ</td>
<td>188</td>
<td>PPIS1 VV0</td>
<td>489</td>
<td>AT1 NN1 IO</td>
<td>197</td>
<td>VD0 XX VV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>197</td>
<td>NN1 CC J#</td>
<td>235</td>
<td>NN1 , RR</td>
<td>368</td>
<td>PPIS1 RR</td>
<td>187</td>
<td>NN1 , CCB</td>
<td>442</td>
<td>RR PPIS1</td>
<td>195</td>
<td>NN1 , CS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>194</td>
<td>PPIS1 VBM</td>
<td>228</td>
<td>VM XX VV</td>
<td>364</td>
<td>VBZ RG J</td>
<td>185</td>
<td>AT NN1</td>
<td>441</td>
<td>NN1</td>
<td>189</td>
<td>PPIS1 VM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192</td>
<td>NN1 NN1 VBZ</td>
<td>225</td>
<td>WO TO VV</td>
<td>342</td>
<td>CS PPIS1 VV0</td>
<td>186</td>
<td>NN1 , CS</td>
<td>437</td>
<td>CS PPIS1 VV0</td>
<td>178</td>
<td>VBZ RG J</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>191</td>
<td>RO DA1</td>
<td>223</td>
<td>II APPGE NN1</td>
<td>341</td>
<td>JJ , PPIS1</td>
<td>184</td>
<td>J# PPIS1</td>
<td>424</td>
<td>PHB1 VY2</td>
<td>177</td>
<td>PPIS1 VM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>187</td>
<td>AT NN1</td>
<td>210</td>
<td>AT NN1</td>
<td>334</td>
<td>AT1 JJ NN1</td>
<td>181</td>
<td>PPIS1 VV0</td>
<td>407</td>
<td>RR</td>
<td>188</td>
<td>NN1 , NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>186</td>
<td>NN1 , APPGE</td>
<td>207</td>
<td>PPIS1 VM XX</td>
<td>331</td>
<td>NN1 CC NN1</td>
<td>161</td>
<td>VBZ RG J</td>
<td>416</td>
<td>VBDZ RG J</td>
<td>170</td>
<td>PPIS1 VV0</td>
<td>NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>PPIS1 RR</td>
<td>197</td>
<td>CS PPIS1 VV0</td>
<td>314</td>
<td>II AT NN1</td>
<td>153</td>
<td>PPIS1 VBZD</td>
<td>405</td>
<td>JJ NN1</td>
<td>164</td>
<td>NN1 , NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>182</td>
<td>AT NN1</td>
<td>190</td>
<td>NN1 , PPIS1</td>
<td>314</td>
<td>AT1 NN1</td>
<td>153</td>
<td>PPIS1 VBZD</td>
<td>390</td>
<td>NN1 IO NN1</td>
<td>162</td>
<td>PPIS1 VV0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>179</td>
<td>JJ , PPIS1</td>
<td>189</td>
<td>JJ , CCB</td>
<td>312</td>
<td>PPPI1 VBZ</td>
<td>150</td>
<td>NN1 , RR</td>
<td>360</td>
<td>AT NN1</td>
<td>182</td>
<td>AT NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>177</td>
<td>J# PPIS1</td>
<td>189</td>
<td>JJ , CCB</td>
<td>312</td>
<td>PPPI1 VBZ</td>
<td>150</td>
<td>NN1 , RR</td>
<td>360</td>
<td>AT NN1</td>
<td>182</td>
<td>AT NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>168</td>
<td>CCB J#</td>
<td>166</td>
<td>CS PPIS1 VV0</td>
<td>289</td>
<td>II NN1</td>
<td>148</td>
<td>II AT NN1</td>
<td>351</td>
<td>PHIS1 VBZD</td>
<td>159</td>
<td>AT NN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>164</td>
<td>VN1 VBZ J#</td>
<td>159.</td>
<td>NN1 , CS</td>
<td>235</td>
<td>AT NN1</td>
<td>119</td>
<td>VD0 XX VV</td>
<td>314</td>
<td>VM XX VV</td>
<td>133</td>
<td>JJ , PPIS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>141</td>
<td>VN1 VBZ</td>
<td>159</td>
<td>NN1 , CS</td>
<td>235</td>
<td>AT NN1</td>
<td>119</td>
<td>VD0 XX VV</td>
<td>314</td>
<td>VM XX VV</td>
<td>133</td>
<td>JJ , PPIS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MULTIVARIATE ANALYSIS
(CLUSTERING)
Analysis of Spoken Corpus (NICT JLE)

SST level | FREQ List
---|---
1 | 
2 | 
3 | 
4 | 
5 | 
6 | 
7 | 
8 | 
9 | 

Correspondence Analysis: Row Coordinates

Data summarization
• The most frequent 100 words can serve as a useful criterion for distinguishing one level from another.
The use of modal auxiliaries across different proficiency levels.
The use of modal auxiliaries across different proficiency levels.
The use of modal auxiliaries across different proficiency levels.
Kaneko (2006): NP structures

**NP types:**
- N
- num/possessive + N
- det + N
- N (adv)part + N
- (adv) adj + N
- N + (adv) + PP
- N + clause

**Learner Levels:**
- IL = Intermediate (low)
- IM = Intermediate (mid)
- IH = Intermediate (high)
- A = advanced
- NS = native speaker
Error freq’s & distributions
行ポイントと列ポイント
対称的正規化
次元 1
1.0 .5 0.0 -.5 -1.0 -1.5
次
元 2
1.0
.5
0.0
-.5
-1.0
-1.5
VAR00001
VAR00002
sst8/9
sst7
sst6
sst5
sst4
sst2/3
v_lxc
v_fin
v_mo
v_ng
v_vo
v_inf
v_cmp
v_tns
v_fm
v_agr
n_lxc
n_dprp
n_gen
n_inf
n_num
n_cnt
n_agr

Dimension 1（64.2%） Dimension 2（寄与率 20.6%）
Abe and Tono (2005)

Dimension 1 (65.5% of Inertia)

Dimension 2 (20.0% of Inertia)

WR mode SP

WR error rate pattern SP

Low Verb

High Noun
Error types across POS (Abe & Tono 2005)
AUTOMATIC ERROR IDENTIFICATION
Automatic identification of learner errors

- JEFLL Corpus ➔ The error-corrected version is now ready.

- We are working on the program that can compare the original and corrected versions of the sentence and automatically identify the patterns of deviation from the corrected sentence in terms of the following 3 types of errors (James 1998):
  - Addition/ omission/ misformation
Upper pain: original vs. lower pain: corrected
Errors involved in copula “be”

Friday, October 23rd and Saturday, October 24th were my school festival days. I am in the on-stage performance club. (?)

Our class had a mosaic exhibition.

We sang “kimi wo nosete” in the song contest.

My class took part in the song and art contests. (?)

I enjoyed the festival!

1C sang the song “Kiminosete”.

My class made a 3-D artwork. (?)

We were very busy doing rehearsals for it. (?)

We exhibited a big artwork in our classroom. (?)

We sang a song called “kimi wo nosete”.

The themes were flight, revolution and nature.

Our school festival had collage and song contests.

The performance days were October the twenty third, and twenty fourth.

Our school festival days were October 23rd and 24th.

My class sang “Kimi wo nosete”.

But neither 1-A, 1-B, 1-C nor 1-D won any prizes. (?)

Many club’s gave a performance.
DP matching

INPUT (corrected):

\[ W-a \quad W-b \quad W-c \quad \ldots \quad W-i \]

INPUT (original):

\[ W-a \quad W-b' \quad W-d \quad \ldots \quad W-i \]

ANALYSIS:

\[ W-a \quad W-b \quad W-c \quad W-d \quad \ldots \quad W-i \]

[msf] [oms] [add]
Automatic identification of learner errors

The first reason is every member of my family is busy in the morning.

Looking at n-grams for maximum match and analyse the unmatched elements:

- The member of my family is busy in the morning.
Automatic identification: output

T: My mother cooks very well ← corrected sentence
O: mother is cook very well ← original sentence
A: <oms>My</oms> mother <add>is</add> cook[*]:msf very well ← identifying differences

☐ Correspondence ratio:
  ■ Word level: 3/5
  ■ Character level: 3.80/5 (76%)

Notes: T = target; O = original; A = analysis
Looking for criterial features
Distributions of error types

- S3
- S2
- S1
- J3
- J2
- J1

Legend:
- omission
- addition
Distributions of error types

- S3
- S2
- S1
- J3
- J2
- J1

Legend:
- omission
- addition
Omission errors are significantly more frequent than addition errors. Errors will not decrease across proficiency.
Preposition errors (to/of)

More nominalization at advanced level, which increases the number of “of-addition/omission” errors.
Use of modals

Marked omissions at the very beginning stages
Later, more use of modals lead to more addition errors
Errors related to ‘have’

N-word clusters of “have”
Errors related to ‘have’

☐ The n. of article additions (218) is almost the same as that of omissions (213):
  ▪ “have a …” forms an unanalyzed chunk
  ▪ “have *a breakfast”/ “have *a time to …”

☐ Also the negation errors are very frequent:
  T: So I don't have time to eat breakfast
  O: So I have n't time to eat breakfast
  A: So I <oms>don't</oms> have <add>n't</add> time to eat breakfast
Supervised vs. unsupervised learning

Automatic extraction of error patterns from LC

Multivariate Analysis

Supervised Learning
- Classification
  - Advanced
  - Intermediate
  - Novice

Unsupervised Learning
- Clustering
New project: ICCI

- International Corpus of Crosslinguistic Interlanguage
- TUFS Global-COE Projects (5-year government-funded project)
- Aims: compiling corpora of young learners of English, comparable to JEFLL
- 7 countries (China; Taiwan; Israel; Spain; Poland; Austria; Singapore) at the moment
- Looking for more partner countries

Tokyo University of Foreign Studies
ICCI: Comparable English learner corpora

JEFL1
- beginning – intermediate levels
- JH1 (year 7) – SH3 (year 12)
- 10,000 subjects; 670,000 words

JEFL2

Spain, Austria, Israel, Poland, Taiwan, Hong Kong, Singapore

JEFL3

Korea, China, Russia, France, etc.