A phonological learner corpus of L2 French in Japan: from compilation to pedagogical use

Sylvain Detey¹ & Yuji Kawaguchi²
¹SILS, Waseda University, ²Tokyo University of Foreign Studies

3rd International Workshop on Advanced Learning Sciences
IWALS 2015 – TUFS – 2nd August 2015
Outline

I. Introduction: the IPFC corpus - origins and objectives

II. Methodology: protocol, coding procedure and Dolmen

III. Corpus and education: teacher’s training and in-class use

IV. Conclusion
I. Introduction: origins

Many studies in L2 pronunciation, but (Gut 2009, 2014; Zampini 2008):

- Mainly "laboratory speech"
  - limited number of speakers, structures and tasks under scrutiny
- Few studies with languages other than English (L2 or L1)
- Lack of disciplinary bridges (Escudero & Boersma 2004; Wauquier 2005)
  - phonetics vs phonology vs psycholinguistics vs education…

- At the same time, development of oral corpora and corpus-based phonetics-phonology: in L1, and more recently in L2
I. Introduction: origins

To study L2 phonology ⇒ existing corpora with:

• L2 Dutch (Neri et al. 2006)
• L2 Polish (Cylwik et al. 2009)
• L2 German and English (Gut 2009)
• L2 English in Asia (Visceglia et al. 2009)
• etc.

For L2 French?


But for L1 French: PFC (Phonologie du Français Contemporain)
(Durand, Laks, Lyche 2014)
I. Introduction: origins

Phonologie du Français Contemporain: usages, variétés et structure
M.-H. Côté (Laval), J. Durand (Toulouse), B. Laks (Paris) & C. Lyche (Oslo)
(large sociophonological survey on spoken French in the world with a common methodology)

In PFC: plurilingual speakers (Africa, Canada, Louisiana…) → native?
I. Introduction: origins

InterPhonology of Contemporary French (IPFC)
(Detey & Kawaguchi 2008; Racine et al. 2012; Detey & Racine 2012, Detey et al. to appear)

non-native part of the PFC programme

IPFC: http://cblle.tufs.ac.jp/ipfc
I. Introduction: objectives

General objectives of the IPFC programme:

- Describing non-native French pronunciation worldwide with a common methodology inspired by corpus phonology, comparable with ‘native’ pronunciation (i.e. PFC corpus)

- Using the results for:
  - Research purposes (e.g. testing L2 phonological models)
  - Applied purposes (e.g. pronunciation training syllabus design)
II. Methodology

- Recording protocol adapted from PFC: similar protocol for all L1s ⇒ data comparability
  - Between native (PFC) and non-native speakers (IPFC)
  - Between different groups of learners (e.g. Japanese and Spanish)

- Protocol: 6 tasks (≈ 1h of recording / learner)
  1. Repetition of a wordlist specific to each L1 including
     - 34 common words for all L1 (e.g. nasal vowels, rounded vowels, etc.)
     - 25-35 words specific to each L1
       (some of them shared among linguistically-related populations)
  2. Reading of the PFC wordlist (94 words)
  3. Reading of the L1-specific wordlist
  4. Reading of the PFC text
  5. Interview with a native speaker (two types according to students’ proficiency: A1-B1 & B2-C2)
  6. Semi-constrained interaction between two learners
     + a sociolinguistic questionnaire (22 questions)
II. Methodology

Data processing:

**Step 1: Orthographic transcription** of the data in Praat (Boersma & Weenink 2014) with sound-to-text alignment and transcription conventions adapted to L2 speech (Racine et al. 2011)
II. Methodology


- Coding a linguistic structure in order to:
  1. Automatically process large amount of data
  2. Include descriptive elements (target segment, left/right context, etc.)
  3. Use perceptive evaluation (targetlike vs. non targetlike quality of the vowel, nasality of the vowel, presence/absence of liaison/schwa, etc.)

- Coding procedure \(\Rightarrow\) intermediate stage between fine-grained acoustic analyses and coarse-grained phonological categorization in a perceptual approach (e.g. substitution / deletion / insertion)

(for details see Detey 2012, 2014)
II. Methodology

• IPFC coding procedure for:
  • Nasal vowels (Detey, Racine & Kawaguchi 2014)
  • Oral vowels (Detey & Racine 2013)
  • Liaison (Racine & Detey 2012; Detey et al. to appear; Racine 2014)
  • Consonants (Detey & Racine 2014)
  • Consonantal clusters (Detey et al. 2014; Detey & Racine 2014)
  • Schwa (Racine, Detey & Andreassen, in preparation)

• Important:
  • Since the assessment rests on a perceptual analysis, we need to keep a permanent link between sound, transcription and coding:
    ➤ coding procedure in Praat, on separate tiers
  • We need a consistent coding procedures for all the phenomenon
II. Methodology
II. Methodology

DOLMEN:

- A code… needs a decoding tool: *Dolmen*, developed by Julien Eychenne

(Eychenne & Paternostro to appear)

⇒ For automatic data processing and comparability on the basis of descriptive statistics
II. Methodology

Research domains:

• Nasal vowels /ɔ̃ ã̃ ɛ̃/ (Detey et al. 2010; Racine et al. 2010; Detey et al. 2014)
• Rounded vowels (Racine 2012; Racine et al. 2012)
• Voiced plosives /b d g/ (Racine et al. 2011)
• /CC/ clusters and liquid consonants (Detey et al. 2014)
• Stress and hierarchical prosodic organisation (Barquero 2012, 2013; Schwab 2012, 2013)
• Liaison (Racine & Zay 2012; Racine & Detey 2012; Falbo et al. 2013; Detey et al. to appear; Racine 2014; Racine & Detey in preparation)
• Schwa (Racine, Detey & Andreassen in preparation)
II. Methodology

- 15 different L1 are now represented in the IPFC project:
  1. German (Osnabrück, Munich, Wien, Zürich)
  2. English (Western Ontario)
  3. Danish (Copenhagen)
  4. Spanish (Geneva & Madrid)
  5. Greek (Poitiers & Cyprus/Thessaloniki)
  6. Italian (Milan, Trieste, Rome, Venice & Ticino)
  7. Japanese (Tokyo)
  8. Dutch (Nijmegen & Groningen)
  9. Norwegian (Tromsø & Oslo)
 10. Portuguese (Brasil, Santa Catarina)
 11. Russian (Moscow)
 12. Swedish (Dalarna)
 13. Turkish (Marmara)
 14. Arabic (France-Morocco-Lebanon-Palestine)
 15. Korean (Seoul)

The two more advanced corpora (in terms of methodology and studies carried out)
II. Methodology

Samples of the Japanese corpus are available here: http://cblle.tufs.ac.jp/ipfc/ipfcsearch/
III. Corpus and Education

- Over the past 15 years, great developments in the use of corpora for language education (written *and* oral):
  - Data (in France: ‘portail de la parole’ and beyond)
  - Tools (concordancers)
  - Procedures (data-driven learning, lexical chunk approach)
  - Diversity (oral, written, multimodal, authentic-pedagogical…)

- But usually: lexical, grammatical, pragmatic purposes
  (e.g. O’Keeffe, McCarthy and Carter 2007)

- And still: mostly *native* corpora

- What can we do with a non-native phonological corpus?
III. Corpus and Education

One example: Gut 2005 ‘Corpus-based pronunciation training’

The LeaP corpus: [http://www.phonetik.uni-freiburg.de/leap/](http://www.phonetik.uni-freiburg.de/leap/)

(Learning Prosody in a Foreign Language): acquisition of prosody by non-native speakers of German and English

4 speech styles: 1) readings of nonsense word lists, 2) reading passage, 3) retellings of a story, 4) free speech in an interview situation

Pedagogical use:

→ in a German pronunciation training course (perception exercises, theoretical input and practical exercises based on LeaP): some improvement (e.g. perception of stress)

→ In a data-driven approach as a tool for inductive learning (course entitled “Phonetic properties of non-native speech” with learners of English): learned about foreign accent and became more aware of foreign accents
III. Corpus and Education

Similarly for IPFC:

- Teachers’ training and support

- In-class use
III. Corpus and Education

For L2 (and L1) French teachers’ training (and practice)

(to appear in 2016)

With audio data from the IPFC corpus
Part 1: pronunciation of ‘native’ French

A) Pronunciation of native French
1. *Eléments de linguistique générale* (Durand & Eychenne)
2. *Le français de référence et la norme* (Detey & Lyche)
3. *Les accents et la variation* (Boula de Mareüil)

B) Geographic variation
4. *Le français méridional* (Durand & Eychenne)
5. *Le français en Belgique* (Bardiaux, Racine, Simon & Hambye)
6. *Le français en Suisse* (Racine)
7. *Le français au Canada* (Côté)
8. *Le français en Louisiane* (Dajko, Klingler & Lyche)
11. *Le français dans les DROM* (Pustka & Ledegen)

12. *Les autres types de variation* (Laks)

13. *La prononciation du français natif : pour aller plus loin* (Côté)
Part 2: pronunciation of L2 French learners

14. L’apprentissage de la prononciation du FLE (Detey & Racine)

15. Les anglophones (Tennant)
16. Les arabophones (Embarki, Abou Haidar, Zeroual & Naboulsi)
17. Les bosno-croato-serbophones (Pillot, Brjkan & Marijanovic)
18. Les coréanophones (Han & Eychenne)
19. Les danophones (Hansen & Johnsson)
20. Les germanophones (Pustka & Meisenburg)
21. Les hellénophones (Valetopoulos & Lamprou)
22. Les hispanophones (Racine)
23. Les italophones (Murano & Paternostro)
24. Les japonophones (Kamiyama, Detey & Kawaguchi)
25. Les lusophones (Seara, Nunes, Farias da Silva & Gonçalves de Andrade)
26. Les malaisophones (Hassan, Riget & Sévery)
27. Les néerlandophones (Berns & Nouveau)
28. Les norvégophones (Andreassen, Bordal & Lyche)
29. Les russophones (Boubnova & Ratnikova)
30. Les sinophones (Landron, Gao, Chang & Tian)
31. Les suédoophones (Stridfeltdt)
32. Les turcophones (Akinci, Kawaguchi & Yilmaz)
33. Les vietnamophones (Lethi)
Part 3: from education to research

D) Pronunciation teaching and corrective phonetics
34. L’enseignement de la prononciation : petit historique (Galazzi)
35. La correction phonétique : le rôle de la formation (Gil Fernández)
36. Enseignement de la prononciation et correction phonétique : principes essentiels (Detey)

E) Pronunciation of L2 French learners: a synthesis
(Detey, Racine, Kawaguchi, Eychenne)
37. Le domaine segmental (vocalique et consonantique)
38. Le domaine suprasegmental (syllabique, accentuel, rythmique et intonatif)

F) Pronunciation of L2 French learners: to go beyond
39. La prononciation des apprenants de FLE et la phonétique expérimentale (Kamiyama & Vaissière)
40. La prononciation des apprenants de FLE et la phonologie développementale (Shoemaker & Wauquier)
41. La prononciation des apprenants de FLE et la prosodie (Baqué)
42. La prononciation des apprenants de FLE et la multimodalité expressive (Shochi et Rilliard)
III. Corpus and Education

In-class use (perspectives):

- **Perceptual training:**
  - thanks to the coding system, access to native ratings and hence perceptual norms (subsets)
  - thanks to PFC : native vs non-native
  → not only the target, but also the divergent forms, to help learners perceive the differences
  → to be able to assess the well-formedness of the productions (native-like judgements)

- **Data-driven approach:**
  - Explicit approach of L1/L2 phonetics and phonology
  - Corpus-based metalinguistic awareness development (transcription, coding, etc.)

- **IPFC-Japanese:**
  - Focus on the specific features of Japanese learners (liquids, /CC/ clusters, nasal vowels, etc.)
  - But also use other data to expose students to ‘international French’ (continuum native-non-native)

- **Beyond phonetics-phonology** (more advanced level):
  - lexicon, morphology, grammar (spontaneous speech)

→ New project (Kaken (B) n° 15H03227 – S. Detey):
A corpus-based multi-level analysis of spoken French produced by pre-advanced Japanese learners of French
IV. Conclusion

Development of the IPFC corpus (including IPFC-Japanese):

- (i) a full-fledged searchable database (@Geneva University)
  → also for easier pedagogical use
- (ii) automatic functions in Dolmen to provide richer descriptions of the learners’ productions (+ variable assessment depending on the region)
  → link to perceptual norms and regional variation among natives
- (iii) applications for syllabus design and pronunciation training.
  → to be tested in class

Overall: integrate non-native material into the corpus-based resources already developed in PFC-EF and at TUFS: from pedagogical to authentic material, from native to non-native speech and vice versa, for beginners (at segmental level) and advanced learners (at discourse level)
We would like to thank:

- Isabelle Racine, Julien Eychenne, Mariko Kondo, Jacques Durand, Bernard Laks, Chantal Lyche and all of our IPFC colleagues

- The Japanese Society for the Promotion of Science – Grants-in-Aid for Scientific Research (B) N° 23320121 and 15H03227 (S. Detey).
Thank you for your attention
A phonological learner corpus of L2 French in Japan: from compilation to pedagogical use

Sylvain Detey¹ & Yuji Kawaguchi²
¹SILS, Waseda University, ²Tokyo University of Foreign Studies

3rd International Workshop on Advanced Learning Sciences
IWALS 2015 – TUFS – 2nd August 2015