



WASEDA University



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
(Granger & Goosens: <http://www.uclouvain.be/en-cecl-lcworld.html>).

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?

⇒ nothing comparable to ESF (*European Science Foundation Second Language*, Perdue, 1993), LANCOM (*LANgue et COMMunication*, Debrock & Flament-Boistrancourt, 1996) or FLLOC (*French Learner Language Oral Corpora*, Myles & Mitchell 2007) in the field of phonetics-phonology.

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)

www.projet-pfc.net

The screenshot shows the homepage of the PFC website. At the top, there's a navigation bar with tabs: ACCUEIL, PFC RECHERCHE (base de données en ligne), PFC ENSEIGNEMENT (le français expliqué), PFC PRÉSENTATION (le français oral dans le monde), IPFC, and PFC PUBLICATIONS (bulletins, colloques, logiciels). Below the navigation bar are four main content boxes:

- PARUTION:** « Les variétés du français parlé dans l'espace francophone : ressources pour l'enseignement » (avec DVD) (S. Detey, J. Durand, B. Laks & C. Lyche éds) dans la collection L'Essentiel français, aux Editions Ophrys [Lire la présentation](#)
- RECHERCHE:** Si vous êtes chercheur, universitaire ou étudiant en linguistique française. Utilisez la base de données schwa, liaison et prosodie pour vos recherches.
- ENSEIGNEMENT DU FRANÇAIS:** Vous êtes enseignant de français (FLE/FLS/FLM), concepteur pédagogique ou étudiant de français? Allez au module PFC Enseignement du Français.
- PRÉSENTATION DU FRANÇAIS:** Vous vous intéressez au français oral contemporain dans l'espace francophone en général ? Cette rubrique est pour vous...

At the bottom of the page, there's a section titled "Base de données sur le français oral contemporain dans l'espace francophone" with a brief description of the project's goals and its relevance to various stakeholders.

Sister project:
***Phonology of
Contemporary
English***
(Carr, Durand,
Przewozny)

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>

Interphonologie du Français Contemporain (IPFC)

IPFC

- Actualité
- Cadre IPFC
- Participants
- Descriptif
- Corpus
- Références
- Thèses et Mémoires
- Projets IPFC
 - IPFC-allemand
 - IPFC-anglais canadien
 - IPFC-espagnol
 - IPFC-grec chypriote
 - IPFC-italien
 - IPFC-japonais
 - IPFC-néerlandais
 - IPFC-norvégien
 - IPFC-portugais brésilien
 - IPFC-suédois
 - IPFC-turc
- Colloques
 - IPFC2011-Paris
 - IPFC2011-Tokyo
 - IPFC2010
- Sites partenaires

What's New

- 2012.07.20 Références
- 2012.07.11 Colloques
- 2012.05.25 IPFC-espagnol
- 2012.05.22 Projets IPFC
- 2012.05.20 IPFC-japonais

Bienvenue sur le site du projet IPFC (Interphonologie du français contemporain), piloté par:

Sylvain Detey (Université Waseda & Université de Rouen)

Isabelle Racine (Université de Genève)

Yuji Kawaguchi (Tokyo University of Foreign Studies)

Jacques Durand (Université de Toulouse & IUF)

Ce projet est dédié à l'étude des systèmes phonético-phonologiques des locuteurs non-natifs du français, pour lesquels le français est une langue étrangère (FLE) ou seconde (FLS). Il s'agit donc de populations d'apprenants qui peuvent faire usage du français dans diverses situations et appartiennent de ce fait au monde francophone.

Par interphonologie, on désigne généralement le nouveau système (phonético-)phonologique des apprenants d'une langue étrangère en cours de construction ou dans un état stabilisé.

Par-delà l'interphonologie, le projet IPFC concerne tous ceux qui s'intéressent à la production (et la perception) orale en français langue étrangère, puisque, à terme, le corpus IPFC devrait pouvoir être, au moins en partie, exploité pour des analyses multi-niveaux (morphologie, lexique, syntaxe, pragmatique).

- Pour une vision globale des objectifs et des enjeux du projet, consulter la section Cadre IPFC.
- Pour une vision plus précise de chacun des sous-projets de IPFC, consulter la section Projets IPFC.

Carte des enquêtes



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)

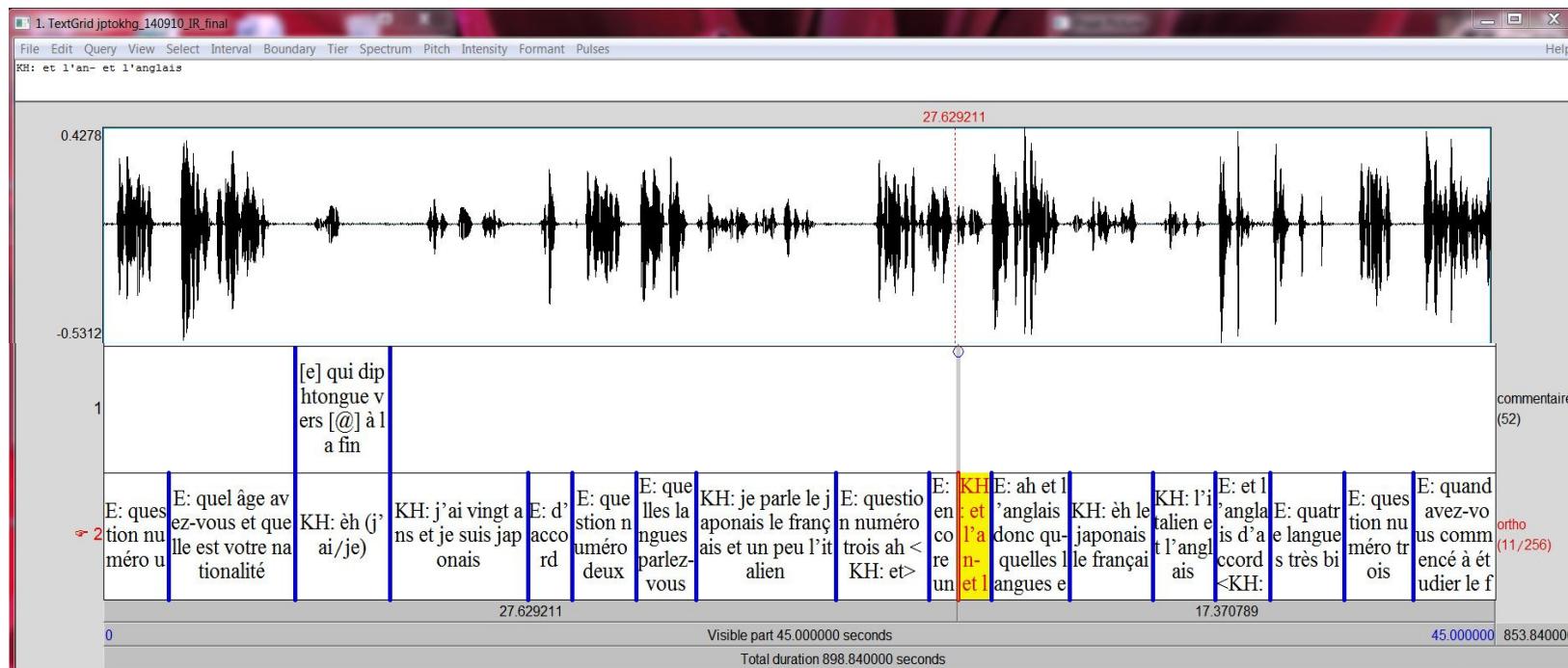
II. Methodology

- Recording protocol adapted from PFC:
similar protocol for all L1s \Rightarrow data comparability
 - Between native (PFC) and non-native speakers (IPFC)
 - Between different groups of learners (e.g. Japanese and Spanish)
- Protocol: 6 tasks (≥ 1 h 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

Step 2: coding (= PFC, Durand *et al.* 2009, 2014)

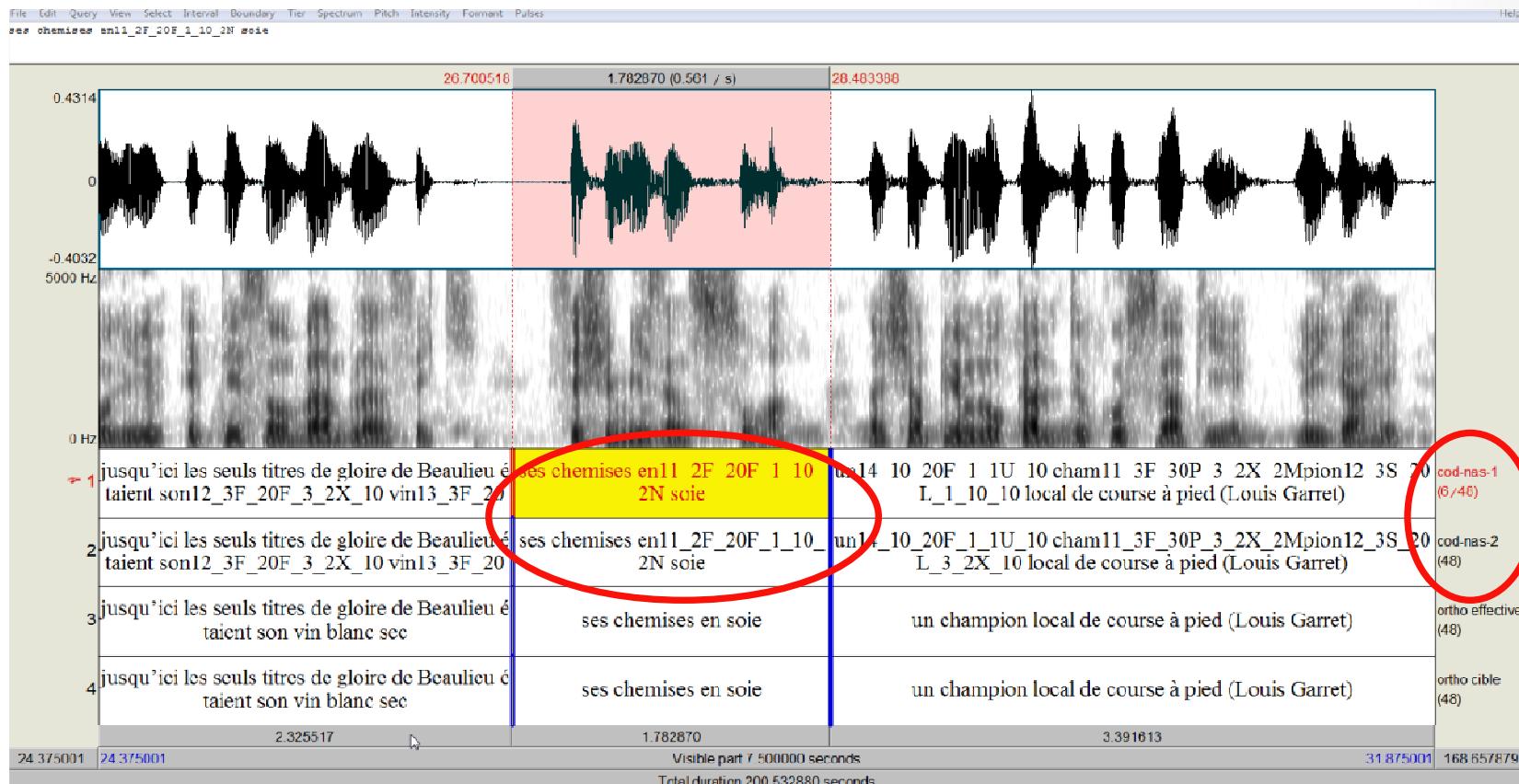
- Coding a linguistic structure in order to:
 1. Automatically process large amount of data
 2. Include descriptive elements (target segment, left/right context, etc.)
 3. Use perceptive evaluation (targetlike *vs.* non targetlike quality of the vowel, nasality of the vowel, presence/absence of liaison/schwa,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

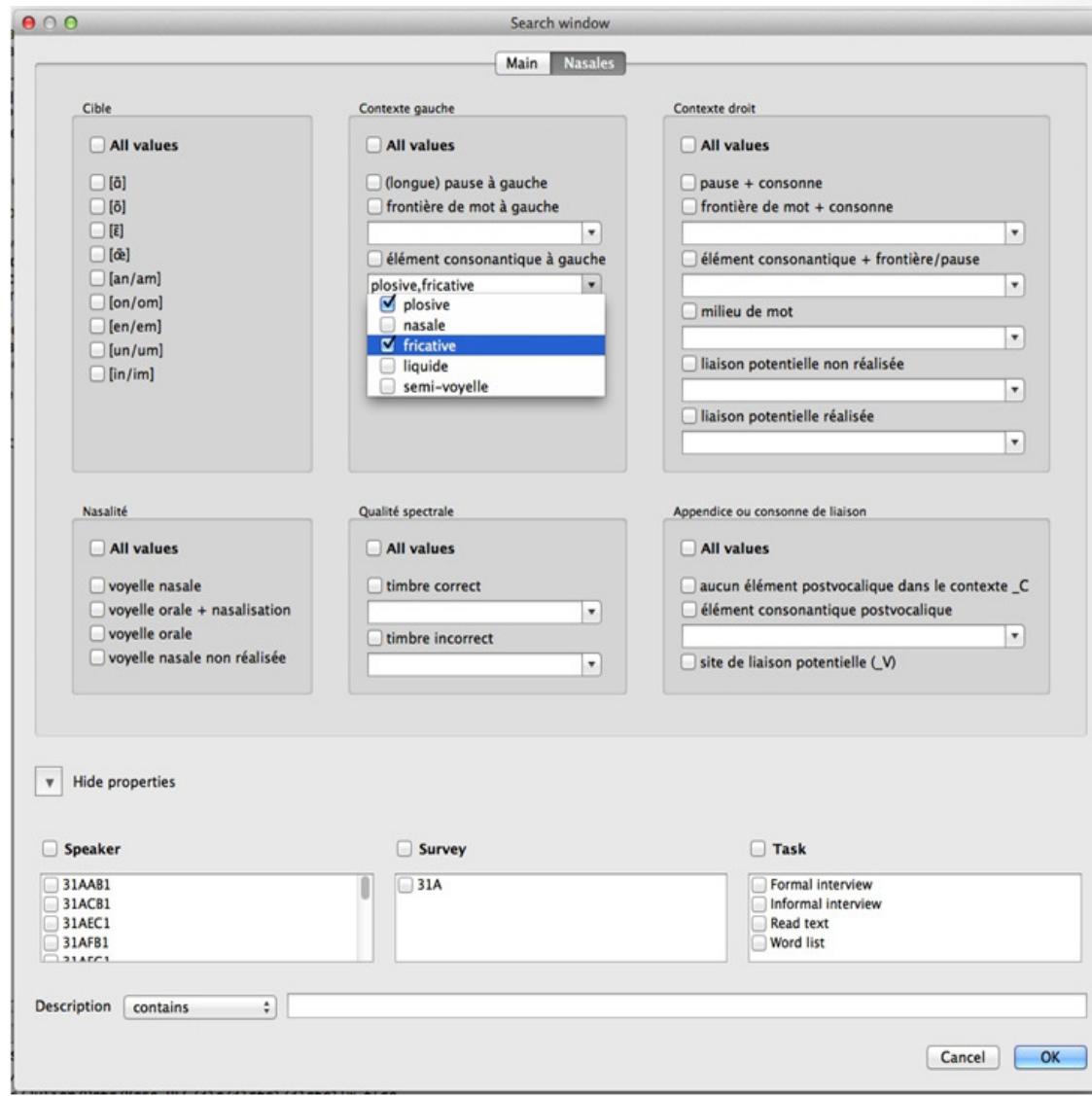
DOLMEN:

- 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 /œ ~ œ̃ ~ ε̃/ (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/>

The screenshot shows the IPFC website interface. At the top, there is a navigation bar with links: ACCUEIL, PFC RECHERCHE (base de données en ligne), PFC ENSEIGNEMENT (le français expliqué), PFC PRÉSENTATION (le français oral dans le monde), IPFC (highlighted in purple), and PFC PUBLICATIONS (bulletins, colloques, logiciels). Below the navigation bar is a search form with dropdown menus for 'Année de naissance', 'Études de français (en mois)', 'Séjours en pays francophone (en mois)', and 'Sexe', followed by a 'Recherche' button. A table below the search form displays user data and audio samples. The table has columns: No, Userid, Année de naissance, Lieu de naissance, Sexe, Nationalité, Langue maternelle, Spécialité, Études de français (en mois), Certificat, Séjours en pays francophone (en mois), and Séjours en pays étrangers non-francophone (en année). The first user entry is: 1, 275, 1986, Yamanashi, F, japonaise, japonais, linguistique, 60, 0, 12, 0. Below the table is a section titled 'LECTURE LISTE GÉNÉRIQUE' containing words like 'compagnie', 'médecin', 'baignoire', 'peur', 'moule', 'la vase', and 'all'. There is also a 'RÉPÉTITION LISTE SPÉCIFIQUE' section and a 'LECTURE LISTE SPÉCIFIQUE' section. At the bottom, there is a 'LECTURE TEXTE' section with the text: 'Le Premier Ministre ira-t-il à Beaulieu? « Le village de Beaulieu est en grand émoi. Le Premier Ministre a en effet décidé de faire étape dans cette commune au cours de sa tournée de la région en fin d'année.' Below this text are five more user entries in a table.

No	Userid	Année de naissance	Lieu de naissance	Sexe	Nationalité	Langue maternelle	Spécialité	Études de français (en mois)	Certificat	Séjours en pays francophone (en mois)	Séjours en pays étrangers non-francophone (en année)
1	275	1986	Yamanashi	F	japonaise	japonais	linguistique	60	0	12	0
2	271	1988	Osaka	F	japonaise	japonais	linguistique	96	APEF Jun1	1	6
3	267	1989	Okinawa	M	japonaise	japonais	linguistique	24	APEF 3	0	0
4	266	1943	Shizuoka	M	japonaise	japonais	linguistique	48	APEF Jun1	6	0
5	263	1986	Fukuoka	F	japonaise	japonais	linguistique	48	APEF 2	10	0

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/>
(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

LA PRONONCIATION
du FRANÇAIS
dANS LE MONDE

Du NATIF à l'APPRENANT

Sous la direction de :
S. DETEY, I. RACINE, Y. KAWAGUCHI, J. EYCHENNE

CLE
INTERNATIONAL

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)
9. *Le français au Maghreb et au Machrek* (Abou Haidar, Zeroual, Embarki & Naboulsi)
10. *Le français en Afrique Subsaharienne* (Boutin & Cissé)
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 suédophones (Stridfeldt)*
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**



WASEDA University



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