Corpus-based L2 phonological data and semi-automatic perceptual analysis: the case of nasal vowels produced by beginner Japanese learners of French

**BACKGROUND**

Using L2 oral corpora for L2 phonology studies (Gut 2009)
- Still fairly new & methodologically challenging, still most studies on L2 English
- Useful for research but also applied linguistics (language education and ASAP)

For L2 French, very few oral corpora – the first phonology-oriented: 
**InterPhonology of Contemporary French (IPFC)** (Detey & Kawaguchi 2008)

**InterPhonology of Contemporary French** (Detey & Racine 2012)
- Based on PFC project (native French corpus, Dudz. Laks. Lyche 2009)
- 4 native-nasal oral corpus, unique to L2 French (Racine et al 2012)

- A 6-task common protocol to all surveys:
  - Wordlist repetition & reading (3 tasks)
  - Text reading (1 task)
  - Conversations with native & non-native (2 tasks)
- A generic variationalist approach (no pre-categorization, Detey 2012)
- Orthographic transcriptions aligned (Pratt; Racine et al 2011)
- Manual coding system (vowels, consonants, liaison, etc.)
- Dedicated software (Dolmen) for code-based descriptive statistics

15 different L1-speaking groups
- Learning L2 French: Arabic, Danish, Dutch, English, German, Greek, Italian, Japanese, Korean, Norwegian, Portuguese, Russian, Spanish, Swedish, Turkish:
  - http://cette.tute.ucp.jp/pc/

- The present study: part of CLIJAF (Corpus Longitudinal Interphonologique d’Apprenants Japonais de Francais), a perception-production longitudinal study in the framework of IPFC with beginner Japanese students. Data used here correspond to the first stage of learning (out of 4 or 2 years).

**French nasal vowels**
- Acoustically complex, phonologically marked & morphological alternations
- Variation in the French-speaking world (learning input)
- Few available studies
- A good benchmark for the methodological approach we develop in IPFC

**Japanese learners**
- No nasal vowels
- Underspecified nasal mosaic /n/.
- Rich loanword lexicon and adaptation rules
- Few studies about L2 French nasal vowels acquisition

**OBJECTIVES**

- Present a procedure for corpus-based L2 phonological data in an applied perspective through a code-mediated perceptual analysis
- Illustrate with French nasal vowels produced by beginner Japanese learners
- Assess /N/ vs /ŋ/ in 2 tasks: wordlist repetition and reading

**METHOD**

**Participants**
- 22 Japanese students (m. age 19) learning French in Tokyo (4 months of study).
- 2 native French-speaking trained phonetic coders.

**Material**
- 12 monosyllabic words containing a nasal vowel (6 /ɔ̃/, 3 /ɛ̃/ 3 /ɑ̃/) in 6 pairs of words:  
  - anise-once, panise-ponce, pan-pon. Andres-Ande, lante-lante, tante-tant

**Perceptual analysis and coding procedure**

- **Reception tasks**
  - Wordlist repetition: listen to each word produced twice by a native and repeat.
  - Wordlist reading: read aloud the word displayed on computer screen.

- **Orthographic transcription**
  - Align with signal (Textgrid)

- **Coding**
  - For both tasks, 520 vowels coded
  - Double-blinded alphanumeric coding by 2 trained coders (auditory evaluation)
  - Code for nasal vowels: 6 fields (descriptive: a, b, c & 3 evaluative: d, e, f)
    - Target segment: a; Left & Right target segment contexts: b, c
    - Nasality: d (nasal vs. aspirated, nasal: aspirated, nasal: unaspirated)
    - Quality assessment (nasal vs. aspirated, nasal; unaspirated, nasal)
    - Consonantal excrescence assessment (appendix or not)

**RESULTS**

**Software used for code analysis** (Eychemen & Patrofostrochnic)
- Dolmen-IPFC, original open-source application for corpus linguistics, with dedicated IPFC plugins to analyse the IPFC coded data.
- Inter-coder reliability
  - ICC coefficient 0.369 (p < 0.001)
- Statistical analyses
  - For each characteristic (nasality, quality, excrescence): target-like assessment rate calculated as a function of vowel (/ɔ̃, /ɛ̃, /ɑ̃/) and task (repetition vs. reading).
- Mixed-effects regression models analyses conducted for each coder separately (participants and stimuli: as random terms)
- **Global analysis** (for each coder):
  - Vowel & nasal task (+ interaction vowel x task) for nasality & quality
  - Except /æ/ for nasality /æ/ vs. /ɑ̃/, /ɛ̃/ vs. /ɒ̃/ no vowel effect but task effect (+ interaction vowel x task) for excrescence

**Results 1: Nasality analysis** (nasal vs non-nasal productions, including subsequent)
- High rates of nasal for both coders (93.06% & 84.6%)
  - For each coder: bipartition
    - /ɔ̃, /ɛ̃, /ɑ̃/
    - Repetition vs Reading
  - Overall:
    - Nasality well acquired.
    - Better results in the repetition task.
    - Better productions of /ɑ̃/ and /ɛ̃/ over /ɔ̃/ in the reading task.

**Results 2: Quality analysis** (target-like vs non-target-like)
- High rates of target-like for both coders (76.30% & 67.12%) but lower than for nasality
  - For each coder, similar bipartition
    - Vowel effect: /ɔ̃ > /ɛ̃ > /ɑ̃/
    - Task effect: Repetition > Reading
  - Overall:
    - Repetition task: no vowel better than another quality-wise
    - Reading task: better productions of /ɔ̃/ and /ɛ̃/ over /ɑ̃/ (82.9% & 75.75% > 59.9% for Cod. 1 and 86.11% & 59.09% > 34.84% for Cod. 2)

**Results 3: Excrescence analysis** (without or with excrescence)
- High rates of target-like for both coders (76.30% & 65.58%)
  - For each coder:
    - No vowel effect
    - Task effect: Repetition > Reading
  - Task x vowel interaction:
    - Cod. 1: no difference between vowels for both tasks
      - Coherence with Cod. 2 with lower rates for /æ/ in reading.

**SUMMARY**

- Higher rates of target-like accent for nasality than for quality.
- Little difference between /ɔ̃, /ɛ̃, /ɑ̃/ but lower rates for /æ/ in reading.
- The distribution between the 3 rates is not significant in repetition (except in one case).
- Overall: better productions for the 3 rates in repetition rather than in reading.

**DISCUSSION**

- These results (beginner level) contrast with previous studies (advanced level – Racine, Deley, Buehler, Schweb, Zay, Kawaguchi 2010):
  - BOTH beginner & advanced: more excrescences in reading task
  - BUT different vowel quality ranking:
    - Advanced: /ɔ̃ > /ɛ̃ > /ɑ̃/
    - Beginner: /ɔ̃ > /ɛ̃ > /ɑ̃/ in reading but no distinction in repetition
  - AND different task effect: qualities:
    - Advanced: reading > repetition
    - Beginner: repetition > reading

- Hyp 1: novice acquire inter-category contrast oral/nasal before intra-category nasal quality distinction /ɑ̃/ vs. /ɒ̃/ vs. /ɛ̃/.
- Hyp 2: task impact changes with development (different production strategies).

**FUTURE RESEARCH**

- This study is part of a 2-year longitudinal perception-production study
- Final stage’s results will correspond (or not) to previous results (advanced level).
- Analyze the parallel development of the perception grammar.

**ACKNOWLEDGEMENT**

This study was supported by JSPS KAKENHI Grant B No. 23320121 (S. Deley).
We would like to thank Maniko Kondo, Jacques Durand, Marie-Laure Sandoz, Marion Diseut, M. Masuzawa, Tsuyoshi Umeno, Kaho Oomura. All the students participated in the study, as well as three anonymous reviewers.