



# Coder les consonnes dans IPFC: segments et syllabes

Sylvain Detey<sup>1</sup> & Isabelle Racine<sup>2</sup>

<sup>1</sup>SILS, Université Waseda & <sup>2</sup>ELCF, Université de Genève

Rencontres FLORAL 2014 - «Corpus oraux et enseignement de la prononciation en FLE & Interphonologie et corpus oraux»

*Maison de Norvège, Cité Internationale, 8-9 décembre 2014*



FONDS NATIONAL SUISSE  
DE LA RECHERCHE SCIENTIFIQUE



UNIVERSITÉ  
DE GENÈVE

FACULTÉ DES LETTRES



# Plan

- ▶ L'approche par codage dans IPFC – Rappel
  - ▶ Codages et décodage
  - ▶ Deux nouveaux codages IPFC:
    1. Le codage des consonnes
    2. Le codage des groupes consonantiques
  - ▶ Perspectives et conclusion
- 

# L'approche par codage dans IPFC - Rappel

- ▶ Stade 1: transcription orthographique des données (☛ *modification dans les conventions de transcription pour les conversations!*)
- ▶ Stade 2: codage (cf. codages PFC)
- ▶ Coder un élément permet de combiner:
  1. **Éléments descriptifs** (ex. cible, contexte gauche/droit)
  2. **Évaluation perceptive** par un/des codeur(s) (ex. degré de conformité de la réalisation par rapport à la cible, degré de nasalité d'une voyelle, présence/absence de la liaison/schwa, etc.)
- ⇒ Traitement automatique et comparable des données et obtention de statistiques descriptives
- ▶ Approche perceptive à mi-chemin entre analyse phonologique grossière (ex. substitution / effacement / insertion) et analyse phonétique (acoustique) fine du signal (pour un développement plus détaillé, cf. Detey, 2012).

# L'approche par codage dans IPFC - Rappel

- ▶ Deux points à souligner:
  - ▶ Comme il s'agit d'une analyse perceptive, importance cruciale de garder un **lien permanent entre son et transcription-codage**
    - ↳ codage sous Praat (Boersma & Weenink, 2014), dans des tires dédiées (idem PFC)
  - ▶ Importance d'avoir une **cohérence globale** entre les différents codes IPFC
    - ▶ Pour les codeurs/codeuses, 1 seule «logique» de codage
    - ▶ Pour le développement de l'outil d'extraction, 1 seule «logique»
    - ▶ Recoupements entre certains champs appartenant à deux codes différents (p. ex. voyelles nasales et liaison)

# L'approche par codage dans IPFC - Rappel

File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help

éviter les manifact- manifestation13\_3S\_20P\_1\_10\_10s qui on12\_2V\_50V\_1\_10\_30t eu ten11\_3P\_40P\_1\_20\_2Ndan11\_3P\_30F\_1\_10\_10ce à se mu- multiplier

112.947775 11.535959 (0.087 / s) 124.483734

0.5415  
0  
-0.6456

(To see the analyses, zoom in to at most 10 seconds,  
or raise the "longest analysis" setting with "Show analyses" in the View menu.)

1	comment	éviter les manifact- manifestations qui ont eu2_20V_2P_100 tendance à se mu- mu	lors des visit	cod-y (59)
2	comment	éviter les manifact- manifestations qui ont eu tendance à se mu- multiplier	lors des visit	cod-u (59)
3	comment	éviter les manifact- manifestations qui ont31_AUX_PPA_100_00_0_1 eu tendance	lors des visit	cod-liaison (59)
4	comment	éviter les manifact- manifestation13_3S_20P_1_10_10s qui on12_2V_50V_1_10_	lors des visit	cod-nas (24/59)
5	comment	éviter les manifact- manifestations qui ont eu tendance à se mu- multiplier	lors des visit	ortho effective (59)
6	comment	éviter les manifestations qui ont eu tendance à se multiplier	lors des visit	ortho cible (59)

1.572767 11.535959 1.891274

111.375008 111.375008 Visible part 15.000000 seconds 126.375008 191.305536

Total duration 317.680544 seconds

# L'approche par codage dans IPFC - Rappel

- ▶ Objectif: développer et mettre à disposition un système de codage – et de décodage – pour tous les aspects intéressants en phonologie L2 ⇒ permet d'assurer la comparabilité des données à grande échelle (cf. travail sur la liaison).
- ▶ Dans cette optique, développement successif de différents codes alphanumériques dans IPFC pour:
  - ▶ Les voyelles nasales (cf. Detey *et al.*, 2014; Detey, Racine & Kawaguchi, 2014)
  - ▶ Les voyelles orales (cf. Detey & Racine, 2013)
  - ▶ La liaison (cf. Racine & Detey, 2014; Racine & Detey, to appear)
  - ▶ Les consonnes
  - ▶ Les groupes consonantiques (Detey *et al.*, 2014)
  - ▶ Le schwa (à venir...)

# Codages et décodage

Search window

Main Nasales Liaison IPFC Nasales

Cible

- All values
- [z] singulier
- [z] pluriel
- [n]
- [t] singulier
- [t] pluriel
- [R]
- [p]
- [g]
- pas de consonne cible

Nb. de syll. et V. du mot liaisonnant

- All values
- monosyllabe
- polysyllabe
- monosyllabe avec liaison possible en [n] précédée de V. orale
- monosyllabe avec liaison possible en [n] précédée de V. nasale
- monosyllabe avec liaison possible en [n] précédée de V. orale + nasalisation ou incertain
- polysyllabe avec liaison possible en [n] précédée de V. orale
- polysyllabe avec liaison possible en [n] précédée de V. nasale
- polysyllabe avec liaison possible en [n] précédée de V. orale + nasalisation ou incertain

Pause/hésitation/coup de glotte

- All values
- absence de pause/hésitation/coup de glotte
- présence de pause/hésitation/coup de glotte

Catégorie du mot liaisonnant

- All values
- Catégorie

déterminant

Toutes

Catégorie du mot suivant

- All values
- Catégorie

nom propre, nom

Toutes

Réalisation

- All values
- absence de liaison
- liaison enchaînée
- liaison non enchaînée
- enchaînement incertain

Type de consonne de liaison

- All values
- absence de liaison
- conforme à la cible
- presque conforme mais différence phonétique
- non conforme mais C. présente dans le mot
- non conforme avec C. épenthétique

Compare annotators Reference annotator: Select fields

Hide properties

Description contains

OK Cancel

**Dolmen-IPFC:**  
(Eycheenne & Paternostro,  
à paraître)

- Outil permettant d'effectuer des requêtes sur la base du codage
- Interaction avec Praat  
(Lecture/écriture de fichiers TextGrid  
Ouverture de fichiers dans Praat.

Voir présentation J. Eycheenne



---

# Deux nouveaux codages IPFC

---



# 1. Le codage des consonnes

▶ Les **7 champs** du codage des consonnes:

- ▶ Champ 1: Consonne cible
- ▶ Champ 2: Position dans le mot
- ▶ Champ 3: Position syllabique à l'intérieur du mot
- ▶ Champ 4: Contexte phonologique gauche
- ▶ Champ 5: Contexte phonologique droit

Éléments  
descriptifs

- ▶ Champ 6: Qualité globale de réalisation
- ▶ Champ 7: Caractéristique phonétique  
(diffère en fonction des consonnes)

Évaluation perceptive

# 1. Le codage des consonnes

---

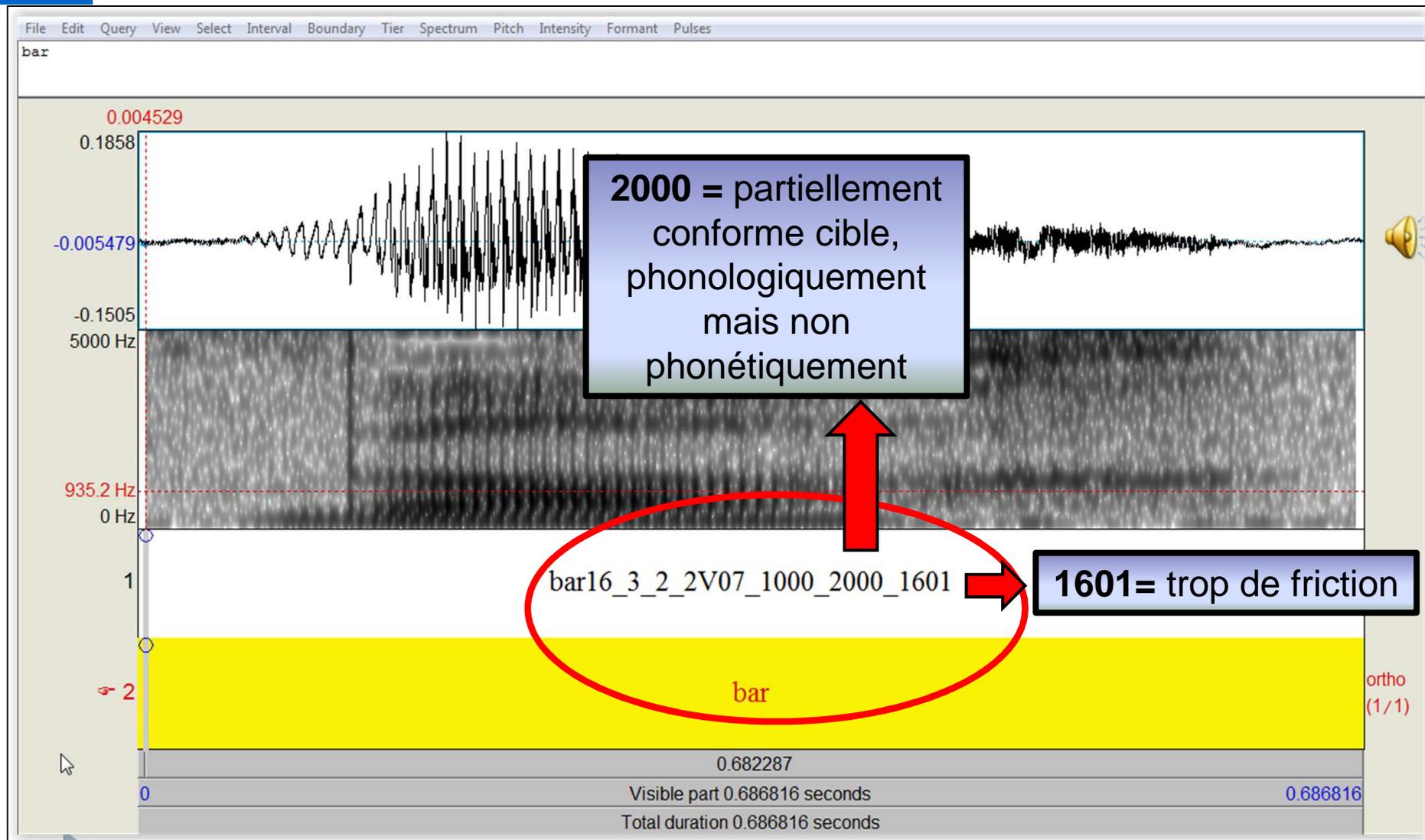
## ▶ Champ 6:

- ▶ **1000** = conforme cible (phonologiquement et phonétiquement)
- ▶ **2000** = partiellement conforme cible, phonologiquement mais non phonétiquement
- ▶ **30+00-16**= non conforme cible et réalisée comme :
  - ▶ 00= non identifiable ou consonne non française
  - ▶ 01-16= autre consonne française (= indiquer la consonne en ajoutant le code consonne)
- ▶ **4000**= effacement (ex. tigre prononcé [tiR], après prononcé [apE])
- ▶ **4100**= insertion consonantique (ex : rat prononcé [gRa])
- ▶ **4200** = insertion vocalique (ex. rat prononcé [aRa])

# 1. Le codage des consonnes

- ▶ **Champ 7: Caractéristique phonétique** (varie en fonction de la consonne cible)
  - ▶ **Pour // (= 09)**
    - ▶ **0901**= réalisation battue
    - ▶ **0902**= réalisation vélarisée
  - ▶ **Pour /R/ (=16):**
    - ▶ **1601**= trop de friction
    - ▶ **1602**= pas assez de friction
    - ▶ **1603**= réalisation apicale / battue / latérale
    - ▶ **1604**= réalisation apicale rétroflexe
  - ▶ **Pour /b/ (= 02):**
    - ▶ **02..** = ...
    - ▶ **02..** = ...

# 1. Le codage des consonnes



# 1. Le décodage des consonnes

The screenshot shows a 'Search window' application with the following sections:

- Search window** (Title bar)
- Navigation tabs:** Main, **Consonnes IPFC**, Liaison IPFC, Nasales IPFC
- Cible:**  All values; checkboxes for [p], [b], [m], [f], [v], [t], [d], [n], [l], [s], [z], [ʃ], [ʒ], [k], [g], [r].
- Position orth.:**  All values; checkboxes for Début de mot, Interne de mot, Fin de mot.
- Position syll.:**  All values; checkboxes for Attaque, Coda.
- Contexte gauche:**  All values; checkboxes for Pause, Voyelle (with dropdown), Consonne (with dropdown).
- Contexte droit:**  All values; checkboxes for Pause, Voyelle (with dropdown), Consonne (with dropdown).
- Qualité:**  All values; checkboxes for Conforme, Partiellement conforme (with dropdown), Non conforme (with dropdown), Effacement.
- Compare annotators:**  Compare annotators; Reference annotator: [text input]; Select fields [button].
- Show properties:**  Show properties.
- Description:** contains [dropdown] [text input].
- Buttons:** Cancel, OK.

## 2. Le codage des groupes consonantiques

- ▶ Les **10 champs** du codage des groupes consonantiques:

- ▶ Champ 1-4: Consonnes cibles du groupe consonantique
- ▶ Champ 5: Contexte phonologique gauche
- ▶ Champ 6: Contexte phonologique droit

Éléments  
descriptifs

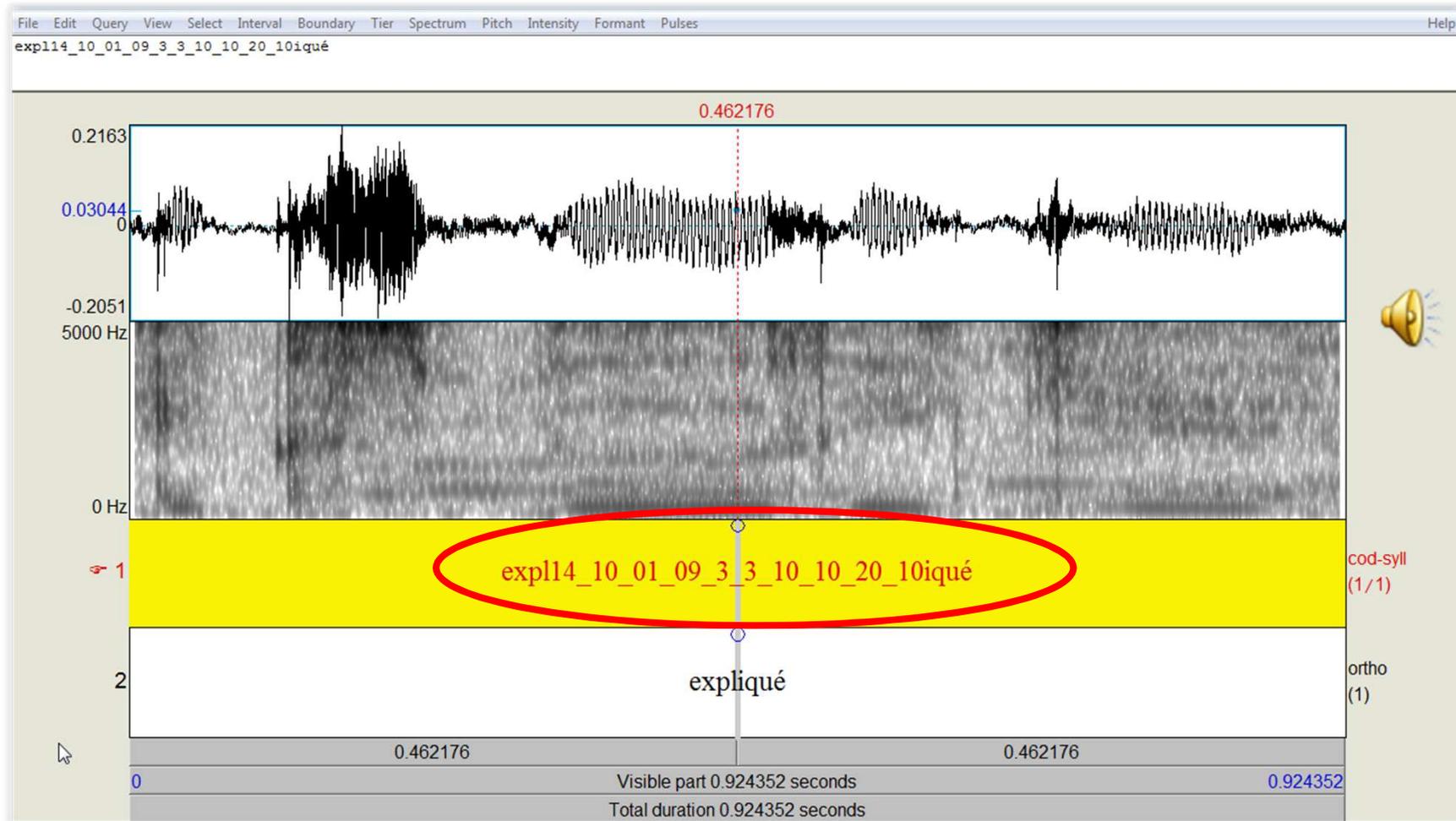
- ▶ Champ 7: Qualité globale de réalisation de la 1<sup>ère</sup> consonne
- ▶ Champ 8: Qualité globale de réalisation de la 2<sup>ème</sup> consonne
- ▶ Champ 9: Qualité globale de réalisation de la 3<sup>ème</sup> consonne
- ▶ Champ 10: Qualité globale de réalisation de la 4<sup>ème</sup> consonne

Évaluation  
perceptive

## 2. Le codage des groupes consonantiques

- ▶ **Evaluation perceptive (champs 7): la première consonne est:**
  - ▶ **10**= Conforme à la cible
  - ▶ **15**= Prothèse : insertion d'une voyelle avant C1 (ex : « spécial » /espesjal/)
  - ▶ **20**= Epenthèse (après C1) sans changement: insertion d'une voyelle après C (ex : « spécial » /səpesjal/)
  - ▶ **30**= Effacement : effacement de la consonne (ex : « spécial » /pesjal/)
  - ▶ **4X**= Métathèse : inversion des consonnes, le X indique si l'inversion est avec C2 (=2), C3 (=3), C4 (=4) (ex : « spécial » /psesjal/)
  - ▶ **50**= Changement : modification de la consonne (ex : « spécial » /fpesjal/)  
(changement de catégorie phonémique)
  - ▶ **51**= Changement avec épenthèse (ex : « spécial » /fəpesjal/)
  - ▶ **60**= XEffacement : il ne reste qu'une consonne, différente des deux (ex : « spécial » /fesjal/) (C1 et C2 sont codées 60)

## 2. Le codage des groupes consonantiques



## 2. Le décodage des groupes consonantiques

Search window

Main Consonnes IPFC Liaison IPFC Nasales IPFC Voyelles IPFC syllabe IPFC

Cible C1

- All values
- [p]
- [b]
- [m]
- [f]
- [v]
- [t]
- [d]
- [n]
- [l]
- [s]
- [z]
- [ʃ]
- [ʒ]
- [k]
- [g]
- [ʀ]

Cible C2

- All values
- null
- [p]
- [b]
- [m]
- [f]
- [v]
- [t]
- [d]
- [n]
- [l]
- [s]
- [z]
- [ʃ]
- [ʒ]
- [k]
- [g]
- [ʀ]

Cible C3

- All values
- null
- [p]
- [b]
- [m]
- [f]
- [v]
- [t]
- [d]
- [n]
- [l]
- [s]
- [z]
- [ʃ]
- [ʒ]
- [k]
- [g]
- [ʀ]

Cible C4

- All values
- null
- [p]
- [b]
- [m]
- [f]
- [v]
- [t]
- [d]
- [n]
- [l]
- [s]
- [z]
- [ʃ]
- [ʒ]
- [k]
- [g]
- [ʀ]

Contexte gauche

- All values
- ##
- #V
- V
- C#

Contexte droit

- All values
- ##
- #V
- V
- #C

C1

- All values
- conforme
- prothèse
- épenthèse
- effacement
- Métathèse
- 
- changement
- changement + épen.
- effac. X

C2

- All values
- null
- conforme
- épenthèse
- effacement
- Métathèse
- 
- changement
- changement + épen.
- effac. X

C3

- All values
- null
- conforme
- épenthèse
- effacement
- Métathèse
- 
- changement
- changement + épen.
- effac. X

C4

- All values
- null
- conforme
- épenthèse
- effacement
- Métathèse
- 
- changement
- changement + épen.
- effac. X

Compare annotators Reference annotator:

Show properties

Description

## Perspectives et conclusion

---

- ▶ Interface Dolmen pour ces deux nouveaux codages
- ▶ Affinage du champ 7 du code des consonnes pour les autres consonnes

**À venir!**

- ▶ Diffuser ces deux nouveaux codages
- ▶ Tester ce codage/décodage par divers apprenants
- ▶ Coder des données à grande échelle

**Permet des comparaisons inter-tâches  
et inter-populations**

---



---

# Merci de votre attention!

Ce projet a bénéficié et bénéficie du soutien:

- de la Société Japonaise pour la Promotion de la Science (JSPS)  
(Grant-in-Aid for Scientific Research B n°23320121)
- de l'Université Waseda (Special Research Grant, 2011B-297)
- du Fonds national suisse de la recherche scientifique (132144/1)
  - de la Faculté des Lettres de l'Université de Genève
  - de la Société Académique de Genève (Fonds Ch. Bally)

Un merci tout particulier à **Romain Isely** pour le travail de « débroussaillage » des données japonaises!





# Références

---

- ▶ Boersma, P. & Weenink, D. (2014). *Praat: doing phonetics by computer*. <http://www.praat.org>.
- ▶ Detey, S. (2012). Coding an L2 phonological corpus: from perceptual assessment to non-native speech models – an illustration with French nasal vowels, in Y. Tono, Y. Kawaguchi & M. Minegishi (eds.), *Developmental and Crosslinguistic Perspectives in Learner Corpus Research*. Amsterdam and Philadelphia: John Benjamins, 229-250.
- ▶ Detey, S., Kondo, M., Racine, I. & Kawaguchi, Y. (2014). A preliminary investigation of /CC/ clusters acquisition by Japanese learners of French using oral corpora: methodological insights, Workshop *Learner Corpus Studies in Asia and the World (LCSAW)*, Kobe University, May 31<sup>st</sup> 2014.
- ▶ Detey, S., Racine, I., Eychenne, J. & Kawaguchi, Y. (2014). Corpus-based L2 phonological data and semi-automatic perceptual analysis: the case of nasal vowels produced by beginner Japanese learners of French, in *Proceedings of Interspeech2014* (Singapore): 539-544.
- ▶ Detey, S., Racine, I. & Kawaguchi, Y. (2014). Des modèles prescriptifs à la variabilité des performances non-natives : les voyelles nasales des apprenants japonais et espagnols dans le projet IPFC, in J. Durand, G. Kristoffersen & B. Laks (eds). *La phonologie du français : des normes aux périphéries* (Festschrift pour Chantal Lyche). Paris : Presses Universitaires de Paris Ouest, 197-226.
- ▶ Detey, S. & Racine, I. (2013). L2 oral corpus data processing at the segmental level : methodological challenges. Workshop “*Cross Cultural Research on Speech Communication & Second Language learning processing*”, Université de Bordeaux III, 15 mars 2013.





# Références

---

- ▶ Eychenne, J. & Paternostro, R. (to appear). Analyzing transcribed speech with Dolmen, in S. Detey, J. Durand & C. Lyche (eds). *Varieties of Spoken French: a source book*. Oxford: Oxford University Press.
- ▶ Racine, I. & Detey, S. (en préparation). L'apprentissage de la liaison en français par des locuteurs non natifs : éclairage des corpus oraux. *Bulletin VALS-ASLA* (n° 102, hiver 2015).
- ▶ Racine, I. & Detey, S. (2014). La liaison dans un corpus d'apprenants : le projet « InterPhonologie du Français Contemporain » (IPFC), in Colloque « *Corpus des français parlés et français parlés des corpus* », University of Neuchâtel, 8-9 mai 2014.

