

Coder les voyelles nasales dans le corpus IPFC: résultats préliminaires

^{1,2}Sylvain Detey, ³Isabelle Racine, ³Françoise Zay,
⁴Yuji Kawaguchi, ¹Mariko Kondo & ⁵Laurie Buscaïl

¹ SILS, Waseda University & ² LiDiFra, Université de Rouen,
³ ELCF, Université de Genève,
⁴ Tokyo University of Foreign Studies
⁵ CLLE-ERSS, Université de Toulouse

Journées IPFC2011-Paris

« Interphonologie du français contemporain : des voyelles nasales à la prosodie »

6-7 décembre 2011



WASEDA University





Plan

I. Introduction:

coder un corpus phonologique en L2

II. Les voyelles nasales dans IPFC:

objectifs et enjeux

III. Un code pour les voyelles nasales:

illustrations et premiers résultats descriptifs

IV. Conclusion:

développements et perspectives



I. Introduction: coder un corpus phonologique en L2

- Parmi les enjeux de l'analyse sur corpus:
 - La représentation visuelle des données
 - Le traitement automatique des données
 - Au sujet de la représentation visuelle et de la discussion transcription phonétique / orthographique en L2 voir Racine, Zay, Detey & Kawaguchi (2011 – *Travaux du Cerlico* 24).
 - Notre base: corpus transcrit orthographiquement avec alignement texte-son.
- Traitement automatique des données ?



I. Introduction: coder un corpus phonologique en L2

L'exemple de PFC (Durand, Laks, Lyche 2009): utilisation d'un code (liaison – schwa)

- Ex. Liaison: *Le village de Beaulieu est11t en grand11t émoi*
- Codage : insertion de symboles alphanumériques dans la transcription orthographique pour automatiser – à des fins descriptives – la quantification et la qualification de phénomènes linguistiques de surface.
- Deux tâches:
 - Conception du code
 - Conception d'un outil de décodage



I. Introduction: coder un corpus phonologique en L2

- Coder un élément permet, pour son analyse, de:
 - 1) Intégrer son contexte d'occurrence
 - 2) Intégrer son évaluation perceptive (*rating*)
 - 3) Traiter automatiquement les données de manière comparable (surtout en tâche de lecture/répétition)
 - Obtention automatique de statistiques descriptives
- Coder: à mi-chemin entre analyse phonologique grossière (ex. Substitution / Effacement / Insertion) et analyse acoustique fine du signal



II. Les voyelles nasales dans IPFC: objectifs et enjeux

- En L2: trois types d'évaluation (Munro 2008):
 - évaluation perceptive non-experte
 - évaluation perceptive experte
 - analyse acoustique
- Dans une perspective appliquée: importance de la perception et de l'évaluation par la communauté des locuteurs
 - Ex: John Lennon vs Jean Réno dans une publicité japonaise en 2011...



II. Les voyelles nasales dans IPFC: objectifs et enjeux

- Etudes précédentes sur les nasales dans IPFC: trois volets (2x perception + acoustique) (Detey & al. 2010, Racine & al. 2010)
- MAIS études sur mot isolés (comparabilité).



II. Les voyelles nasales dans IPFC: objectifs et enjeux

- Pour parole continue (en particulier entretien): nécessité d'adopter une autre procédure → approche par codage.
- Quel(s) paramètre(s) coder?
 - Degré de nasalité de la voyelle
 - Qualité spectrale de la voyelle
 - Présence d'un appendice consonantique
 - Influence du contexte
 - Influence de la tâche
 - Influence du groupe (L1, Sexe, Age, ...)
 - ...

III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

- Le code initial

First field (syllabic context)	Second field (vocalic nasality)	Third field (spectral quality)	Fourth field (postvocalic consonantal excrecence)
1-V	1-nasal vowel	1-targetlike	1-no postvocalic consonant
2-VC	2- oral vowel with nasalisation	2-uncertain	2- postvocalic consonant or appendix
3-CVC		3-off-target	
4-CV	3-oral vowel		
5-CCVC			

III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

- Le code révisé

1- Target (2 elements)	2- Left context (2 elements)	3- Right context (3 elements)	4-Vocalic nasality (1 element)	5-Spectral quality (2 elements)	6-Postvocalic consonantal excrescence or nasal liaison (2 elements)	7-Word target (1 element)
<p>1x = v nasal 11 = [a~] 12 = [o~] 13 = [e~] 14 = [ɤ~]</p> <p>2x = v oral corresponding 21 = [an/am] 22 = [on/om] 23 = [en/em] 24 = [un/um] 25 = [in/im]</p>	<p>10 = ##_ Ex: (C'est##) en ...</p> <p>2 = #_ *Add P/N/F/L/S/V Très[z] inquiet Ou Très [z] inquiet</p> <p>3 = C_ *Add P/N/F/L/S Ex: blanc: 3L</p>	<p>100 = _## C Ex: ... pain. C'est...</p> <p>20 = _# C (with C NOT liaison consonant) *Add P/N/F/L/S Ex: pain dur</p> <p>30 = _C#(≠) (with C NOT liaison consonant) Ex: grande (porte)(,). *Add P/N/F/L/S</p> <p>40 = (C)(C)_C(C) Ex: amplifier, grandement. *Add P/N/F/L/S</p> <p>5- Potential liaison sites: *50: no liaison Ex : bon // été ; bon // enfant ; mon // enfant ; ton // haricot ; grand // enfant ; fins // haricots ; ...le coin. On est... ; comment // en plus *51 : liaison or consonantal appendix with or without enchaînement Ex : bon[n]enfant ; mon[n]enfant ; ton[n]haricot ; (including liaison « abusive », Ex: gran(d)[n]enfant ; fins[n]haricots) « un membre de l'opposition[n] / aurait déclaré » = 51V</p> <p>*Add V if oral N if nasal Ex: mon // enfant = 50N bon[n]été = 51V</p>	<p>1- nasal vowel Ex: pain = [pa~]</p> <p>2- oral vowel + nasalisation (or uncertain) Ex: pain = [pa~]</p> <p>3- oral vowel Ex: pain = [pa]</p> <p>4 – nasal vowel non realised Ex. téressé (intéressé)</p>	<p>1- targetlike for AN/ON/IN/UN=[e~] or [ɤ~] *Add 0 if AN/ON/IN *Add I if UN=[e~] *Add U if UN=[ɤ~]</p> <p>2- off-target +A/O/I/U/X ([a~], [o~], [e~], [ɤ~], X = other vowel)</p>	<p>10- no postvocalic consonant (in context_C)</p> <p>2- postvocalic consonant or appendix (in context_C) *Add N, M ou G according to the nature of the sound * Add X if other</p> <p>30-Potential liaison site (context_V)</p>	<p>1 – nasal vowel in word lists</p> <p>2– nasal vowel in the text or conversation</p>



III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

Champ 1: Voyelle cible

- 1x = V nasale
 - 11 = [a~]
 - 12 = [o~]
 - 13 = [e~]
 - 14 = [ɥ~]
- 2x = séquence Vocale+C (étude nasalisation abusive)
 - 21 = [an/am]
 - 22 = [on/om]
 - 23 = [en/em]
 - 24 = [un/um]
 - 25 = [in/im]

III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

Champ 1: Voyelle cible – Exemples

- ses chemises en **11_2F_20F_1_10_10_2** soie 🗣️
- son **12_3F_51V_3_2X_30_2** usine de pâtes italiennes 🗣️
- in **13_10_40P_1_10_10_2** diqueraient 🗣️
- d'un **14_3P_51V_2_2U_30_2** autre côté 🗣️






III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

Champ 4: Nasalité vocalique

- 1- voyelle nasale
 - Ex: pain = [pa~]
- 2- voyelle orale + nasalisation (ou incertaine)
 - Ex: pain = [paa~]
- 3- voyelle orale
 - Ex: pain = [pa]
- 4- voyelle nasale non réalisée
 - Ex. téressé (intéressé)

III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

Champ 4: Nasalité vocalique - Exemples

- mille neuf cen11_3S_20P_1_10_2N_2t trente-six 
- cam11_3P_40P_2_2A_2M_2pagne profonde 
- gran11_3L_51V_3_2X_30_2d émoi 





III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

Champ 5: Qualité spectrale

- 1- Conforme à la cible pour AN/ON/IN/UN=[e~] ou [ɔ~]
 - *Ajouter 0 si AN/ON/IN
 - *Ajouter I si UN=[e~]
 - *Ajouter U si UN=[ɔ~]
- 2- non-conforme +A/O/I/U/X
([a~], [o~], [e~], [ɔ~], X = autre voyelle)

III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

Champ 5: Qualité spectrale - Exemples

- j'ai vin13_3F_30P_1_10_10_2gt-cin13_3F_30P_1_10_10_2q ans 
- j'ai vingt-cinq an11_2P_50V_1_20_30_2s 






III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

Champ 6: **Excroissance consonantique ou liaison nasale**

- 10- pas de consonne postvocalique (en contexte_C)
- 2- consonne postvocalique ou appendice (en contexte _C)
 - * Ajouter N, M ou G selon la nature du segment
 - * Ajouter X si autre segment
- 30-site de liaison potentielle (contexte _V)

III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

Champ 6: Excroissance consonantique ou liaison nasale - Exemples

- ses chemises en11_2F_20F_1_10_10_2 soie 
- en fin13_3F_20P_3_2X_2N_2 d'année 
- est en gran11_3L_51V_3_2X_30_2d émoi 

III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

- Une fois codé... il faut décoder: → plateforme **Dolmen** (Julien Eychenne)

The screenshot shows the Dolmen software interface. The main window displays a text grid with phonetic annotations for a French text. The text is: "tité risquent de provoquer une explosion un jeune membre s'électionn comment en plus éviter les manifestation n23P4P32X2N2don23P2P1102M2s pas de la réaction des gen113F2P1102N2s du pays nous av la côte escarpée du mon23N2F1102N2t sain33F2P1102M2t-pierre qui mène au village com t des barrages chaque fois que les opposan113F2P1102N2ts de tous les bords manifestent leur mpagne profonde le maire de Beaulieu Marc Blan113L10110102c est en112V2L1102N2 revan113F3F110102che vu à Jon123F4P110102quièr et des vérification123S2P32X102s d'iden113P4P110102tité risquent de provoq comme on122N5N110302 en11105V32X312 a vu à jon123F4P110102quièr et des vérification123S2P32X2N tout est fait pour le protéger or un14102P11U2G2 gros détachemen113N2P1102N2t de police ros détachemen113N2P110102t de police comme on122N5N110312 en112V5V32X312 a vu à Jon123F4P110102quièr Ministre ne cesse de baisser depuis les élection123S10110102s comment113N10110102t en11102P110102 plus iter les manifestation123S10110102s qui on122V3P32X2N2t eu ten113P4P32X2N2dan113P3F110102ce à s 'opposition123S5V32X302 aurait déclaré dan113P2L32X102s le coin133S10110102 on12105V32X312 détachemen113N2P1102N2t de police comme on122N5N110302 en11105V32X312 a vu à jon123F4P110102quièr arc Blan113L10110102c est en112V2L1102N2 revan113F3F110102che très in132F4P110102quiet la cote du P 2P110102s pas de la réaction123S2P32X102 des gen113F2P110102s du pays nous avon123F2L32X102s le souti est jaloux de notre liberté s'il faut mon123N4P32X2N2trer patte blan113L3F32X2N2che pour circule aire de Beaulieu Marc Blan113L10110102c est en112V2L1102N2 revan113F3F110102che très in132F4P110102qu quièr et des vérification123S2P32X102s d'iden113P4P110102tité risquent de provoquer une explosion123 3F2L110102s le soutien133S101102N2 du village en112F4P32X2N0tier F110102che pour circuler nous ne répond- répon123P4P32X2N2don23P2P110102s pas de la réaction123S2P32 berté s'il faut mon123N4P110102trer patte blan113L3F110102che pour circuler nous ne répond- répon12 on depuis les élection123S10110102c comment113N10110102t en11102P32X2N2 plus éviter les manifestat

The interface includes a project list on the left, a search bar, and an output window at the bottom showing saved files.



III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

Résultats quantitatifs à partir du champ 1 :

- Texte japonophones :
 - 11 : 142 occurrences
 - 12 : 85
 - 13 : 48
 - 14 : 13

En tout, 288 nasales codées

- Guidée japonophones :
 - 11 : 222
 - 12 : 134
 - 13 : 59
 - 14 : 21

En tout, 436 nasales codées

- Texte hispanophones :
 - 11 : 187
 - 12 : 111
 - 13 : 63
 - 14 : 16
 - 23 : 1

En tout, 378 nasales codées

288 + 436 + 378 + 180 mots

= **1282** voyelles nasales codées



III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

Résultats des études précédentes (sur les mots seulement):

A) *Le timbre est meilleur chez les japonophones que chez les hispanophones* → **Qualité du timbre:**

→ recherche dans Dolmen = .._..._..._.._1_..._..

Résultat:

- Japonophones : 144 (texte) + 38 (LSL) + 6 (LSR) = 188
- Hispanophones : 78 (texte) + 27 (LSL) + 10 (LSR) = 115

188 (jap) > 115 (hisp) → concordant a priori !



III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

B) La voyelle ON est la plus fidèlement réalisée

→ recherche dans Dolmen = **12_..._..._._1_..._.**

Japonophones :

- réussite texte: 49% (AN) ; **58% (ON)** ; 50% (IN) ; 23% (UN)
- réussite LSL: 73 % (AN) ; **100 % (ON)** ; 80 % (IN) ; UN absent
- réussite LSR: 0 % (AN) ; **60 % (ON)** ; 50 % (IN) ; UN absent

Hispanophones :

- réussite texte: 26% (AN) ; **37% (ON)** ; 22% (IN) ; 0% (UN)
- réussite LSL: 73% (AN) ; **93% (ON)** ; 13% (IN) ; UN absent
- réussite LSR: **100% (AN)** ; **80% (ON)** ; 50% (IN) ; UN absent

→ concordant a priori !



III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

C) *Excroissances plus nombreuses en lecture qu'en répétition*

→ **recherche dans Dolmen** = .._..._..._..._2._.

Hispanophones LSL : 18 ; LSR : 13 => concordant a priori

Japonophones LSL : 12 ; LSR : 4 => concordant a priori

→ concordant a priori !



III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

ET

D) Timbre est meilleur en lecture qu'en répétition

→ recherche dans Dolmen = .._..._..._.._1_..._.

- Hispanophones LSL : 27 ; LSR : 10 => concordant
- Japonophones LSL : 38 ; LSR : 6 => concordant

→ *4/4 résultats concordants entre les deux démarches*

Codage (1 codeuse)

vs

évaluation non-exp. (30 éval.) & évaluation exp. (4 éval.)



IV. Conclusion: développements et perspectives

- Malgré la complexité du codage (Gut à paraître: “The higher the number of different categories in an annotation scheme, the lower the [inter-rater] agreement”), des résultats concordants avec les précédentes études.
- Mais la question du/des codeur(s) reste importante (si natif: variété en L1; non-natif?)
 - nécessité d'évaluateurs multiples (cf. transcription phonétique multiple à l'aveugle dans PHON)
 - Calcul de coefficient Kappa et développement de Dolmen



IV. Conclusion: développements et perspectives

Dans cette étude:

- Explorer un mode d'exploitation de corpus phonologique en L2 utile pour la recherche:
 - Fondamentale (analyse interphonologique)
 - Appliquée (transformation d'un corpus en corpus *évalué* ("rated database"))
- Obtenir des résultats préliminaires sur les voyelles nasales
 - Concordants avec les résultats de nos études expérimentales
 - Encouragent le développement du codage
- Préparer un codage plus extensif (voyelles; consonnes; structures syllabiques) avec outil adéquat (Dolmen-IPFC)



Merci!

Ce projet a bénéficié du soutien de:

- Deux subventions de recherche au Japon (S. Detey):
 - Special Research Grant (2011B-297) de l'Université Waseda.
 - Research Grant 'B' (2011B-297) du Ministère japonais de l'éducation, de la science et de l'industrie
- Trois subventions de recherche à Genève (I. Racine):
 - Projet FNS (100012_132144/1)
 - Subvention de recherche de la Faculté des Lettres de l'Université de Genève
 - Subvention de recherche de la Société Académique de Genève (Fonds Ch. Bally)

Remerciements particuliers à Laurie Buscail, Julien Eychenne et Sandra Schwab...



Références

- Detey, S., Racine, I., Kawaguchi, Y., Zay, F., Buehler, N. & Schwab, S. (2010). Evaluation des voyelles nasales en français L2 en production: de la nécessité d'un corpus multitâches. In Neveu, F., Durand, J., Klingler, T., Prévost, S. & Muni-Toké, V. (eds), *Actes de CMLF'10*. Paris: ILF. 1289-1301.
- Durand, J., Laks, B. & Lyche, C. (2009). Le projet PFC: une source de données primaires structurées. In J. Durand, B. Laks et C. Lyche (eds). *Phonologie, variation et accents du français*. Paris: Hermès. 19-61.
- Gut, U. (à paraître). The LeaP corpus. In Durand, J., Gut, U. and Kristoffersen, G. (eds.). *Handbook of Corpus Phonology*. Oxford: Oxford University Press.
- Munro, M. J. 2008. Foreign accent and speech intelligibility. In Hansen Edwards, J. G., Zampini, M. L. (eds), *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins. 193-218.
- Racine, I., Detey, S., Buehler, N., Schwab, S., Zay, F. & Kawaguchi, Y. (2010). The production of French nasal vowels by advanced Japanese and Spanish learners of French: a corpus-based evaluation study. In Deziubalska-Kolaczyk, K., Wrembel, M. & Kul, M. (eds), *Proceedings of New Sounds 2010 - Sixth International Symposium on the Acquisition of Second Language Speech*. Poznan: Adam Mickiewicz University. 367-372.
- Racine, I., F. Zay, S. Detey & Y. Kawaguchi (2011). De la transcription de corpus à l'analyse interphonologique: enjeux méthodologiques en FLE. In G. Col & S.N. Osu, (eds.), *Transcrire, écrire, Formaliser (1)* Rennes: PUR. *Travaux Linguistiques du CerLiCO* 24, 13-30.