FONDS NATIONAL SUISSE
DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ DE GENÈVE

Coder les voyelles nasales dans le corpus IPFC: résultats préliminaires

[1,2]Sylvain Detey, [3]Isabelle Racine, [3]Françoise Zay,
[4]Yuji Kawaguchi, [1]Mariko Kondo & [5]Laurie Buscail

[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

UFR UNIVERSITE DE ROUEN

東京外国語大学 Tokyo University of Foreign Studies

WASEDA University

Université de Toulouse

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'occurence
  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 excrescence)
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-oral vowel	3-off-target	
4-CV			
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)

1x = v nasal
11 = [a~]
12 = [o~]
13 = [e~]
14 = [9~]
 	
2x = v oral corresponding
21 = [an/am]
22 = [on/om]
23 = [en/em]
24 = [un/um]
25 = [in/im]
	10 = ##_
	Ex: (C'est##) en ...
	
	2 = #_
	*Add P/N/F/L/S/V
	Très[z] inquiet
	Ou
	Très [z] inquiet
	
	3 = C_
	*Add P/N/F/L/S
	Ex: blanc: 3L
		100 = _## C
		Ex: ...pain. C'est...
		
		20 = _# C (with C NOT liaison consonant)
		*Add P/N/F/L/S
		Ex: pain dur
		30 = _C#(#) (with C NOT liaison consonant)
		Ex: grande (porte)(.),
		*Add P/N/F/L/S
		
		40 = (C)(C)_C(C)
		Ex: amplifier, grandement.
		*Add P/N/F/L/S
		
		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
		
		*Add
		V if oral
		N if nasal
		Ex:
		mon // enfant = 50N
		bon[n]été = 51V
			1- nasal vowel
			Ex: pain = [pa~]
			2- oral vowel + nasalisation (or uncertain)
			Ex: pain = [paa~]
			3- oral vowel
			Ex: pain = [pa]
			
			4 – nasal vowel non realised
			
			Ex. téressé (intéressé)
				1- targetlike for AN/ON/IN/UN=[e~] or [9~]
				*Add 0 if AN/ON/IN
				*Add I if UN=[e~]
				*Add U if UN=[9~]
				
				2- off-target +A/O/I/U/X ([a~], [o~], [e~], [9~], X = other vowel)
					10- no postvocalic consonant (in context_C)
					
					2- postvocalic consonant or appendix (in context _C)
					*Add N, M ou G according to the nature of the sound
					* Add X if other
					
					30-Potential liaison site (context _V)
						1 – nasal vowel in word lists
						
						2– nasal vowel in the text or conversation

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 = [9~]

- 2x = séquence Vorale+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_2diqueraient
- 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 [9~]

  *Ajouter 0 si AN/ON/IN

  *Ajouter I si UN=[e~]

  *Ajouter U si UN=[9~]

- 2- non-conforme +A/O/I/U/X

  ([a~], [o~], [e~], [9~], 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_**2O**_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)

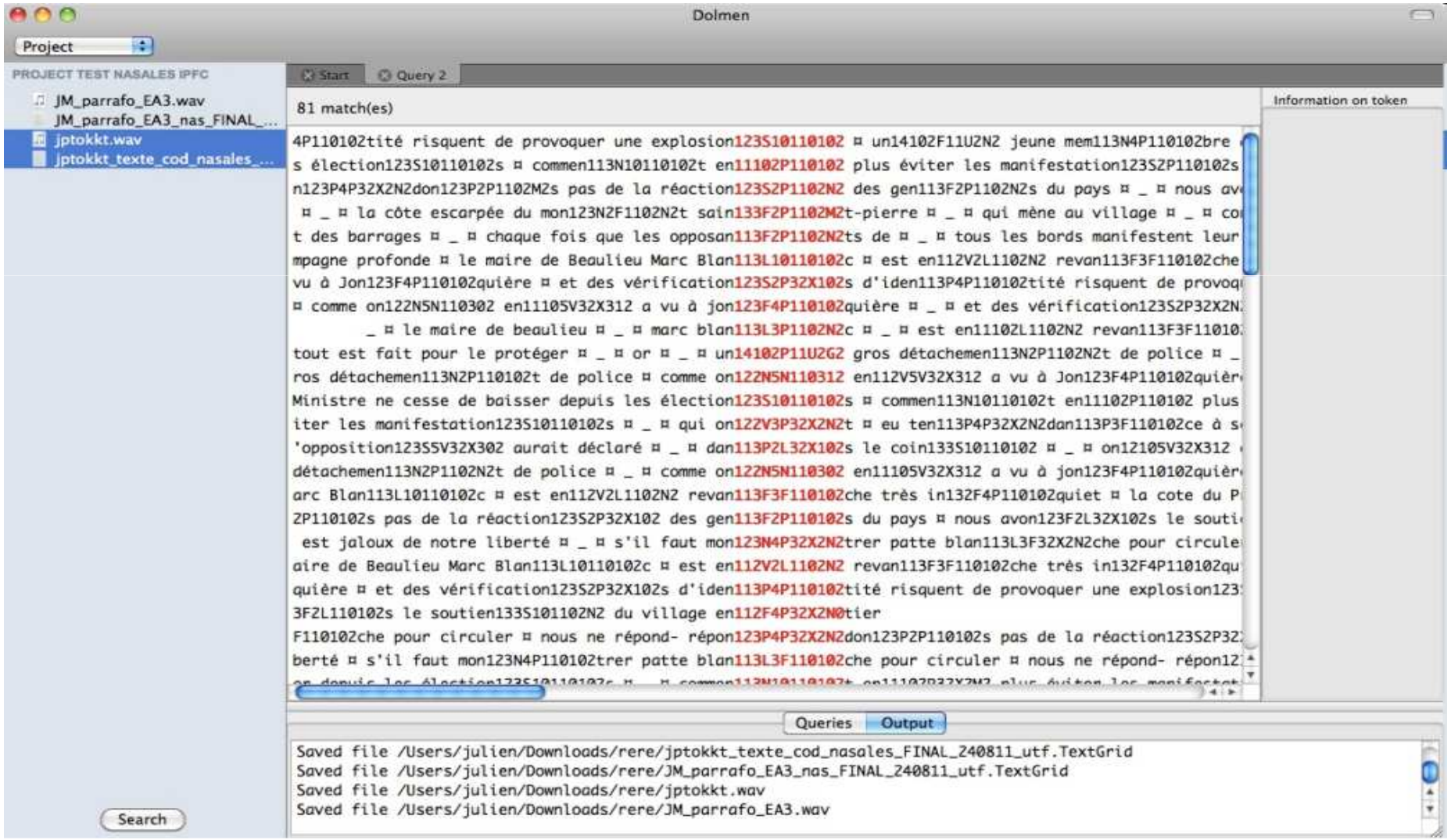
III. Le code pour les voyelles nasales: illustrations et premiers résultats descriptifs

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)



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 répétition qu'en lecture

→ **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). 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.