

Le projet « Interphonologie du français contemporain » : réflexions méthodologiques et premières données d'apprenants hispanophones et japonophones

Isabelle Racine, Sylvain Detey, Françoise Zay & Yuji Kawaguchi

ELCF, Université de Genève,
SILS, Waseda University & LiDiFra, Université de Rouen,
ELCF, Université de Genève,
Tokyo University of Foreign Studies

« Langue française en contexte », colloque AFLS 2009
Université de Neuchâtel, 4 septembre 2009



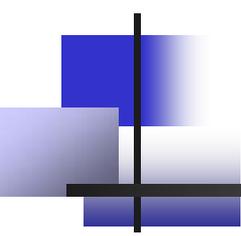
UNIVERSITÉ DE GENÈVE



WASEDA University



東京外国語大学 Tokyo University of Foreign Studies



Documenter la phonologie du FLE : enjeux

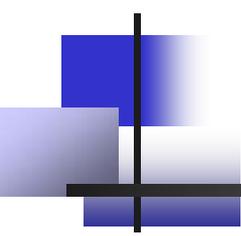
- Nouvel intérêt pour l'enseignement de la prononciation:
 - Aspects communicatifs longtemps privilégiés au détriment de la prononciation
 - Un bon niveau de prononciation en L2 permet d'être intelligible et assure un certain confort d'écoute à l'auditeur (Celce-Murcia et al., 1996 ; Lauret, 2007 ; Morley, 1991).
- Recherches sur l'interlangue des apprenants : interlangue phonique ou interphonologie
- Nécessité de disposer d'inventaires précis et conséquents des productions des apprenants
- Travail sur corpus

Les corpus de FLE : le corpus FLLOC

1) Le projet FLLOC :

- ⇒ « **French Learner Language Oral Corpora** », (Florence Myles, Newcastle Univ. & Rosamond Mitchell, Univ. of Southampton, <http://www.flloc.soton.ac.uk/>)
- Jeunes apprenants FLE anglophones (11-16 ans) de différents niveaux
 - Dimension longitudinale
 - Transcriptions et codage selon normes et outils CHILDES

CHILDES	Child Language Data Exchange System	
CHILDES is the child language component of the TalkBank system. TalkBank is a system for sharing and studying conversational interactions.		



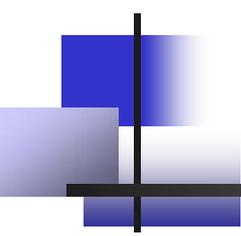
Les corpus de FLE: LANCOM et ESF

2) Le projet LANCOM (Université de Leuven)

- Cf. <http://bach.arts.kuleuven.be/elicop/ProjetLANCOM.htm>
- Apprenants FLE néerlandophones
- Dimension comparative

3) Le corpus ESF (European Science Foundation Second Language)

- Cf. <http://www.mpi.nl/world/tg/lapp/esf/esf.html>
- Ouvriers immigrés en interaction avec des natifs
- Données pour cinq L2 (allemand, anglais, néerlandais, suédois et français)
- Pour le français, 40 sujets arabophones ou hispanophones

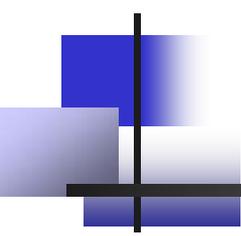


Les corpus de FLE : conclusion

- Recherches axées sur le lexique, la morpho-syntaxe, la syntaxe et la pragmatique;
- Bases de données transcrites dans en fonction de ces objectifs-là;
- Pas de corpus FLE multilingue (plusieurs L1) dédié à l'interphonologie du français avec transcriptions appropriées.
- Besoin de données diversifiées

↙ **Projet**

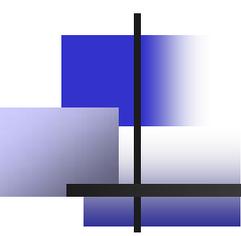
« Interphonologie du français contemporain »



Le projet IPFC (Detey & Kawaguchi, 2008)

- Constituer et mettre à disposition une base de données orales issues d'apprenants de diverses L1.
- Avec un protocole de données calqué sur celui du projet PFC et identique pour toutes les L1.
 - Permet des comparaisons entre apprenants de diverses L1.
 - Permet des comparaisons entre apprenants et natifs de différentes variétés

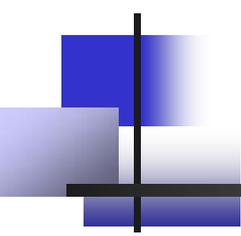
(base de données PFC, Durand, Laks & Lyche, 2002 et 2005, www.projet-pfc.net).



De nouvelles données pour L2

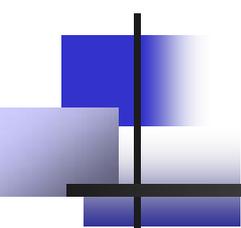
→ besoin de données diversifiées :

- Dans le domaine de l'interphonologie
 - Insuffisances des approches contrastives classiques des 50s et 60s
 - Travaux basés sur des données essentiellement issues de l'anglais langue étrangère
 - Regain d'intérêt dans les années 90 avec les modèles à contraintes (p. ex. OT)
- Dans le domaine de l'enseignement de la prononciation
 - Baqué (2004, 2005): Pour améliorer le matériel didactique, nécessité de disposer d'une vision très précise de l'interlangue des apprenants dans son intégralité et en tenant compte de la L1
 - Développement de systèmes d'entraînement à la prononciation assistée par ordinateur (CAPT) qui intègre un système de reconnaissance automatique (cf. Byrne et al., 1998).
- Plus largement, dans le domaine de l'ingénierie des langues
- Travaux sur des langues autres que le français :
 - ⇒ Neri, Cucchiarini & Strik (2006)
 - ⇒ Gut (2005, 2006 et 2009) : le corpus LeaP



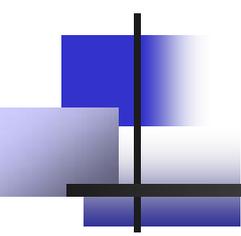
IPFC - Objectifs

- Constitution d'un corpus de recherche FLE varié (différentes L1, différentes tâches) destiné aux linguistes, psycholinguistes et didacticiens s'intéressant à l'apprentissage du français langue étrangère, et plus spécifiquement de son système phonologique.
- Permettra des travaux portant sur :
 - le rôle des facteurs spécifiques aux langues
 - le rôle des facteurs universaux
 - la dimension longitudinale
- Permettra de documenter et d'étudier la variation :
 - intégration de la variation FLE dans le domaine variationnel francophone
 - rôle de la variation en L1 dans l'apprentissage de la L2
 - variation en L2
- Proposera des activités autour de la prononciation du français et de son apprentissage (difficultés typiques et moins typiques)
- Devrait promouvoir / jouer un rôle dans l'image du français langue internationale (observation et familiarisation)



Protocole IPFC

- Constitution de « points d'enquête » formés d'apprenants d'une même variété de L1, de même niveau CECR (20 locuteurs minimum).
- Tâches:
 - Répétition d'une liste spécifique à la L1 concernée, lue par un locuteur natif
 - Lecture de la liste PFC (94 mots)
 - Lecture de la liste spécifique
 - Lecture du texte PFC (« Le Premier Ministre ira-t-il à Beaulieu? »)
 - Entretien guidé (15mn)
 - Conversation entre deux apprenants (25mn)
- Passation:
 - 2 séances d'une heure environ (+ questionnaire biographique à remplir entre les deux séances)



Premières données japonophones et hispanophones

Deux illustrations:

a) Avantages d'un corpus multitâches

- 1. Données japonophones**
- 2. Données hispanophones**

b) Un exemple d'interférence: données hispanophones

Les premières données japonophones

- **Date**: juillet 2009
 - **Lieu**: Tokyo University of Foreign Studies (Tokyo)
 - **Participants**:
 - 1) 34 étudiants de 1ère année non-spécialistes de français
 - 2) 27 étudiants de 1ère année spécialistes de français
 - 3) 28 étudiants de 2ème année spécialistes de français
 - 4) 11 étudiants de 3ème et 4ème année, Master 1, spécialistes de français
- ↳ **TOTAL**: n = **100** sujets (des - experts (1) aux + experts (4))
- **Tâches**: Répétition & Lecture (liste PFC, liste spécifique et texte)
 - **Passation**:
 - Passation automatisée sur Moodle
 - Paramètres d'enregistrement: format PCM, Mono, 22KHz, 16 bits

a) Avantages d'un corpus multitâches:

1. l'exemple des nasales

- Listes de mot:
un mot = un fichier son
⇒ *Multiples comparaisons possibles (inter-, intra-...)*

Exemples de voyelles nasales:

Liste spécifique IPFC Japon

ã/ẽ	Andes/Inde	tante/teinte	tant/teint
ã/õ	anse/once	panse/ponce	pan/pont

Liste générique PFC

ã	9. nous prendrions, 24. bêtement, 59. blanc, 72. vous prendriez
ẽ	8. intact, 20. médecin, 22. infect, 51. influence, 57. brin, 68. cinquième, 94. brin
õ	15. compagnie, 42. blond, 49. million, 50. explosion, 65. compagne
œ	27. brun, 93. brun

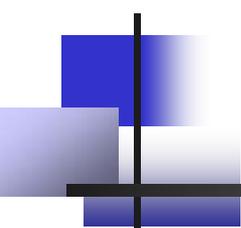
Rendu informatique après passation:

11	repetition	34	Inde	tonique	95-...033-00.wav
12	repetition	59	anse	tonique	95-...058-00.wav
13	repetition	60	Andes	tonique	95-...059-00.wav
14	generique	8	intact	pretonique	96-...007-00.wav
15	generique	9	nous prendrion:	tonique	96-...008-00.wav
16	generique	9	nous prendrion:	pretonique	96-...008-00.wav
17	generique	15	compagnie	pretonique	96-...014-00.wav
18	generique	20	medecin	tonique	96-...019-00.wav
19	generique	24	betement	tonique	96-...023-00.wav
20	generique	27	brun	tonique	96-...026-00.wav
21	generique	41	blond	tonique	96-...042-00.wav
22	generique	49	million	tonique	96-...048-00.wav
23	generique	50	explosion	tonique	96-...049-00.wav
24	generique	51	influence	pretonique	96-...050-00.wav
25	generique	51	influence	tonique	96-...050-00.wav
26	generique	57	brin	tonique	96-...056-00.wav
27	generique	59	blanc	tonique	96-...058-00.wav
28	generique	65	compagne	pretonique	96-...064-00.wav
29	generique	68	cinquieme	pretonique	96-...067-00.wav
30	generique	72	vous prendriez	pretonique	96-...071-00.wav
31	generique	93	brun	tonique	96-...092-00.wav
32	generique	94	brin	tonique	96-...093-00.wav
33	lecture	4	teinte	tonique	97-...003-00.wav
34	lecture	9	teint	tonique	97-...008-00.wav
35	lecture	11	tant	tonique	97-...010-00.wav

a) Avantages d'un corpus multitâches:

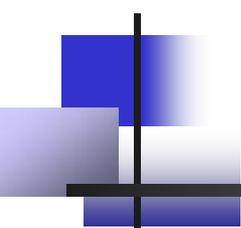
1. l'exemple des nasales

Comparaison		Novices (1 ^e - non-spécialistes)		Experts (3 ^e -4 ^e -5 ^e -spécialistes)	
		Homme	Femme	Homme	Femme
TEINTE vs TANTE	Répétition (A)				
	Lecture (B)				
TANTE vs TANT	Répétition (C)	9	10	11	12
	Lecture (D)	13	14	15	16
REPETITION vs LECTURE	Teinte, Tante, Tant (E)				



Les premières données hispanophones

- **Date**: juin 2009
- **Lieu**: ELCF, Université de Genève
- **Participants**:
 - 4 étudiants espagnols de la filière DEFLE (niveau B2-C1)
- **Tâches**: protocole complet
- **Passation**:
 - Passation semi-automatisée en laboratoire de langue
 - Paramètres d'enregistrement: format WAV, Mono, 22KHz, 16 bits
- **Traitement des données**:
 - Listes et texte transcrits orthographiquement avec alignement son-texte sous Praat (Boersma & Weenink, 2009)



a) Avantages d'un corpus multitâches: 2. confusion b/v

- Systèmes de la L1 et de la L2:
 - Système espagnol
 - Un phonème avec 2 réalisations allophoniques:
 - /b/ réalisé [β] en position intervocalique
Ex. vaca [**b**aka] vs una vaca [unaβaka]
 - Pas de /v/
 - Système français
 - 2 phonèmes:
 - /b/ (occlusive bilabiale sonore)
 - /v/ (fricative labio-dentale sonore)
- Quelques occurrences issues des 4 tâches:
 - Répétition de la liste spécifique
 - Lecture de la liste spécifique
 - Lecture du texte
 - Conversation

a) Avantages d'un corpus multitâches: 2. confusion b/v

	Source: base	Source: vase
Répétition	 	  
Lecture	 	 
Lecture texte	  	
Conversation		

Remarques:

- Mauvaise discrimination des deux phonèmes.
- Bonne articulation des deux phonèmes sur une base graphique après un an de travail régulier en laboratoire de langue.
- Réalisations correctes, erreurs et autocorrections dans la lecture du texte.
- Erreurs sans autocorrection en conversation.

⇒ **La mauvaise discrimination n'a pas pour conséquence une mauvaise articulation pour autant qu'il y ait une base graphique identifiée.**

b) Un exemple d'interférence

Quel est le mot?



l'arabe



dorer



adorer

⇒ **Confusion [g, ʁ, ʁ]**



b) Un exemple d'interférence

- En espagnol:
 - /r/ roulé apical
 - /g/ occlusive vélaire sonore
 - /g/ réalisé [ɣ] (spirante vélaire sonore) en position intervocalique.
 - Ex. gato [gato] vs hago [aɣo] 
- En français:
 - /R/ réalisé [ʁ] = fricative uvulaire sonore
 - /g/ occlusive vélaire sonore
- Lors de la mise en place du système:
 - Recul du lieu d'articulation pour produire le R français
 - L1 = 2 réalisations sonores dans cette zone [g, ɣ]
 - Mode d'articulation proche pour [ʁ] et [ɣ]
 - Pour acquérir le [ʁ], l'apprenant part de [ɣ] et recule le point d'articulation (Companys, 1966).

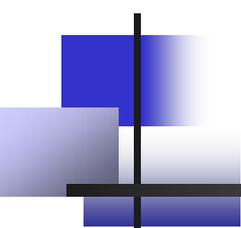
b) Un exemple d'interférence

	PALATAL	VELAR	UVULAR	PHARYNGEAL	GLOTTAL
PLOSIVE	c ʃ	k ɡ	q ɢ		ʔ
NASAL	ɲ	ŋ	ɴ		
TRILL			ʀ		
TAP OR FLAP					
FRICATIVE	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
LATERAL FRICATIVE					
APPROXIMANT	j	ɥ			
LATERAL APPROXIMANT	ʎ	ʟ			

Tableau tiré de <http://www.phonetics.ucla.edu/course/chapter1/consonants2.html>

b) Un exemple d'interférence

- Mais cela engendre des interférences entre [g, ʁ, ʁ]:
 - [g] est réalisé [ʁ]
⇒ possibles difficultés de compréhension     
 - [ʁ] est réalisé [ʁ]
⇒ Quelle catégorisation? 
⇒ Différence hors contexte/en contexte?
 - [ʁ] est réalisé [g]
⇒ compréhension très difficile voire impossible à comprendre 
 - Problème dans le groupe [gʁ]  
⇒ difficultés de compréhension

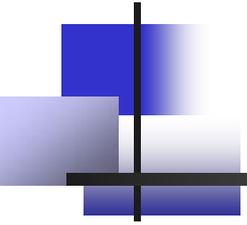


Conclusion

- Ces données posent des questions intéressantes au niveau de la perception par des natifs:
 - Le remplacement d'un phonème de L2 par un autre phonème de L2 est problématique: [ɣ] par [g]
 - Le remplacement d'un phonème de L2 par un allophone avec le même mode et un lieu d'articulation similaire semble moins problématique, notamment en contexte large: [ɣ] par [ɣ̞].

Perspectives:

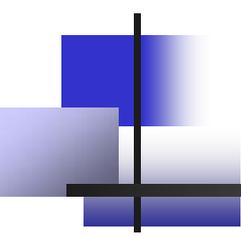
- Que faire au niveau de la transcription du corpus:
 - Viser un niveau phonologique?
 - Viser un niveau phonétique?
 - Les deux?



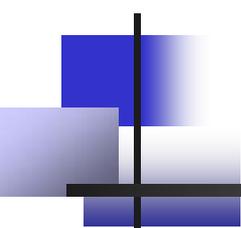
Merci!

Références

- Baqué, L. (2004). Analyse de l'interlangue de productions phoniques déviantes : un préalable pour la programmation de correction phonétique. Dans : J. Suso López et R. López Carrillo (eds.), *Le français face aux défis actuels : Histoire, langue et culture*. Granada, vol. 1, 293-302.
- Baqué, L. (2005). Effet sur la prononciation d'apprenant de FLE de la tâche et de la focalisation sur la forme. *Actes du 1er Colloque International de Didactique Cognitive*, Toulouse, France.
- Boersma, P., Weenink, D. (2009). *Praat : doing phonetics by computer* (version 5.0), <http://www.praat.org>.
- Celce-Murcia, M., Brinton, D. M. & Goodwin, J. M., (1996). *Teaching Pronunciation*, Cambridge : CUP.
- Companys, E. (1966). *Phonétique française pour hispanophones*, Paris : Hachette/Larousse.
- Detey, S. et Kawaguchi, Y. (2008). Interphonologie du Français Contemporain (IPFC) : récolte automatisée des données et apprenants japonais. *Journées PFC : Phonologie du français contemporain : variation, interfaces, cognition*, Paris, 11-13 décembre 2008.
- Durand, J., Laks, B. & Lyche, C. (2002). La phonologie du français contemporain: usages, variétés et structure. In : C. Pusch & W. Raible (eds.), *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language*. Tübingen : Gunter Narr Verlag, 93-106.
- Durand, J., Laks, B. & Lyche, C. (2005). Un corpus numérisé pour la phonologie du français. In G. Williams (ed.), *La linguistique de corpus*. Rennes : Presses Universitaires de Rennes, p. 205-217. Actes du colloque 'La linguistique de corpus', Lorient, 12-14 septembre 2002.
- Gut, U. (2005). Corpus based pronunciation training, PTLC2005.
- Gut, U. (2006). DGfS Bielefeld, 22th February 2006.
- Gut, U. (2009). Non-native Speech: A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German. Wien: Peter Lang.
- Lauret, B. (2007). *Enseigner la prononciation du français: questions et outils*. Paris : Hachette.
- Morley, J. (1991) The pronunciation component in teaching English to speakers of other languages, *TESOL Quarterly*, 25, 481-519.
- Neri, A., Cucchiarini, C. & Strik, H. (2006). Selecting segmental errors in non-native Dutch for optimal pronunciation training, *IRAL - International Review of Applied Linguistics in Language Teaching*, 44, 357-404.

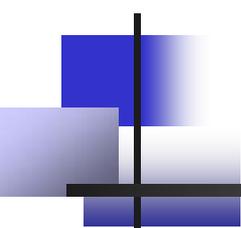


Dias supplémentaires



1. Liste spécifique

- Répétition puis lecture d'une liste comprenant:
 - 36 mots avec difficultés communes aux différentes L1 (ex. u/y, voyelles nasales, R, etc.)
 - Difficultés : u/y, voyelles nasales, R/l, etc.
 - Testées dans plusieurs contextes: **Andes/Inde**, **tante/teinte**, **tant/teint**
 - 25-30 mots avec difficultés spécifiques à la L1:
 - Hispanophones: b/v, occl. sonores en position initiale/finale vs. intervocalique, etc.
 - Japonophones: f/l + u, s/j + i, groupes de 4 consonnes
 - Difficultés testées dans plusieurs contextes :
 - Hispanophones: **base/vase**, la **base/la vase**, l'**arabe/la rave**
 - Japonophones: **foule/houle**, la **foule/la houle**
 - Tâches (entraînement préalable de 3 mots):
 - Pour la répétition, mots lus 2x par un natif (1 sec. d'intervalle entre les deux occurrences, 4 sec. entre les mots)
 - Pour la lecture, mots présentés
 - Sur support papier pour la préparation (5mn)
 - Sur support informatique avec défilement (1 mot toutes les 4 secondes)

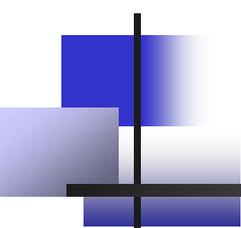


2. Texte

- 391 mots divisés en 3 paragraphes
- 20mn de préparation
- Présentation d'un bref résumé du texte en L1
- Extrait:

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. Jusqu'ici les seuls titres de gloire de Beaulieu étaient son vin blanc sec, ses chemises en soie, un champion local de course à pied (Louis Garret), quatrième aux jeux olympiques de Berlin en 1936, et plus récemment, son usine de pâtes italiennes. Qu'est-ce qui a donc valu à Beaulieu ce grand honneur ? Le hasard, tout bêtement, car le Premier Ministre, lassé des circuits habituels qui tournaient toujours autour des mêmes villes, veut découvrir ce qu'il appelle "la campagne profonde". [...]



3. Entretien guidé

- Questions présentées sur support papier en ordre fixe
- Durée : env. 15mn

Liste fermée:

1. Quel âge avez-vous et quelle est votre nationalité ?
2. Quelles langues parlez-vous ?
3. Quand avez-vous commencé à étudier le français ?
4. Où avez-vous commencé à étudier le français ?
5. Pourquoi avez-vous commencé à étudier le français?

Liste ouverte (niveaux A1-B1):

6. Décrivez le chemin pour aller de chez vous à la gare la plus proche.
7. Avez-vous déjà voyagé dans un pays étranger ?
8. Si oui, racontez votre/vos voyage(s).
9. Si non, dans quels pays aimeriez-vous voyager et pourquoi ?
10. Quels sont vos projets pour le futur (professionnels, personnels, etc.) ?
etc.

Liste ouverte (niveaux B2-C1):

6. Pouvez-vous raconter votre première leçon de français ?
7. (Si immersion) Pourquoi avoir choisi [lieu d'accueil] pour étudier le français ?
8. Avez-vous l'impression que le français de [lieu d'accueil] est identique à celui que vous avez appris dans votre pays ? Si non, quelles sont les différences que vous avez remarquées ?
9. Avez-vous déjà rencontré des difficultés à comprendre certaines personnes qui parlent français et qui sont francophones ? Si oui, où ?

4. Conversation entre 2 apprenants

- Objectif: obtenir une interaction entre les apprenants
- Discussion entre deux apprenants sur un thème choisi par chacun d'entre eux avec l'enquêteur parmi la liste fermée suivante :

Liste :

1. Un film vu
 2. Un livre vu
 3. Un endroit visité (pays, ville, musée, etc.)
 4. Un aspect de votre pays (plat, coutume, manière de vivre, fête)
 5. Une différence culturelle entre votre pays et un autre pays
 6. Un souvenir marquant (enfance, famille, école, voyage, anecdote, 1^{ère} leçon de français)
 7. Un thème d'actualité
 8. Un fait divers
- Durée: env. 25 mn (temps de préparation: 10mn)
 - Remarques:
 - Le choix du thème doit être communiqué à l'enquêteur avant la préparation
 - La liste de thèmes peut être traduite en L1 si nécessaire.
 - Consigne aux apprenants : utilisation de la L1 uniquement en cas de blocage conversationnel.
 - Pas d'utilisation de notes pendant la présentation du thème.